Justifying the Construct Definition for a New Language Proficiency Assessment: The Redesigned *TOEIC Bridge*® Tests — Framework Paper

Jonathan Schmidgall
Maria Elena Oliveri
Trina Duke
Elizabeth Carter Grissom

December 2019

ETS Research Report Series

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.

Justifying the Construct Definition for a New Language Proficiency Assessment: The Redesigned *TOEIC Bridge*® Tests—Framework Paper

Jonathan Schmidgall, Maria Elena Oliveri, Trina Duke, & Elizabeth Carter Grissom

Educational Testing Service, Princeton, NJ

One of the most critical steps in the test development process is defining the construct, or the knowledge, skills, or abilities, to be assessed. This foundational step provides the basis for initial assumptions about the meaning of test scores and serves as a reference for subsequent validity research. In this paper, we describe the purpose of the redesigned *TOEIC Bridge*® 4 skills assessments and elaborate the theoretical basis of its construct definition. We also describe how an evidence-centered design (ECD) approach was used to develop the redesigned TOEIC Bridge assessments and the first stage of that approach, the domain analysis. The domain analysis begins by elaborating a clearer definition of the context in which language is evaluated by the redesigned TOEIC Bridge assessments, "everyday adult life." Next, we review research literature and relevant language proficiency standards to highlight the knowledge, skills, and abilities relevant to beginner to low-intermediate English proficiency for everyday adult life. This information is synthesized in the construct definitions for reading, listening, speaking, and writing ability for beginner to low-intermediate levels of general English proficiency in the context of everyday adult life.

Background

In this framework paper, we describe the purpose of the redesigned *TOEIC Bridge*® tests and justification of their construct definitions. In doing so, we elaborate the rationale for the interpretation and use of test scores. This is a foundational step in the test design process that provides the basis for initial assumptions about the meaning of test scores and serves as a reference for subsequent validity research (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Bachman & Palmer, 2010).

We begin with a discussion of the purpose and intended uses of the assessment and key stakeholder groups and propose a logic model that outlines the relationships among assessment components, intended uses, and intended outcomes. This forms the basis of a mandate for test design. It also establishes connections among test purpose, test design, and validation (Fulcher, 2013).

We contextualize the rest of the framework paper within an evidence-centered design (ECD) approach to test design and development (Mislevy, Almond, & Lukas, 2003). Although the ECD approach consists of five layers of analysis, the framework paper focuses primarily on the first layer, domain analysis.

Our approach to domain analysis reflects an interactionalist approach to construct definition, in which context and abilities interact to form the construct (Bachman, 2007). Thus, we begin by elaborating a clearer definition of our language use domain, "everyday adult life." Next, we survey research literature and relevant developmental proficiency standards to highlight the knowledge, skills, and abilities relevant to beginner to low-intermediate general English proficiency. This information is synthesized in our definitions of the constructs of reading, listening, speaking, and writing ability for beginner to low-intermediate levels of general English proficiency in the context of everyday adult life.

Corresponding author: Jonathan Schmidgall, E-mail: jschmidgall@ets.org

Test Purpose and Intended Uses

The redesigned TOEIC Bridge tests measure beginning to low-intermediate English language proficiency in the context of everyday adult life. In order to accommodate the particular needs of score users, the redesigned TOEIC Bridge tests will include modules for listening and reading, speaking, and writing. If score users are interested in an evaluation of overall language proficiency or communicative competence, all four skills should be tested.

The tests are primarily intended to be used for selection, placement, and readiness purposes. Some score users may wish to use the test to determine whether applicants to vocational or training institutions have a threshold level of English proficiency that is needed or desirable (i.e., selection) to benefit from further English language training. Other score users may use information about English proficiency for the purpose of placing students or employees into English language training courses or programs of study at beginner to low-intermediate proficiency levels. Additionally, some score users (i.e., test takers) may wish to use the information obtained about their English proficiency to determine their readiness to take TOEIC tests or for more advanced study.

Several secondary uses of the test were also considered in the design of the test. Some score users may want to use test section scores to track or benchmark development or improvement over time in order to monitor growth in language skills or overall proficiency. Others may wish to use subscores or other performance feedback in order to identify their relative strengths and weaknesses with respect to different language skills.

Stakeholders

The stakeholders of a test are those who are either directly affected (primary stakeholders) or indirectly affected (secondary stakeholders) by the use of the test (Bachman & Palmer, 2010). Those directly affected—primary stakeholders—are the individuals whose proficiency is being evaluated (test takers) and those who use the scores to make important decisions (score users, including teachers). Those indirectly affected—secondary stakeholders—are the individuals who may have a stake in the use of the test due to its impact on their work or experience (e.g., teachers who are not necessarily score users).

Test takers are young adults (high school/secondary school and older) and adults for whom English is a second or foreign language, and their nationalities and native languages (L1) will vary. Test takers' educational backgrounds and purpose for learning English (e.g., general purposes, academic purposes, occupational purposes) may also vary. Score users will typically be administrators (e.g., at vocational training institutions) and managers (e.g., at organizations and institutions). Teachers may be primary or secondary stakeholders and will be affected if the redesigned TOEIC Bridge tests are used for placement into language training courses. Teachers may also benefit from the use of the test to track proficiency and potentially monitor progress and the use of any information provided by the test to inform remedial instruction.

A Logic Model for Redesigned TOEIC Bridge Tests

Ultimately, tests are used to promote particular outcomes, effects, or consequences. With this in mind, intended outcomes should be elaborated from the beginning of a test design project and inform the design of the test itself (Bachman & Palmer, 2010; Norris, 2013). Bachman and Palmer (2010) advanced this view through the use of an argument-based approach to test use, which begins with test developers consulting with score users to establish claims about desirable outcomes (e.g., hiring employees with appropriate English language skills). Test developers then work backward to determine the types of decisions that facilitate the intended outcomes (e.g., a selection decision), the interpretations about abilities needed to facilitate equitable decisions, the scores that are needed to facilitate meaningful and impartial interpretations, and finally, characteristics of test performances needed to produce scores that are reliable or consistent.

Another approach that establishes a link between test components, intended uses, and outcomes is the theory of action (Bennett, 2010; Patton, 2002, pp. 162–164). The theory of action uses a logic model to illustrate how components of the test (such as scores) are expected to facilitate particular actions (i.e., decisions), which in turn are intended to produce particular effects (i.e., outcomes or consequences). In the logic model, arrows indicate hypothesized causal links: For example, an arrow between test components and a particular action mechanism implies a claim about the relevance of the test for a particular use. When fully developed, the logic model is expanded to a theory of action by providing documentation that explicitly states each claim and provides a summary of the evidence backing the claim.
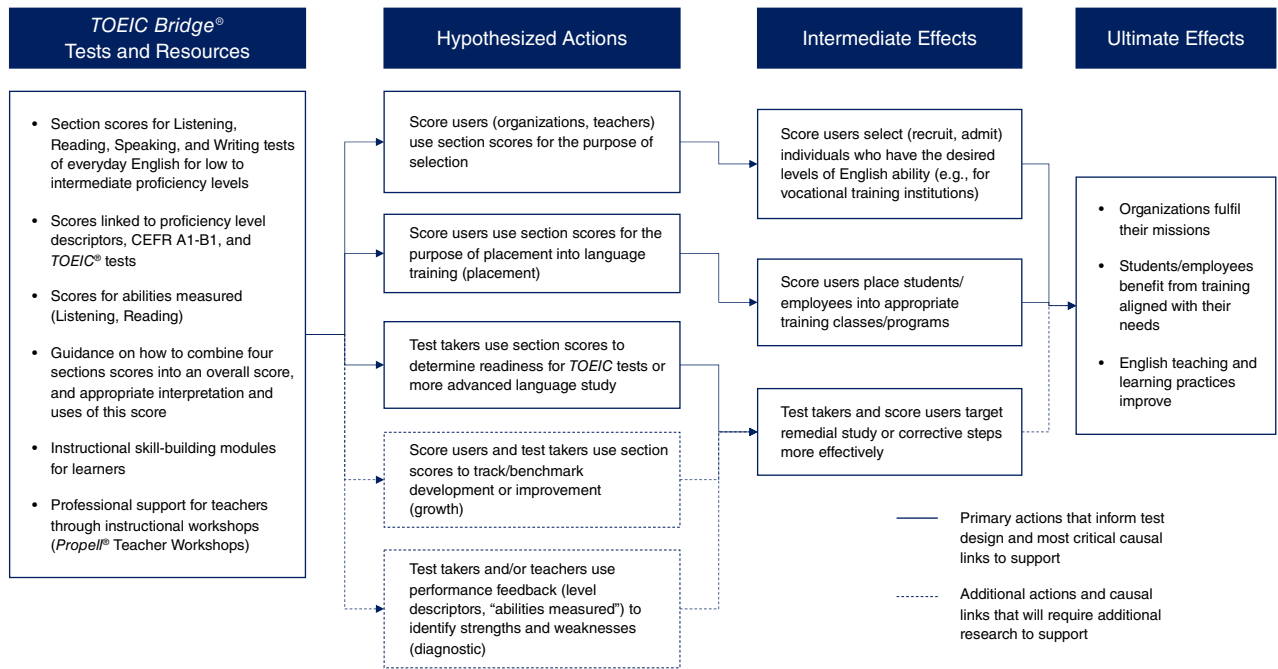
TOEIC Bridge® Tests and Resources	Hypothesized Actions	Intermediate Effects	Ultimate Effects

