



## Research Memorandum

ETS RM-23-04

# Design Framework for the TFI™ Test

Shinhye Lee  
Saerhim Oh  
Jonathan Schmidgall

May 2023



# ETS Research Memorandum Series

---

## EIGNOR EXECUTIVE EDITOR

Daniel F. McCaffrey  
*Associate Vice President*

## ASSOCIATE EDITORS

Usama Ali  
*Senior Measurement Scientist*

Beata Beigman Klebanov  
*Principal Research Scientist*

Brent Bridgeman  
*Distinguished Presidential Appointee*

Heather Buzick  
*Senior Research Scientist*

Tim Davey  
*Director Research*

Larry Davis  
*Director Research*

Jamie Mikeska  
*Senior Research Scientist*

Gautam Puhan  
*Director Psychometrics & Data Analysis*

Jonathan Schmidgall  
*Senior Research Scientist*

Jesse Sparks  
*Senior Research Scientist*

Michael Walker  
*Distinguished Presidential Appointee*

Klaus Zechner  
*Senior Research Scientist*

## PRODUCTION EDITORS

Kim Fryer  
*Manager, Editing Services*

Ayleen Gontz  
*Senior Editor*

---

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

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

## Design Framework for the TFI™ Test

Shinhye Lee, Saerhim Oh, and Jonathan Schmidgall  
ETS, Princeton, New Jersey, United States

May 2023

Corresponding author: S. Oh, E-mail: SOH@ets.org

Suggested citation: Lee, S., Oh, S., & Schmidgall, J. (2023). *Design framework for the TFI™ test* (Research Memorandum No. RM-23-04). ETS.

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

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

**Action Editor:** Tim Davey

**Reviewers:** Katerina Duke and Spiros Papageorgiou

Copyright © 2023 by Educational Testing Service. All rights reserved.

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

TFI is a trademark of ETS. All other trademarks are the property of their respective owners.

## **Abstract**

In this paper, we describe key components of the TFI™ test, including how the constructs of listening and reading comprehension are defined and operationalized in the test. The TFI test was developed to facilitate the assessment of French listening and reading comprehension for aspiring and working professionals who are expected to operate within settings where the French language is a major channel of communication. In this paper, we describe the following aspects of the TFI test: (a) the purposes and intended uses for test scores, (b) the target population, (c) the language use domain, (d) the test construct of the TFI test and its theoretical basis, and (e) the operationalization of the construct in test tasks. Overall, we seek to position this paper as a frame of reference that attests to the test’s appropriateness and adequacy as an assessment of French language proficiency.

*Keywords:* TFI™ test, test framework, construct definition, operationalization of the construct

## **Acknowledgments**

The authors would like to thank the reviewers for contributing their expertise and insightful comments, which helped us improve the quality and clarity of this paper.

In the increasingly globalized workplace of the 21st century, language has become a vital tool for facilitating international business and communication. Within this setting, a common language of communication, or *lingua franca*, is often adopted to bring together individuals and professionals from different linguistic and sociocultural backgrounds. The widespread use of English as *lingua franca* notwithstanding (Crystal, 2012), other contact/working languages are used within increasingly diverse workforces and clientele in international settings (Rogerson-Revell, 2007). Consequently, the demand for both aspiring and existing professionals to become equipped with functional language skills other than English is expanding (Ginsburgh *et al.*, 2017; Taillefer, 2007).

The Chauncey Group International, a subsidiary of ETS at the time, developed the TFI™ test in 2000 to address this need, particularly in service of markets and regions where the French language is an integral part of daily life and workplace communication. The TFI test was conceived as a standardized assessment of proficiency of French as a second language (L2) or foreign language (FL). Specifically, the TFI test was designed on the model of the TOEIC® test as a measurement of French-language listening and reading skills relevant to everyday and professional contexts.<sup>1</sup> In 2001, ETS adopted the TFI test and has continued to administer the test worldwide primarily through its representatives in local markets but also through corporations and institutions on an on-demand basis (ETS, 2022b).

The primary purpose of this framework paper is to describe key components of the TFI test, to present the theoretical and empirical basis justifying the construct definition of the test, and to demonstrate how the construct is operationalized in the test by presenting the connection between the test design and the test construct. Overall, we seek to position this paper as a reference that supports claims about the TFI test's appropriateness and adequacy as an assessment of French language proficiency.