- Section scores for Listening, Reading, Speaking, and Writing tests of everyday English for low to intermediate proficiency levels
- Scores linked to proficiency level descriptors, CEFR A1-B1, and *TOEIC®* tests
- Scores for abilities measured (Listening, Reading)
- Guidance on how to combine four sections scores into an overall score, and appropriate interpretation and uses of this score
- Instructional skill-building modules for learners
- Professional support for teachers through instructional workshops (*Propell®* Teacher Workshops)

Score users (organizations, teachers) use section scores for the purpose of selection

Score users use section scores for the purpose of placement into language training (placement)

Test takers use section scores to determine readiness for *TOEIC* tests or more advanced language study

Score users and test takers use section scores to track/benchmark development or improvement (growth)

Test takers and/or teachers use performance feedback (level descriptors, "abilities measured") to identify strengths and weaknesses (diagnostic)

Score users select (recruit, admit) individuals who have the desired levels of English ability (e.g., for vocational training institutions)

Score users place students/ employees into appropriate training classes/programs

Test takers and score users target remedial study or corrective steps more effectively

- Organizations fulfil their missions
- Students/employees benefit from training aligned with their needs
- English teaching and learning practices improve

——— Primary actions that inform test design and most critical causal links to support

- - - - - Additional actions and causal links that will require additional research to support

Figure 1 A logic model for the redesigned TOEIC Bridge tests.

As a preliminary step, we specified a logic model for the redesigned TOEIC Bridge tests that reflects their purpose and intended uses (see Figure 1). These uses are formalized in the diagram as hypothesized actions. Each hypothesized action is expected to produce intermediate and ultimate effects. Based on the actions and effects we intend to support and promote, we specified components of the tests that we believe are necessary.

In the logic model, there are three primary hypothesized actions that the test will be designed to support: selection, placement, and determining readiness for TOEIC tests or more advanced study. There are two additional hypothesized actions that the test developer would like to support, identified in dashed boxes in the logic model: monitoring growth or progress and using test information to identify learners' strengths and weaknesses. Several components of the redesigned TOEIC Bridge tests will be necessary to support these actions: test section scores, and mapping or concordance with external standards (e.g., Common European Framework of Reference [CEFR] A1 to B1) and TOEIC tests. We intend these actions to have specific intermediate and ultimate effects.

Evidence-Centered Design and Test Development

With the intended uses, effects, and test components specified in the logic model, we began to conceptualize the design of the test within an ECD framework (Mislevy, Almond, & Lukas, 2003; Mislevy, Steinberg, & Almond, 2003). ECD is a systematic approach to test design that helps identify, map, and categorize activity patterns associated with a particular context or practice to render test takers' implicit behaviors and attitudes observable and assessable in an operational assessment. Although conceived as a general approach to test design and development, ECD has been utilized by several language assessment programs (Chapelle, Enright, & Jamieson, 2008; Hines, 2010; Kenyon, 2014).

The ECD model has five layers: (a) domain analysis, (b) domain modeling, (c) the conceptual assessment framework (CAF), (d) assessment implementation, and (e) assessment delivery (Mislevy & Yin, 2012). Each layer includes different concepts and entities, representations, purposes, and questions. There is an implied iteration between these layers as developers move back and forth between the layers. Figure 2 illustrates the roles, associated activities, and resulting activity for the first three layers of ECD (Riconscente, Mislevy, & Corrigan, 2015). The red boxes identify the aspects of the ECD process that are addressed by this framework paper.

The purpose of the first layer, domain analysis, is to identify the key attributes that define the constructs of interest. In language assessment, construct definition typically entails elaborating ability-in-context (Bachman, 2007): knowledge, skills, and abilities (KSAs) and the target language use (TLU) domain. Activities at this stage of ECD typically include
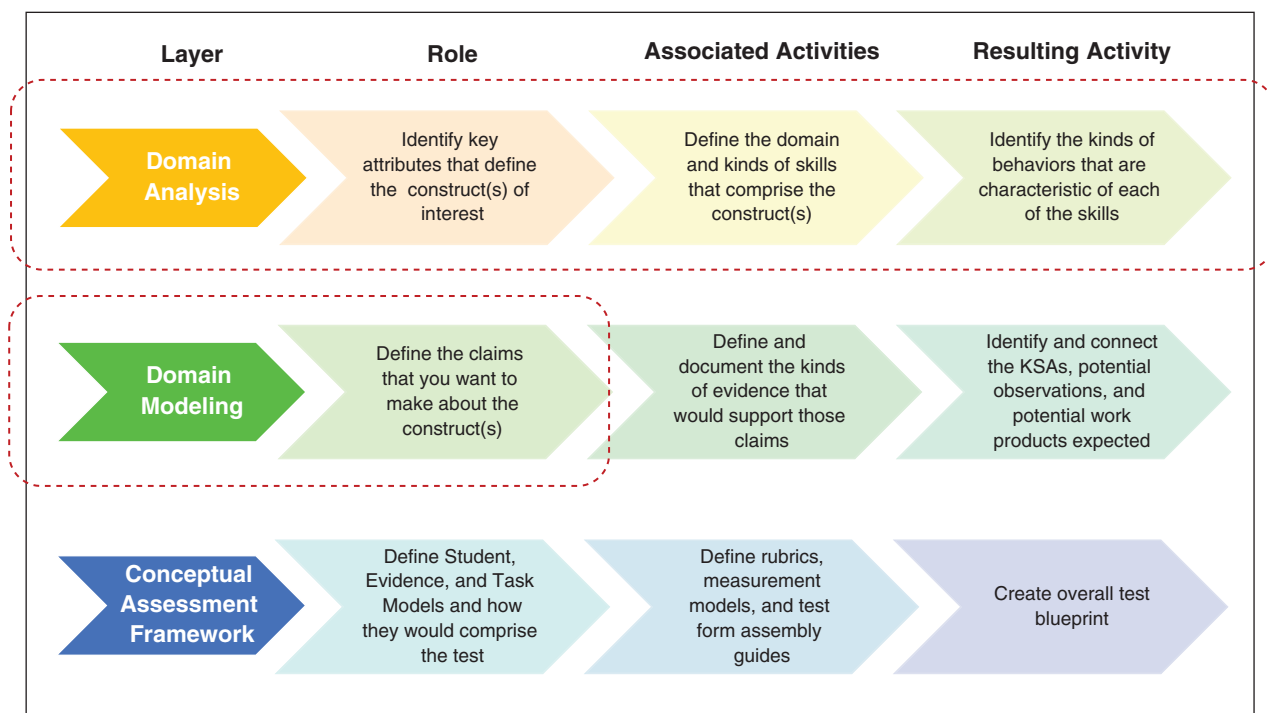
Figure 2 Activities within the first three layers of the evidence-centered design assessment development process, and the focus of the framework paper.

conducting systematic literature reviews of frameworks, taxonomies, and assessments and may include consulting with subject-matter experts and industry-related stakeholders to identify the key features of the construct(s) of interest, the kinds of skills that comprise it, and the kinds of behaviors that characterize each skill.

In the second layer, domain modeling, the information gleaned in the domain analysis is parsed into assessment design patterns (Wei, Mislevy, & Kanal, 2008). Design patterns elaborate key attributes of the test, including its rationale, focal KSAs, potential observations, characteristic features, and variable features. They form the initial narrative for the design of the test and the basis for the development of test specifications in subsequent ECD layers.

The third layer of ECD is the CAF, which is used for the assembly of the entire assessment by generating a test blueprint (which should include the desired performances to elicit and work products to capture, the features of tasks or items, and constraints for the development of the assessment). The CAF includes the student, evidence, and task models that specify the elements of an operational assessment design (Mislevy, Steinberg, & Almond, 2003). The student model is conceptualized in terms of the construct, assessment purpose, and the target population(s). The evidence model structures thinking about the kinds of performances (their salient features captured as observable variables) that provide evidence of a test taker's standing on the KSAs as deemed important for the construct. Considerations for how to elicit the desired evidence about the defined construct occur in the task model. These considerations include identifying the types of situations necessary to best elicit behaviors that demonstrate proficiency in the desired KSAs. All of the information from the design patterns is brought together to populate the student, evidence, and task models. The assessment is specified in terms of its content, how it will be delivered, features of the test-taking environment, and test administration instructions. The CAF documents how items/tasks can be varied to create additional test forms. It also documents how test developers update their beliefs about test takers' proficiency based on their work products. In other words, the CAF specifies the operational elements, models, and data structures that instantiate the assessment argument. It structures the data that will be produced and makes sense of them in a way that permits interpretable and meaningful score-based inferences, in accordance with the assessment argument. The CAF also serves another purpose: examining the impact the assessment may have on test takers and different populations. Reviewing the elements of the operational assessment at this stage helps the developer ensure that inferences from the overall performances are appropriate and the construct coverage is adequate.

After the assessment is deployed operationally (see Mislevy & Yin, 2012, for a discussion of the assessment delivery and assessment implementation layers), the ECD-based assessment argument can be extended into an assessment use argument using a formal argument-based approach to validation (e.g., Bachman & Palmer, 2010; Kane, 2011). Evidence collected throughout the ECD process can provide initial backing to support claims about test scores, score interpretations, and test use.

Domain Analysis: Conceptualizing Beginner to Low-Intermediate General English Proficiency for Everyday Adult Life

Language proficiency may be conceptualized as ability-in-context or from an interactionalist perspective (e.g., Bachman, 2007; Chalhoub-Deville, 2003; Chapelle, 1998; Xi, 2015). This involves three essential components: the language knowledge required to facilitate performance, communicative strategies to support performance, and a description of the performance context itself. The performance context is often referred to as the TLU domain (Bachman & Palmer, 2010). Once the TLU domain is broadly defined (e.g., everyday adult life), communicative tasks that are typical of the domain are identified and their features are elaborated using a task characteristic framework (e.g., Bachman & Palmer, 2010) or another principled approach to specifying contextual features of tasks (e.g., Xi, 2015). The underlying language knowledge (e.g., lexical knowledge) and processes (e.g., lexical retrieval) needed to successfully perform tasks in the domain form another component. Communicative strategies are often linked to particular tasks and reflect the use of language to achieve a communicative purpose or functional goal and may be articulated more broadly (e.g., reading to find information) or narrowly (e.g., ability to identify essential information in complex sentences in text). Documentation of these components is the product of the domain analysis stage and provides the basis for domain modeling, the next layer in the ECD process.

Figure 3 illustrates how the stages of the domain analysis described in this section were structured to provide the basis for construct definition for redesigned TOEIC Bridge tests.

The domain analysis began with a review of literature that may inform the definition of the TLU domain, English for everyday adult life (see the subsection entitled Defining the TLU Domain of Everyday Adult Life). The purpose of this activity was to elaborate the contextual features of the TLU domain (i.e., general features of the setting) relevant to construct definition and test design. We then summarized literature related to the constructs of listening comprehension, reading comprehension, speaking proficiency, and writing proficiency for second or foreign language (L2) learners in the subsection entitled Defining English Reading, Listening, Speaking, and Writing Proficiency. These initial reviews provided the theoretical basis for construct definition within an interactionalist approach, highlighting relevant abilities and contexts that should be incorporated into construct definitions for redesigned TOEIC Bridge tests. Given the mandate to evaluate proficiency at beginner to low-intermediate levels—and map test-based interpretations about proficiency with levels of the CEFR, Canadian Language Benchmarks (CLB), and American Council on the Teaching of Foreign Languages (ACTFL) language proficiency standards—we then conducted a thorough review of relevant levels of these standards with our definitions of proficiency in mind in the subsection entitled Defining Beginner to Low-Intermediate English Proficiency. This evaluation informed the refined version of the construct definition, presented in the section Construct Definition for an Assessment of Beginning to Low-Intermediate English Language Proficiency for Everyday Life.

Defining the Target Language Use Domain of Everyday Adult Life

Broadly, researchers make a distinction between general and specific-purpose TLU domains (Douglas, 2000). This distinction is made based on the degree to which the TLU domain is concretely and narrowly specified; in other words, the communicative context of a general purposes domain is more varied and resistant to precise description (Douglas, 2001). Although definitions of general and specific-purpose domains are typically based on a theoretical model of language ability or acquisition, the nature of specific-purpose domains facilitates a more detailed analysis of relevant communicative tasks and language abilities.

Although a broad distinction between general and specific-purpose domains can be maintained, it might be helpful to view the specificity of TLU domains as a continuum with general purposes on one end and specific purposes on the other (Knoch & Macqueen, 2016). TLU domains that are more narrowly and concretely defined (e.g., English for aviation) will have a higher degree of specificity than those that are more broadly or abstractly defined (e.g., English for the workplace).
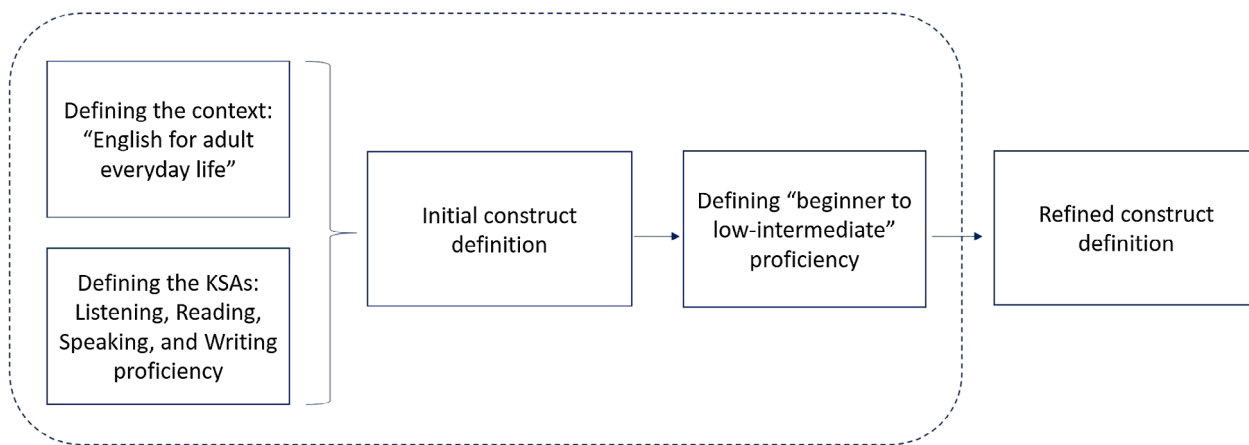
Figure 3 Domain analysis as the basis of construct definition for redesigned TOEIC Bridge tests.