In this document, we describe the following aspects of the TFI test: (a) the test purpose and the intended uses of test scores, (b) the target population, (c) the language use domain, (d) the construct definition, and (e) the operationalization of the construct.

### **Test Purpose and Intended Uses of Test Scores**

The TFI test is designed to measure the ability of French language learners to comprehend spoken and written communication in French relevant to everyday and workplace environments. Test scores are thus intended to provide an indication of French language proficiency, particularly the knowledge, skills, and abilities in French that are essential in these language-use settings.

With the test focusing on communication abilities relevant to domains spanning daily life and general workplace settings, scores on the TFI test can be used to support the decision-making needs of score users. For score users operating within workplace contexts where French is a required job skill (e.g., companies, government agencies), test results are intended to be used to inform various employment decisions (e.g., selection, placement, promotion). For educational authorities and institutional score users (e.g., French language programs, training institutes), test results are intended to be used to measure students' level of French proficiency and to monitor progress over time. Finally, individual test takers can use the TFI test scores to verify their French proficiency level and use this information to make decisions related to advancing their professional careers.

### **Target Population**

The TFI test is designed to be suitable for L2 or FL speakers of French who wish to demonstrate the ability to understand spoken and written French for communication purposes. In particular, the target test-taking population includes those whose first language (L1) is not French and who are from varying educational and vocational backgrounds including (a) language learners in French-language training programs, (b) individuals aspiring to work in organizations where French is used, and (c) professionals who are required to use French in real-life work settings. Though the test is offered worldwide, test takers of the TFI test are typically expected to reside in francophone countries (e.g., Canada, France, and several French-speaking countries in Northern Africa), where the French language is regularly used in the everyday and global workplace environment (ETS, 2022b).



As a standardized assessment of French language proficiency, the TFI test is intended to measure the abilities of test takers across a wide range of proficiency levels. Specifically, scores on the TFI test can be interpreted in relation to the proficiency levels described in the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001), as established by Baron and Tannenbaum (2010). Using a standard-setting procedure<sup>2</sup> (Cizek & Bunch, 2007), Baron and Tannenbaum mapped TFI test scores to CEFR proficiency levels A2 (elementary) through C1 (advanced), thereby underscoring the broad range of proficiency assessed by the test.

### **Target Language Use Domain**

The contexts of language use targeted by the TFI test range from semiformal settings (e.g., dining, traveling) to formal settings (e.g., meetings, presentations) that are reflective of communication commonly occurring in the international workplace (see the Test Content section of the *TFI Test User Guide* for more examples of relevant settings; [ETS, 2022b]). The context of language use targeted by the test, or target language use (TLU) domain (Bachman & Palmer, 2010), includes general exchanges involving L1 and L2/FL speakers of French as professional correspondents and interactants (ETS, 2022b). The TLU domain of the TFI test does not include academic subject matter (e.g., French literature) or interpersonal relationships (e.g., family) and does not exclusively reflect contexts and customs specific to a francophone country (e.g., France). In addition, given the general-purpose focus of the test, TFI test content is not intended to be profession-specific; that is, test content is modeled on general language use situations and tasks where highly specialized knowledge of a given occupation (e.g., healthcare) is not a necessity (ETS, 2022b).

To a large extent, the TLU domain targeted by the TFI test corresponds with the domains of language use specified in the CEFR (Council of Europe, 2009), particularly the *occupational* and *public domains*. The occupational domain refers to the general activities related to one's professional practice, which is relevant to the TFI test's focus on workplace contexts. The public domain emphasizes activities that occur as part of everyday business and

public-service transactions, thereby broadly corresponding to the general-purpose, profession-neutral focus of the test.

### **Construct Definition**

Justification of the design and development of assessment tasks are guided by sound articulation of a test construct and its core aspects that are to be represented (American Educational Research Association [AERA] *et al.*, 2014). A clear articulation of the test construct in test development is critical because the construct definition provides a basis for intended use of the test and the interpretation of the test scores. The construct definition reflects the interpretation of test scores, and the theoretical basis for the construct definition should be elaborated by the test developer.