When the degree of specificity is high, the language abilities and contextual features relevant to the domain can be more clearly articulated. For more general domains where the degree of specificity is low, researchers or test developers may need to rely on taxonomies to describe features of the TLU domain that should be represented in the assessment procedure to facilitate generalizations about language abilities.

The TLU domain of everyday adult life as conceptualized for the redesigned TOEIC Bridge tests is expected to fall toward the general-purposes end of a specificity continuum. Given this conceptualization of the TLU domain, we considered a number of relevant taxonomies to further elaborate what the general, everyday adult life TLU domain may or may not include. Our review of relevant literature and test documentation identified four approaches that could contribute to the conceptualization of everyday adult life: the social-ecological model of concentric circles, the CEFR for languages standards, the ACTFL proficiency guidelines, and the *TOEFL*® family of assessments. Our initial review of the CLB noted that discussion of the context of language use primarily focuses on differentiating *nondemanding* (common everyday activities) and *demanding* (educational and work-related) contexts and was generally aligned with the CEFR's approach; consequently, we did not include it in our summary. Below, we briefly summarize relevant information from each of the four approaches reviewed in depth.

Social-Ecological Model of Concentric Circles

One way to consider the TLU domain of everyday life is through the lens of ecological models that specify a set of nested social contexts. The social-ecological model (Bronfenbrenner, 1979) originated as a model of human development and describes an ecological system composed of five socially organized subsystems that support human development. It is conceptualized as a set of concentric circles that are centered on the individual (the microsystem); extends to family, peers, and other intimates (the mesosystem); then to neighbors, extended family, and less intimate others (the exosystem); and beyond that to a context that reflects norms from cultural values, customs, and laws (the macrosystem). Given that this model was conceived in the context of development, changes that occur in individuals or the environments within these subsystems over time are accounted for in a fifth subsystem (the chronosystem). According to Bronfenbrenner (1979), human development occurs through progressively more complex interactions between the individual and the people, objects, and symbols in the individual's environment. These interactions are called *proximal processes*. Together, process, person, and context form the core of the ecological model.

Although originally conceived for general child development, this model has been applied to many other fields, including L2 development. Van Lier (2000) related the model to Vygotsky's sociocultural theory (Vygotsky, 1978). Learners develop by engaging in different learning contexts, or proximal processes, analogous to Vygotsky's zone of proximal development (ZPD). These subsystems may roughly translate to a variety of TLU domains or subdomains, each engaging the learner in a different set of proximal processes. As proficiency increases, test takers are able to interact with the increasingly less familiar, moving from their immediate social network to the broader community or culture and social norms, and from concrete ideas to more abstract concepts.

The Council of Europe Framework of Reference Standards

The CEFR standards describe four broad domains of language use: personal, public, educational, and occupational. The personal domain involves "family relations and individual social practices," whereas the public domain involves "ordinary social interaction (business and administrative bodies, public services, cultural and leisure activities of a public nature, relations with media, etc.)" (CEFR, 2009, p. 15). The educational domain relates to "the learning/training context … where the aim is to acquire specific knowledge or skills," and the occupational domain focuses on "a person's activities and relations in the exercise of his or her occupation" (Council of Europe, 2009, p. 15).

In practice, these domains may overlap in various ways. For retail workers, the occupational and public domains may largely overlap. For teachers, the educational and occupational domains may overlap. Even in these cases, however, distinctions between these broad domains may be useful to maintain. For example, the communicative skills needed by retail workers to interact in the occupational domain differ somewhat from the skills needed to interact in the public domain, given the differences between the roles and responsibilities of employees and customers. The communicative skills required by students in training courses—even teacher training courses—differ somewhat from those required by the teachers of those training courses.

The American Council on the Teaching of Foreign Languages Proficiency Guidelines

The ACTFL proficiency guidelines (ACTFL, 2012) do not formally define language use domains or subdomains, although they provide descriptions of relevant contexts of language use at each level of proficiency. The notion of "everyday contexts" is elaborated in terms of topics of communication related to survival in the target language culture, such as communicating basic personal information, basic objects, and a limited number of activities, preferences, and immediate needs as well as responding to simple, direct questions or requests for information.

The guidelines note that everyday tasks and communicative functions might be expressed in different forms depending on whether speech or writing is presentational (one-way, noninteractive) or interpersonal (i.e., interactive, two-way communication). For example, for writing, tasks and communicative functions may include lists, short messages, postcards, and simple notes (presentational) or they may include instant messaging, e-mail communication, and texting (interpersonal).

The TOEFL Family of Assessments Approach to Domain Definition

The TOEFL family of assessments includes the *TOEFL iBT*®, *TOEFL ITP*®, *TOEFL Junior*®, and *TOEFL Primary*® assessments. Although these assessments are designed to evaluate English proficiency in the context of English-medium education (i.e., academic TLU domain), their overall approach to conceptualizing the TLU domain was considered for how it may be adapted for our purposes. The TOEFL family of assessments' conceptualization of the academic TLU domain includes subdomains that include social-interpersonal, academic navigational, and academic content (see So et al., 2015). Two of these subdomains—social-interpersonal and academic-navigational—have potential relevance to the domain of everyday language use.

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

A second use is communicating for navigational purposes, such as communicating with peers, teachers, and other school staff about school- and course-related materials and activities but not about academic content. For example, students communicate about homework assignments to obtain and clarify details. In some cases, they need to extract key information from school-related announcements. That is, students need to communicate to navigate school or course information. The second subdomain captures this specific purpose of communication.

Although the TLU domain targeted by the TOEFL Junior test pertains to young learners, language activities are generally meaning focused and intended to replicate a variety of real-life communication contexts. Language activities are
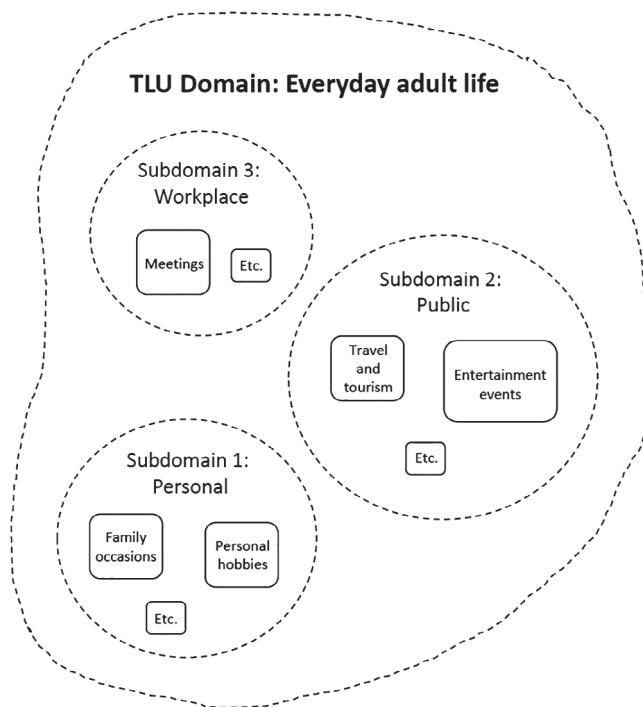
Figure 4 Components of the target language use domain of everyday adult life.

typically organized around a theme (e.g., my weekend) to allow learners to use learned expressions in a variety of settings relevant to young learners (e.g., plan a weekend with a classmate, survey the class on favorite weekend activities). The language use contexts replicated in the English as a foreign language (EFL) classroom are largely social, meaning that learners primarily use language to communicate with people around them (e.g., family, friends, classmates, teachers) on familiar topics (e.g., myself, animals, people) and to obtain basic information from familiar sources (e.g., stories, announcements, directions).

Summary

Although our review did not identify any formal attempt to define the more general-purpose TLU domain of English for everyday adult life, the approach advocated by the authors of the CEFR standards was useful. Specifically, this approach suggested that language is primarily used in personal and public contexts at lower proficiency levels and branches out into specific-purpose domains (academic or occupational) at intermediate to advanced levels. Given the purpose of the assessment—measuring English proficiency at beginning to low-intermediate levels—personal and public contexts should be well represented in the domain definition for everyday adult life. Figure 4 provides a visual representation of this domain.

As shown in Figure 4, the TLU domain of everyday adult life is a more general-purpose domain that emphasizes tasks and contexts that are expected to be familiar to adults and young adults. TLU subdomains include personal, public, and some more general and familiar aspects of occupational or workplace contexts. Within each subdomain, there are settings that are expected to be more familiar. An example of a familiar setting within the personal subdomain might be family occasions or settings that relate to personal hobbies and interests. Within the public subdomain, familiar settings may include travel and tourism, entertainment events, and shopping. Only the most general workplace settings (i.e., those that require no industry-specific experience to understand) would be considered relevant to the workplace subdomain.

In addition to specifying the subdomains and settings typical of the TLU domain of everyday adult life, it is important to consider other contextual features of the setting that may need to be represented in language use tasks included in the assessment. In their framework of language task characteristics, Bachman and Palmer (2010) elaborated characteristics of the setting, rubric, input, expected response, and relationship between input and expected response that should

be considered when describing or developing language use tasks for the purpose of assessment. Several of these characteristics are worth noting, as the degree to which they are represented in assessment tasks may constrain or facilitate generalization about language proficiency to the TLU domain of everyday adult life. For productive language use (i.e., speaking and writing), care should be taken to identify the role of the test taker and his or her intended audience in order to simulate the interactional nature of everyday adult life. As researchers have observed, English communication often occurs between L2 users of English who use English as a lingua franca (McNamara, 2011), and so the intended audience in the TLU domain of everyday adult life may include both native (L1) and nonnative (L2) users of English. In addition, the topical characteristics of tasks should reflect the subdomains (personal, public, workplace) included in the TLU domain.

Defining English Reading, Listening, Speaking, and Writing Proficiency

Our review of the components or elements of English language proficiency is influenced by the module-based design of the redesigned TOEIC Bridge tests, as score users may be interested in more selective (e.g., comprehension skills only, speaking ability only) or more comprehensive (i.e., four skills) information about language proficiency for the purpose of decision-making. The conceptualization of overall language ability as multicomponential, consisting of four modalities (reading, listening, speaking, writing) and linguistic elements (e.g., grammar, vocabulary, phonology, sociopragmatics) reflects the view of many researchers (Purpura, 2004).

Listening comprehension has generally been conceptualized through the use of cognitive processing models of listening comprehension or component models of listening ability. Cognitive processing models attempt to identify the phases of cognitive processing and resources involved between the reception of an acoustic (and potentially visual) signal and a listener's response (e.g., Bejar, Douglas, Jamieson, Nissan, & Turner, 2000; Field, 2013; Rost, 2005). Component models of listening ability are ontological representations that are influenced by models of communicative competence (e.g., Bachman, 1990), and typically include the higher order components of language competence and strategic competence (Buck, 2001; Weir, 2005). In Buck's model — largely based on Bachman and Palmer's (1996) framework of communicative competence — language competence consists of declarative and procedural knowledge related to listening, including grammatical, discourse, pragmatic, and sociolinguistic knowledge. Strategic competence includes cognitive and metacognitive strategies that are related to listening.

Reading comprehension is typically conceptualized as the process or product of a reader's interaction with a text (Alderson, 2000; Koda, 2013). The process-oriented view conceptualizes reading comprehension as the process of a reader interacting with a text, while the product-oriented view focuses on the product of this interaction, typically demonstrated by answering comprehension questions that require readers to recall the product (i.e., the aspect of comprehension elicited by the question) from memory (Koda, 2013). In both views, the reader may comprehend a text to a greater or lesser degree, and the notion of the processing demands of comprehension questions may be useful to differentiate the degree of overall comprehension. In a comprehensive review of the construct of reading comprehension, researchers have also elaborated the "reader purpose" view, largely complementary to both processing and product perspectives (Jamieson, Jones, Kirsch, Mosenthal, & Taylor, 2000). In the reader purpose view, reading comprehension is conceptualized as the interaction between reader linguistic and processing abilities, reader purpose, and text characteristics for a given reading task. Thus, a construct definition for reading should elaborate the range of relevant skills and strategies needed by the reader given the purposes and text characteristics involved for the targeted reading tasks. In a meta-analysis of the relationship between reading component variables and passage-level reading comprehension, Jeon and Yamashita (2014) found that L2 grammar knowledge, L2 vocabulary knowledge, and L2 decoding were the strongest predictors of L2 reading comprehension.

Models of speaking proficiency — much like those of listening comprehension — can be characterized as cognitive processing or component models. Cognitive processing models of speech production typically involve four phases: conceptualization, formulation, implementation, and self-monitoring. One of the best known models in second language speech production is Levelt's (1989) parallel model, which hypothesizes knowledge bases that inform the conceptualization phase (e.g., discourse and background knowledge) and formulation phase (lexical, grammatical, and phonological knowledge). Component models of speaking ability have been much more widely utilized in language assessment and generally correspond to models of communicative competence such as that of Bachman and Palmer (2010), who suggested that language ability reflects an interaction between strategic competencies and language knowledge. In Bachman

and Palmer's model, language knowledge is composed of organizational knowledge (grammatical and textual knowledge) and pragmatic knowledge (functional and sociolinguistic knowledge), and strategic competencies involve goal setting, appraising, and planning. Some researchers have attempted to refine component models that are perceived to lack sensitivity to contextual features by emphasizing the importance of pragmatic competencies (e.g., Purpura, 2004), or greater representation of contextual facets of tasks in the construct definition (e.g., Xi, 2015).

Writing proficiency is often broadly conceptualized using process-oriented cognitive models (Weigle, 2002), and it is considered in second or foreign language contexts using task-based approaches that elaborate important features of writing tasks such as subject matter, discourse mode (genre, audience, purpose), and stimulus materials (Weigle, 2013). In this task-oriented approach, writing ability is essentially defined by the ability to produce written texts in accordance with the purpose of the task (e.g., to inform, to persuade), follow conventions of the genre (e.g., explanatory writing, transactional writing), and consider the needs of the intended audience (e.g., laypersons, academic specialists). The underlying linguistic knowledge and resources needed to demonstrate writing ability may vary by domain, task, and proficiency levels but often refer to elements such as content, organization, vocabulary use, mechanics, and grammar (e.g., Jacobs, Zingram, Wormuth, Hartfiel, & Hughey, 1981; Weir, 1990).

Two overarching themes are present in much of the research related to one or more of these four skills as well as broader conceptualizations of language ability (e.g., Bachman, 1990). The first theme is that language is used with communicative goals in mind. For the modalities of reading and listening, the communicative goal is to understand written or spoken texts with particular characteristics (e.g., a particular genre) for a strategic purpose (e.g., for implied meaning, for the main idea). For speaking and writing, the communicative goal is achieved by successfully performing a specific communicative task (e.g., making a request, describing an activity). The second theme is that linguistic knowledge and subcompetencies (i.e., language knowledge and skills) are used to achieve communicative goals. Most components of linguistic knowledge and subcompetencies are utilized across modalities of communication and have been articulated in more general models of communicative competence or language ability (e.g., Bachman, 1990). These components include lexical, grammatical, discourse, phonological, and orthographic knowledge of language, as well as pragmatic and strategic competencies.