With respect to L2/FL testing, assessment researchers have approached and defined the language ability construct via three perspectives (Bachman, 2007): (a) as a trait, or “the knowledge and fundamental processes of the test taker” (Chapelle, 1998, p. 34); (b) as a behavior, or “what the test takers can do, and under what circumstances they can do it” (Buck, 2001, p. 110); or (c) as some combination of trait and behavior. It is this third and integrative approach to construct definition, also referred to as the *interactionalist approach* (Bachman, 2007; Chapelle, 1998), that is relevant to how the TFI test construct is conceptualized.

Proponents of the interactionalist perspective assert that language performance is a result of the interaction between factors internal and external to the learner (Ellis, 1989); that is, performance not only reflects one’s underlying language ability but is also influenced by the context in which it occurs. A third component that mediates this interaction is one’s strategic competence (Bachman, 1990), which enables a language user to put their linguistic knowledge and skills to use in a communicative context. Thus, in defining a given test construct, the interactionalist approach takes the view that language knowledge—and the strategic competence needed to execute the use of this knowledge—should be specified more precisely in reference to the TLU domain (Chapelle, 1998). In practice, this process involves taking stock of three essential components, namely, “the language knowledge required to facilitate

performance, strategies to support performance, and a description of the performance context itself” (Schmidgall et al., 2019, p. 5). In the case of the TFI test, the test intends to measure the ability to comprehend spoken and written French, particularly in the contexts of everyday and international workplace settings (i.e., the TLU domain of the TFI test). The TFI test construct can be further elaborated as including both selected foundational language ability knowledge (e.g., vocabulary knowledge) and relevant strategies and communication goals (e.g., making inferences) that constitute the language demands representative of tasks common to the TLU domain.

To provide a comprehensive basis for interpreting the TFI test construct, we outline how the key language abilities of interest—listening and reading comprehension—have been theoretically conceptualized by applied linguistics and language assessment researchers. In so doing, we highlight major components—linguistic, strategic, and contextual factors in particular—that are hypothesized and potentially known predictors of L2/FL listening and reading competence and thus provide support for the intended meaning of TFI test scores. This review then forms the basis of the final section in which we present how the highlighted components are operationalized in each test section of the TFI test.

### **Literature on Listening Comprehension**

Listening has been conceptualized in the literature as primarily a cognitive process, one that involves a complex interplay of various types of mental activities (Brindley, 1998; Buck, 2001). Largely based upon principles derived from cognitive psychology and L1 listening comprehension research (e.g., Chaudron & Richards, 1986), early views explain the listening process in terms of two approaches: the *bottom-up* and *top-down* approaches. The bottom-up approach assumes that the starting point of comprehension is the lower level decoding of the spoken input (e.g., sounds, words), whereas the top-down approach specifies the starting point to be the use of other higher order information (e.g., background knowledge, context). Implicit in these approaches is the notion of listening as a unidirectional process (Batty, 2021) whereby

successful comprehension results in executing operations in a sequential manner to construct meaning (Bejar *et al.*, 2000).

Recent views subscribing to a more integrative perspective of the listening process assume the complementary and parallel nature of multiple processes and knowledge sources, as well as their convergence (Flowerdew & Miller, 2010). For instance, Rost (2005) projected four types of processing categories—neurological, linguistic, semantic, and pragmatic processing—onto the continuum of bottom-up and top-down extremes to propose that listeners employ all types of processes simultaneously and continuously to make sense of the spoken language. Vandergrift and Goh (2012) noted the use of different knowledge sources and further emphasized the role of listener strategies to regulate the various cognitive activities and their interactions. Buck (2001) presented the integrative view to describe the listening process as “listeners using whatever information they have available, or whatever information seems relevant to help them interpret what the speaker is saying at the time” (p. 3). Collectively, the integrative view forms the basis of a component skills approach to defining a listening construct (Bachman, 1990), which models listening comprehension as a collection of linguistic and strategic subskills (e.g., Buck, 2001; Weir, 2005).

Researchers have attempted to explain the impact of various linguistic subskills and strategies used to enhance success in listening. The subskills are conceptualized as components of language ability (e.g., vocabulary, grammar, discourse, pragmatic, sociolinguistic knowledge), which are typically explored alongside other learner-internal variables (e.g., working memory, background knowledge). In general, a fairly robust finding in relevant L2/FL listening comprehension research is that linguistic knowledge—vocabulary and, to a lesser extent, grammar—is highly predictive of listening comprehension (Zhang & Zhang, 2022). This finding appears to be primarily observed in studies involving learners of English as a foreign language (e.g., Cheng & Matthews, 2018; Révész & Brunfaut, 2013) but also supported by research focused on learners of Indo-European languages such as Dutch (Andringa *et al.*, 2012), Spanish (Mecartty, 2000), and French (Vandergrift & Baker, 2015, 2018; Vandergrift *et al.*, 2006).

Research conducted by Vandergrift and colleagues suggested that vocabulary knowledge in French is the primary determinant of success in comprehending spoken French.

In regard to listening comprehension strategies, numerous taxonomies have been theorized to describe the mental activities listeners employ when attributing meaning to spoken language (Vandergrift & Goh, 2012). Of particular relevance to the L2/FL listening assessment context are two types of strategy use: cognitive strategies, the mental processes related to comprehending, storing, and linguistic knowledge (e.g., making inferences, making predictions), and metacognitive strategies, the self-management activities of listeners to plan for, monitor, and evaluate one's listening (e.g., planning for a listening task, identifying listening problems; Purpura, 1997, 1998). Theoretically, cognitive and metacognitive strategies are viewed as parallel and separate (O'Malley & Chamot, 1990; Oxford, 1999). Researchers, however, have noted the difficulty of operationalizing and delineating the two processes (Chapelle *et al.*, 1997; Vandergrift *et al.*, 2006), especially "when they are embedded in complex sequences of behavior or hierarchies of decisions" (Paris *et al.*, 1991, p. 610). In modeling the listening construct, Buck (2001) specified the use of both sets of strategies as foundational components of strategic competence (Bachman, 1990), emphasizing the orchestration of the two in maximizing comprehension in a testing context. In a study of the impact of strategy use on IELTS listening scores, Kök (2018) corroborated Buck's view by demonstrating a positive correlation between increased strategy use and greater listening proficiency, with synergized usage of both sets of strategies contributing to higher listening comprehension scores.

Reflecting the importance of task characteristics in assessment (Bachman & Palmer, 2010), a wide range of text- and task-specific parameters have also received much attention in L2/FL listening assessment literature. The following three aspects of task characteristics have been most frequently cited to explain how test takers process and perform on listening tasks (see He & Jiang, 2020, for a comprehensive review): (a) modality and channel of task presentation (e.g., use of videos as test material; Batty, 2015; Suvorov, 2015; Wagner, 2010); (b) types and varieties of spoken texts (e.g., monologic or dialogic texts and scripted or

unscripted texts; Papageorgiou *et al.*, 2012; Wagner & Wagner, 2016); and (c) accents of speakers (e.g., diversity and intelligibility of speaker accents; Kang *et al.*, 2019; Ockey & French, 2016). To date, research on these issues, and in particular on the actual impact of these parameters on listening comprehension scores, is inconclusive at best and highlights the numerous challenges to constructing a sound measurement of listening proficiency (Brunfaut, 2016). Yet, as Wagner (2014) aptly summarized, the task settings and conditions noted in empirical research are useful insofar as they tap into what constitutes authentic listening tasks representative of the TLU domain, which then forms a useful basis for the design and development of relevant test tasks to follow (Bachman & Palmer, 2010).

### **Literature on Reading Comprehension**

Reading comprehension is considered in the review literature as a complex and multifaceted process of extracting and constructing meaning from written text (Grabe, 2009). Various theoretical models have emerged to elucidate what reading comprehension entails and how to elicit evidence of that comprehension. Two contrasting views have dominated the field of reading research in general and reading assessment practices in particular, namely, the product- and process-oriented approaches to reading (Johnston, 1984). Whereas the product model puts an emphasis on *what* understanding of the text was reached, the process model concerns *how* that understanding was reached.