We incorporated these two themes (communicative goals; linguistic knowledge and subcompetencies) into our initial construct definitions for each of the four skills. For example, the initial construct definition for *speaking proficiency in everyday adult life* included a list of important communicative goals for speakers in the TLU domain (e.g., expressing an opinion) and a broad set of linguistic skills and subcompetencies needed to realize various communicative goals (e.g., the ability to use high-frequency vocabulary appropriate to a task, or lexical knowledge and use).

An important aspect of conceptualizing English language proficiency that is often overlooked is whether the underlying proficiency model emphasizes native-like competence or communicative effectiveness (Hu, 2017). This aspect is particularly relevant for conceptualizing and evaluating speaking proficiency, where emphasis may be placed on the accuracy of form in relation to the norms of a particular variety of English (i.e., emphasizing native-like competence) or on the comprehensibility and communicative impact of speech (i.e., communicative effectiveness). Given the recognition that a speaker's or writer's audience in the domain of English for everyday adult life may include native or nonnative speakers of English, an underlying proficiency model based on communicative effectiveness (as opposed to native-like competence) will inform the construct definition, development, and scoring of the redesigned TOEIC Bridge tests.

Defining Beginner to Low-Intermediate English Proficiency

In the third phase of the domain analysis, we closely examined descriptors of language proficiency standards relevant to the modalities (reading, listening, speaking, writing) and components of linguistic knowledge and subcompetencies identified in the previous phase. This analysis served two purposes. First, one of the mandates for test design—as documented in the initial logic model—was the need to map test scores to language proficiency standards to enhance the interpretation of test scores. Incorporating information from language proficiency standards during the test design stage provides stronger evidence of alignment (Council of Europe, 2009). The second purpose of this analysis was to produce artifacts that could inform task and scoring rubric design. Whereas the prior review of theory and research literature helped inform the construct definition for each test section, it provided minimal guidance on the types of communication goals and linguistic knowledge and subcompetencies expected of L2 users at different levels of proficiency (e.g., communicative goals appropriate for a low beginner versus a high beginner).

With this background in mind, we identified levels of the CEFR standards (Council of Europe, 2018), the CLB (Centre for Canadian Language Benchmarks, 2012), and the ACTFL proficiency guidelines (ACTFL, 2012) that were relevant to the range of beginner to low-intermediate English proficiency. Given the mandate to target the assessment of proficiency from CEFR Levels Pre-A1 to B1, we reviewed relevant descriptors across this range. Based on a study that mapped ACTFL proficiency levels to CEFR levels (Bärenfänger & Tschirner, 2012), we reviewed descriptors associated with ACTFL proficiency levels up to the intermediate high level for speaking and writing, and up to the advanced low level for reading and listening. We also reviewed descriptors associated with CLB Levels 1 to 6 based on Vandergrift's (2006) proposed alignment between CLB and CEFR levels. The CEFR and CLB include overall descriptor scales for each modality as well as more detailed scales that describe more specific activities, competencies, or strategies associated with each modality. Since the CEFR includes a wide range of descriptor scales, we restricted our review to scales relevant to the initial construct definition (see Appendix A for a full list of the CEFR descriptor scales that were reviewed).

For each modality (reading, listening, speaking, writing), we aggregated information across standards and relevant descriptor scales that aligned with CEFR Levels Pre-A1, A1, A2, and B1. For example, for the speaking beginner level (CEFR A1, CLB 1–2, ACTFL novice high), we summarized information in relevant descriptors as they pertained to communication goals, topics, characteristics of the input, and linguistic skills and subcompetencies (lexical knowledge, grammatical knowledge, discourse knowledge, phonological knowledge, pragmatic competence). The summary produced for the speaking beginner level is reproduced in Appendix B.

Although the overall structure of our construct definitions was not affected by the analysis of language proficiency standards, the analysis helped us refine some of the language in our construct definitions. The analysis allowed us to cross-validate our lists of communication goals by comparing them to communicative activities highlighted within and across standards. We also refined some of the language used to describe different linguistic skills and subcompetencies based on standards-based descriptors of these skills at the low-intermediate level. Thus, the analysis did not have a major impact on the components of language ability that were included in the construct definitions (e.g., communicative goals, various linguistics skills and subcompetencies) that were theoretically derived; rather, it helped us refine or cross-validate our expectations of how these components would be realized for beginner to low-intermediate learners.

Construct Definition for an Assessment of Beginning to Low-Intermediate English Language Proficiency for Everyday Adult Life

In this section, we present the proposed construct definition for each of the proposed redesigned TOEIC Bridge tests. The construct for each test section is based on the interactionalist approach to construct definition (described in the Background subsection) and reflects a theoretical approach in which language proficiency is demonstrated by using linguistic knowledge and subcompetencies to achieve communication goals. This overall approach of focusing on communication goals and linguistic knowledge and subcompetencies in context (for each of the four language skills) was based on the reviews described in the subsections Defining the Target Language Use Domain of Everyday Life and Defining English Reading, Listening, Speaking, and Writing Proficiency.

The construct definition for each test begins with a broad statement about what the test intends to measure and then lists the communication goals relevant to the use of English at beginning to low-intermediate levels in the context of everyday adult life. This statement is followed by an elaboration of the specific linguistic knowledge and subcompetencies needed to achieve the communication goals. The categories of linguistic knowledge and subcompetencies (i.e., lexical, grammatical, discourse, phonological, and orthographic knowledge; pragmatic and strategic competence) are generally consistent across all four tests. For each test section, the communication goals and linguistic knowledge and subcompetencies listed also reflect our principled analysis of relevant language proficiency standards (CEFR, CLB, ACTFL). As previously described in the subsection Defining Beginner to Low-Intermediate English Proficiency, this analysis helped refine specific elements of each construct definition and produced artifacts that were used to guide the test development process.

The redesigned TOEIC Bridge tests are a measure of the ability of beginning and low-intermediate learners of English to communicate in personal, public, and general workplace contexts and to comprehend and produce basic spoken and

written texts commonly occurring in everyday adult life. The construct definitions for each test section (listening, reading, speaking, and writing) are found in Appendix C.

Concluding Comments

This paper described the process used to produce a construct definition for a new suite of language proficiency tests, the redesigned TOEIC Bridge tests. The process followed a mandate-driven approach to ECD. This approach began by defining the mandate for test design, including test purpose and intended uses, stakeholders, and a logic model that specified assessment components, hypothesized actions (intended uses), and hypothesized intermediate and long-term effects (impact or consequences of test use). Based on this mandate, a domain analysis was conducted that further elaborated the TLU domain (i.e., English for everyday adult life) and targeted language proficiency competencies (i.e., reading, listening, speaking, and writing skills). In order to facilitate alignment between the assessment and language proficiency standards, produce artifacts that could support the next stages of the test development process (i.e., domain modeling and the CAF), and further refine the initial construct definition based on the targeted proficiency levels, we analyzed relevant descriptors from three language proficiency standards (CEFR, CLB, and ACTFL).

The outcome of this work is a proposed construct definition for each test that is based on theory, research, and relevant language proficiency standards. The construct definition reflects an interactionalist approach that specifies characteristics of the TLU domain (e.g., setting, audience, communication goals) and relevant linguistic skills and subcompetencies. These construct definitions provide a basis for the next steps in the ECD process — domain modeling and development of the CAF — as well as justification for the intended meaning of test scores and intended uses of the test. In addition, the construct definitions provide the basis for subsequent evaluations of interpretations and uses — validity research — based on the actual ensuing assessment.

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511732935

American Council on the Teaching of Foreign Languages. (2012). *ACTFL proficiency guidelines 2012*. Alexandria, VA: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Author.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. New York, NY: Oxford University Press.

Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, and C. Doe (Eds.), *Language testing reconsidered* (pp. 41–72). Ottawa: University of Ottawa Press. https://doi.org/10.2307/j.ctt1ckpccf.9

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. New York, NY: Oxford.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. New York, NY: Oxford.

Bärenfänger, O., & Tschirner, E. (2012). *Assessing evidence of validity of assigning CEFR ratings to the ACTFL Oral Proficiency Interview (OPI) and the Oral Proficiency Interview by computer (OPIc)*. Retrieved from http://www.global8.or.jp/OPIc%20CEFR%20Study%20Final%20Report%20pdf.pdf

Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL 2000 Listening framework: A working paper* (TOEFL Monograph Series MS-19). Princeton, NJ: Educational Testing Service.

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives, 8*, 70–91. https://doi.org/10.1080/15366367.2010.508686

Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.

Buck, G. (2001). *Assessing listening*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511732959

Centre for Canadian Language Benchmarks. (2012). *Canadian language benchmarks: English as a second language for adults*. Ottawa, ON: Author.

Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing, 20*(4), 369–383. https://doi.org/10.1191/0265532203lt264oa

Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. Bachman and A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139524711.004

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

Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg, France: Author.

Council of Europe. (2018). Common European Framework of Reference for Languages: Learning, teaching, assessment. *Companion volume with new descriptors*. Retrieved from https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge, UK: Cambridge University Press.

Douglas, D. (2001). Language for specific purposes assessment criteria: Where do they come from? *Language Testing, 18*(2), 171–185. https://doi.org/10.1177/026553220101800204

Field, J. (2013). Cognitive validity. In A. Geranpayeh and L. Taylor (Eds.), *Examining listening* (pp. 77–151). Cambridge, UK: Cambridge University Press.

Fulcher, G. (2013). Test design and retrofit. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. New York, NY: Blackwell Publishing. https://doi.org/10.1002/9781405198431.wbeal1199

Hines, S. (2010). Evidence-centered design: The TOEIC Speaking and Writing tests. *TOEIC Compendium Study*. Princeton, NJ: Educational Testing Service.

Hu, G. (2017). The challenges of world Englishes for assessing English proficiency. In E. Low and A. Pakir (Eds.), *World Englishes: Rethinking paradigms*. New York, NY: Taylor and Francis.

Jacobs, H. L., Zingram, D. R., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.

Jamieson, J., Jones, S., Kirsch, I., Mosenthal, P., & Taylor, C. (2000). *TOEFL 2000 framework: A working paper*. (TOEFL Monograph Series Report No. 16). Princeton, NJ: Educational Testing Service. Retrieved from https://www.ets.org/Media/Research/pdf/RM-00-03-Jamieson.pdf

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning, 64*(1), 160–212. https://doi.org/10.1111/lang.12034

Kane, M. T. (2011). Validating score interpretations and uses. *Language Testing, 29*(1), 3–17. https://doi.org/10.1177/0265532211417210

Kenyon, D. (2014, May). *From test development to test use consequences: What roles does the CEFR play in a validity argument*? Invited keynote presentation at the annual conference of the European Association of Language Testing and Assessment, University of Warwick, Coventry, UK.

Knoch, U., & Macqueen, S. (2016). Language assessment for the workplace. In D. Tsagari & J. Baneerjee (Eds.), *Handbook of second language assessment* (pp. 291–307). Berlin, Germany: De Gruyter Mouton.

Koda, K. (2013). Assessment of reading. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. New York, NY: Blackwell Publishing. https://doi.org/10.1002/9781405198431.wbeal0051

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

McNamara, T. (2011). Managing learning: Authority and language assessment. *Language Teaching, 44*(4), 500–515. https://doi.org/10.1017/S0261444811000073

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design*. (Research Report No. RR-03-16.) Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2003.tb01908.x

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement*, *1*(1), 3–62. https://doi.org/10.1207/S15366359MEA0101_02

Mislevy, R. J., & Yin, C. (2012). Evidence-centered design in language testing. In G. Fulcher and F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 208–222). New York, NY: Routledge.

Norris, J. M. (2013, October). *Reconsidering assessment validity at the intersection of measurement and evaluation*. Invited plenary address at the annual conference of the East Coast Organization of Language Testers, Georgetown University, Washington, DC.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Purpura, J. (2004). *Assessing grammar*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511733086

Riconscente, M. M., Mislevy, R. J., & Corrigan, S. (2015). In S. Lange, T. M. Haladyna, & M. Raymond (Eds.), *Handbook of test development* (2nd ed., pp. 40–63). London, UK: Routledge.

Rost, M. (2005). L2 listening. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*. Mahwah, NJ: LEA.

So, Y., Wolf, M. K., Hauck, M. C., Mollaun, P., Rybinski, P., Tumposky, D., & Wang, L. (2015). *TOEFL Junior® design framework* (Research Report No. RR-15-13). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12058

Van Lier, L. (2000). From input to affordance: Social-interactive learning from an ecological perspective. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 245–259). New York, NY: Oxford University Press.

Vandergrift, L. (2006). *New Canadian perspectives: Proposal for a common framework of reference for languages in Canada*. Ottawa, ON: Canadian Heritage.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

Wei, H., Mislevy, R. J., & Kanal, D. (2008). An introduction to design patterns in language assessment. *PADI Technical Report, 18*, 1–24.

Weigle, S. C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511732997

Weigle, S. C. (2013). Assessment of writing. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. New York, NY: Blackwell Publishing. https://doi.org/10.1002/9781405198431.wbeal0056

Weir, C. J. (1990). *Communicative language testing*. New York, NY: Prentice Hall.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. New York, NY: Palgrave Macmillan. https://doi.org/10.1057/9780230514577

Xi, X. (2015, March). *Language constructs revisited for practical test design, development and validation*. Paper presented at the 37th Annual Language Testing Research Colloquium, Toronto.

Appendix A

Common European Framework of Reference Descriptor Scales Included in the Review

Communicative language activity	Descriptor scales included
Reading comprehension	• Overall reading comprehension • Reading correspondence • Reading for orientation • Reading for information and argument • Reading instructions • Identifying cues and inferring
Listening comprehension	• Overall listening comprehension • Understanding conversation between other speakers • Listening as a member of a live audience • Listening to announcements and instructions • Listening to audio media and recordings • Identifying cues and inferring
Spoken production	• Overall spoken production • Sustained monologue: describing experience • Sustained monologue: giving information • Sustained monologue: putting a case • Public announcements
Spoken interaction	• Informal discussion • Obtaining goods and services • Information exchange • Phonological control
Written production	• Overall written production • Creative writing • Written reports and essays
Written interaction	• Overall written interaction • Correspondence • Notes, message, and forms
Other (interaction strategies, linguistic, sociolinguistic, pragmatic)	• Online conversations and discussion • General linguistic range • Vocabulary range • Grammatical accuracy • Vocabulary control • Thematic development • Coherence and cohesion • Propositional precision • Spoken fluency • Sociolinguistic appropriateness

Appendix B

Summary of Scale Descriptors Relevant to the Speaking Construct Definition at Common European Framework of Reference Level A1 (and CLB Levels 1 to 2, American Council on the Teaching of Foreign Language Level Novice High) From Language Proficiency Standards