By their nature, reading assessments are product-oriented (Koda, 2013), with test questions primarily designed to sample different reading abilities and levels of understanding of text information (e.g., understanding details vs. main idea of a text). Typically, accuracy in reading comprehension (as demonstrated in test scores) is what is thought to provide evidence of successful reading comprehension (Brunfaut, 2022). Yet, “variation in the product” is a given (Alderson, 2000, p. 5); not all readers reach uniform and accurate understandings of the same reading passage (Grabe, 2009). This variation in outcomes (i.e., comprehension) is primarily due to language-ability differences but is also attributable to how readers interact with the text—that is, the process by which they arrive at successfully comprehending a given text (Hoey,

2001; Khalifa & Weir, 2009). This means that the kinds of processing activities, or more precisely, the strategies that readers employ to facilitate and monitor their reading, may influence how well readers understand the text (Cohen & Upton, 2006; Purpura, 1998). Contextual factors may also come into play, as different types of reading tasks impose varying degrees of real-time processing demands on readers (Koda, 2013; Skehan, 1998). Therefore, text- and task-specific characteristics may also need to be accounted for in order to evaluate the degree of successful reading comprehension (Enright *et al.*, 2000). When it comes to conceptualizing a reading construct, researchers thus advocate an integrative point of view that considers both product and process models of reading (Hubley, 2012), as well as the collective impact that linguistic and nonlinguistic components have on comprehension (Alderson, 2005).

In keeping with this integrative view, determining the key variables that affect the product and process of reading (Bernhardt, 2010) has been a frequent point of interest in L2/FL reading literature. Research-driven meta-analyses have proven to be useful for synthesizing the relative contributions that each of these variables make in explaining variance in reading comprehension performance (Zhang & Zhang, 2022). In a meta-analysis of 58 empirical studies, Jeon and Yamashita (2014) identified high-evidence and low-evidence variables (*i.e.*, the variables most and least frequently researched). High-evidence variables included language-knowledge attributes (*e.g.*, grammar and vocabulary knowledge). Low-evidence variables included language-general variables (*e.g.*, working memory, metacognition). The researchers compared the size of the correlation coefficients reported for each of these variables with reading comprehension measures. Findings indicated that variables in the high-evidence group, particularly grammar and vocabulary knowledge, were invariably the stronger correlates with L2/FL reading comprehension. A few language-general variables, on the other hand, demonstrated significant yet moderate correlations, with metacognition being the weakest correlate of all. Notably, as a comprehensive meta-analysis, a wide range of learner population and target languages were represented in the study; French learners, for instance, were the second-most referenced group of learners in the primary research after English learners. Thus,

the study findings underscore the pivotal role that language knowledge plays in particular and appear to encompass various reading/learning contexts including those relevant to French as an L2 or FL (e.g., Brisbois, 1995; Fecteau, 1999; Lefrançois & Armand, 2003; Morrison, 2004; Noonan et al., 1997).

Although research has suggested that linguistic subskills (e.g., vocabulary and grammatical knowledge) are a critical component of reading comprehension, there has been a growing interest in evaluating the role of strategic competence in L2/FL reading comprehension (Carrell & Grabe, 2002; Enright et al., 2000). This has led to a focus on readers' use of metacognitive and cognitive strategies in reading assessment research (Cohen & Upton, 2006; Purpura, 1999). In general, researchers have echoed the "unitary" view of strategy use as put forth by listening assessment research (Buck, 2001; Dalman & Plonsky, 2022), emphasizing how metacognitive and cognitive strategies are closely related and even "form a continuum" (Phakiti, 2003, p. 44). Using a structural equation modeling (SEM) approach, Zhang et al. (2014) empirically tested this hypothesized relationship by constructing strategy-use models based upon the performance and reported strategy use of test takers of the Chinese College English Test (CET) Band 4. The multiple SEM analyses indicated that the unitary model (i.e., metacognitive and cognitive strategies working in synergy) functioned as the best-fitting model over the hierarchical model (i.e., metacognitive strategies operating as higher order factors) and correlated model (i.e., metacognitive and cognitive strategies having a correlated relationship) for explaining performance on the reading test. Based upon the study findings, Zhang et al. further noted the central role that simultaneous use of strategies plays in maximizing reading test performance, corroborating the findings of Kök (2018). Although more research is needed to investigate examinee strategy uses in real world reading contexts, under high-stakes testing circumstances where a wide range of sources of information and test-task demands are presented (as in the case of the CET testing context), it appears to be useful for test takers to make use of multiple strategies in concert.