Category	Summary
Communication goals	● Ask and respond to simple, direct questions and statements (CEFR, CLB, ACTFL) ● Description (CEFR, CLB) ● Read a short, prepared/rehearsed statement (CEFR) ● Use and respond to basic courtesy formulas and greetings (CEFR, CLB) ● Give brief, common, routine instructions (CLB) ● Express basic ability or inability (CLB) ● Limited number of activities and preferences (ACTFL); Express likes and dislikes (CLB)
Topics	● People, places (CEFR) ● Areas of immediate need or very familiar topics, such as asking for assistance, or the time, price, or an amount (CEFR, CLB, ACTFL); Very simple warnings and cautions (CLB) ● Very basic personal information: description, occupation, surroundings (CEFR, CLB, ACTFL) ● Basic everyday, routine communication (CLB) ● Straightforward social situations (ACTFL) ● Basic objects (ACTFL)
Characteristics of the input	● Slower speech rate (CEFR) ● Questions and instructions addressed carefully and slowly; short, simple directions (CEFR) ● Allow rephrasing and repair (CEFR, ACTFL) ● Sympathetic or supportive interlocutor (CEFR, CLB, ACTFL)
Linguistic skills and subcompetencies Lexical knowledge and use	● Common, familiar words (CLB); money, prices, amounts (CLB); sizes, colors, numbers (CLB); concrete objects (CLB); likes and dislikes (CLB); numbers, quantity, cost, time (CEFR) ● Formulaic expressions (CLB); common greetings, introductions, and leave-takings (CEFR, CLB) ● May significantly impede communication (CLB) ● Numbers and dates, name, nationality, address, age, date of birth, etc. (CEFR) ● Basic vocabulary repertoire of isolated words and phrases related to particular concrete situations (CEFR)
Grammatical knowledge and use	● Simple phrases (CEFR) ● Imperative forms (CLB); both positive and negative commands (CLB) ● Tend to use present tense (CLB, ACTFL) ● Little or no control over basic grammar structures and tenses (CEFR, CLB, ACTFL) ● May significantly impede communication (CLB)
Discourse knowledge and use	● Mainly isolated words or phrases, no or little evidence of connected discourse (CEFR, CLB) ● Link words or simple phrases with very basic linear connectors such as "and" or "then" (CEFR) ● Short conversational openings and closings (CEFR, CLB)
Phonological knowledge and use	● Not adequate to sustain simple conversations (CLB, ACTFL) ● Slow speech rate with frequent pauses, hesitations, repetitions; rephrasing and repair (CEFR, CLB, ACTFL) ● Pronunciation difficulties may significantly impede communication (CLB) ● Use alphabet to spell out words, such as name (CLB)
Pragmatic competence	● Use appropriate courtesy words (CLB)

Appendix C

Construct Definitions for the Redesigned TOEIC Bridge Tests:
Listening, Reading, Speaking, and Writing

Listening

The redesigned TOEIC Bridge Listening test measures the ability of beginning to lower-intermediate English language learners to understand short spoken conversations and talks in personal, public, and familiar workplace contexts. This includes the ability to understand high-frequency vocabulary, formulaic phrases, and the main ideas and supporting details of clearly articulated speech across familiar varieties of English on familiar topics. Test takers can comprehend simple greetings, introductions, and requests; instructions and directions; descriptions of people, objects, situations; personal experiences or routines; and other basic exchanges of information.

Communication Goals

In English, test takers can understand commonly occurring spoken texts, demonstrating the ability to

- understand simple descriptions of people, places, objects, and actions;
- understand short dialogues or conversations on topics related to everyday life (e.g., making a purchase); and
- understand short spoken monologues as they occur in everyday life (e.g., an announcement in a public area) when they are spoken slowly and clearly.

Linguistic Knowledge and Subcompetencies

To achieve these goals, beginning and lower-intermediate English language learners need the ability to

- understand common vocabulary and formulaic phrases (lexical knowledge);
- understand simple sentences and structures (grammatical knowledge);
- understand sentence-length speech and some common registers (discourse knowledge);
- recognize and distinguish English phonemes and the use of common intonation and stress patterns and pauses to convey meaning in slow and carefully articulated speech across familiar varieties (phonological knowledge);
- infer implied meanings, speaker roles, or context in short, simple spoken texts (pragmatic competence); and
- understand the main idea and stated details in short spoken texts (listening strategies).

Reading

The redesigned TOEIC Bridge Reading test measures the ability of beginning and lower-intermediate English language learners to understand short written English texts in personal, public, and familiar workplace contexts and across a range of formats. This includes the ability to understand high-frequency vocabulary, formulaic phrases, and the main ideas and supporting details of short written texts dealing with familiar topics. Test takers can comprehend simple texts such as signs, lists, menus, schedules, advertisements, narrations, routine correspondence, and short descriptive texts.

Communication Goals

In English, test takers can understand commonly occurring written texts, demonstrating the ability to

- understand nonlinear written texts (e.g., signs, schedules);
- understand written instructions and directions;
- understand short, simple correspondence; and
- understand short informational, descriptive, and expository written texts about people, places, objects, and actions.

Linguistic Knowledge and Subcompetencies

To achieve these goals, beginning and lower-intermediate English language learners need the ability to

- understand common vocabulary (lexical knowledge);
- understand simple sentences and structures (grammatical knowledge);
- understand the organization of short written texts in a variety of formats (discourse knowledge);
- recognize simple mechanical conventions of written English (orthographic knowledge);
- infer implied meanings, including context or writer's purpose, in short, simple written texts (pragmatic competence); and
- understand the main idea and stated details in short written texts; infer the meaning of unknown written words through context clues (reading strategies).

Speaking

The TOEIC Bridge Speaking test measures the ability of beginning and lower-intermediate English language learners to carry out spoken communication tasks in personal, public, and familiar workplace contexts. This includes the ability to communicate immediate needs, provide basic information, and interact on topics of personal interest with people who are speaking clearly. Test takers can answer simple questions on familiar topics and use phrases and sentences to describe everyday events. They can provide brief reasons for and explanations of their opinions and plans and narrate simple stories.

Communication Goals

In spoken English, perform simple communication tasks, demonstrating the ability to

- ask for and provide basic information;
- describe people, objects, places, activities;
- express an opinion or plan and give a reason for it;
- give simple directions;
- make simple requests, offers, and suggestions; and
- narrate and sequence simple events.

Linguistic Knowledge and Subcompetencies

To achieve these goals, beginning and lower-intermediate English language learners need the ability to

- use high-frequency vocabulary appropriate to a task (lexical knowledge);
- use common grammar structures to contribute to overall meaning (grammatical knowledge);
- use simple transitions to connect ideas (e.g., so, but, after — discourse knowledge);
- pronounce words in a way that is intelligible to proficient speakers of English; use intonation, stress, and pauses to pace speech and contribute to comprehensibility (phonological knowledge); and
- produce speech that is appropriate to the communication goal (pragmatic competence).

Writing

The TOEIC Bridge Writing test measures the ability of beginning and lower-intermediate English language learners to carry out written communication tasks in personal, public, and familiar workplace contexts. This includes the ability to use high-frequency vocabulary and basic grammar structures to produce phrases, sentences, and paragraphs on subjects that are familiar or of personal interest. Test takers can write notes and messages relating to matters of immediate need. They can write simple texts, such as personal letters describing experiences and giving simple opinions.

Communication Goals

In written English, perform simple communication tasks, demonstrating the ability to

- ask for and provide basic information;
- make simple requests, offers, and suggestions, express thanks;

- express a simple opinion and give a reason for it;
- describe people, objects, places, activities; and
- narrate and sequence simple events.

Linguistic Knowledge and Subcompetencies

To achieve these goals, beginning and lower-intermediate English language learners need the ability to

- use high-frequency vocabulary appropriate to a task (lexical knowledge);
- write a sentence using simple word order, such as subject-verb-object, interrogatives, imperatives; use common grammatical structures to contribute to meaning (grammatical knowledge);
- arrange ideas using appropriate connectors (e.g., for example, *in addition, finally*); sequence ideas to facilitate understanding (discourse knowledge);
- control mechanical conventions of English (spelling, punctuation, and capitalization) to facilitate comprehensibility of text (orthographic knowledge); and
- produce text that is appropriate to the communication goal (pragmatic competence).

Suggested citation:

Schmidgall, J., Oliveri, M. E., Duke, T., & Carter Grissom, E. (2019). *Justifying the construct definition for a new language proficiency assessment: The redesigned* TOEIC Bridge® *tests—Framework paper* (Research Report No. RR-19-30). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12267