Researchers have also delineated a range of contextual factors associating with L2/FL reading comprehension. For instance, Alderson (2005) offered a comprehensive review of theoretically acclaimed text characteristics, which widely range from minute linguistic features of a text (e.g., typography) to types and genre of reading passages. Enright et al. (2000) similarly categorized key reading components into largely reader-internal and external factors, of which the latter points to the effects of types of text, task, and topic in relation to achieving a specific reading purpose. Recently, Brunfaut (2022) categorized relevant research endeavors into the following three areas of inquiries: (a) characteristics of the reading texts (e.g., topic, genre, and length of texts; Freedle & Kostin, 1993); (b) characteristics of the test tasks (e.g., types and varieties of response format; In'nami & Koizumi, 2009; Löwenadler, 2019); and (c) the interactions between these two parameters (Freedle & Kostin, 1993; McCray, 2014).

### **Summary**

Although the above depiction of listening and reading comprehension is not an exhaustive one, it encompasses many of the more prominent factors noted in empirical research that are theorized or hypothesized to influence listening and reading comprehension. Ultimately, the collective theme that cuts across this narrative review is that listening and reading comprehension entail the interplay of the various learner-internal and external factors, which broadly tap into a learner's foundational linguistic knowledge, cognitive and metacognitive capacities, and the context in which comprehension occurs. It then follows that test developers need to sample varying texts, tasks, and topics to cover key influential components, while bearing in mind the purpose, learner population, and the TLU domain targeted by a given assessment (Bachman & Palmer, 2010).

In light of the review of theory, in the following section we describe the TFI test sections in relation to how the empirically attested components of listening and reading comprehension are operationalized into test content and items.

## Operationalization of the Construct

In this section, the overall structure of the TFI test is described followed by an explanation of how the underlying construct of the TFI test is operationalized in each test section of the TFI test and a description of the TFI test scores.

### Structure of the TFI Test

The TFI test consists of two main parts: Listening Comprehension and Reading Comprehension. These two parts each include three sections and consist of 90 multiple-choice items. Table 1 presents the overall structure of the TFI test.

**Table 1. Structure of the TFI Test**

Part	Number of items	Estimated duration	Section
Part 1: Listening Comprehension	90	42 minutes	Section I: Question-Response Section II: Conversations Section III: Talks
Part 2: Reading Comprehension	90	68 minutes	Section IV: Error Recognition Section V: Incomplete Sentences Section VI: Reading Comprehension

In the Listening Comprehension part of the test, test takers listen to questions, short dialogues, conversations, and talks, and they then answer questions based on what they hear. In the Reading Comprehension part of the test, test takers are presented with written sentences and texts, and they identify errors and complete sentences and answer questions based on what they read.

### *TFI Test Sections*

As mentioned earlier, the construct targeted by the TFI test includes the linguistic subskills underlying French listening and reading comprehension as well as relevant strategies and, in some cases, higher order communication goals relevant to completing tasks representative of the TLU situations.

Sections I (Question-Response) and II (Conversations) of the Listening Comprehension part of the test comprise items intended to assess the degree to which test takers understand short utterances and conversations. Section III (Talks) consists of extended pieces of spoken texts (e.g., oral presentations), each followed by two to four comprehension questions. The listening texts utilized in Section III typically contain a range of registers and formality and are more varied in content domains than those presented in the preceding sections. The questions in this section target three types of strategic purposes for understanding a spoken/written text: understanding specific details/information, understanding the main idea, and understanding the implied meaning. Depending on the question, test takers may need to employ different sets of cognitive and metacognitive skills to effectively demonstrate comprehension. For instance, for items asking for the extraction of details, test takers would need to prioritize focusing on a specific part of the text (rather than generally trying to understand everything) in order to identify matching information (Buck, 2001). Items asking for the gist/main idea tap into the overall/global meaning of a given text; therefore, test takers would benefit from synthesizing information that is spread over a number of statements and then understanding how each of the statements function in relation to each other to make the focal point (Alderson, 2005). Finally, to understand implied meaning, test takers may need to bring together what is explicitly stated and not stated in the text and make appropriate inferences about what a specific part of the spoken/written text means (Alderson, 2005).

Figures 1 to 3 present examples from Section I (Question-Response), Section II (Conversations), and Section III (Talks). See ETS (2022a) for more examples.

**Figure 1. Example From Section I (Question-Response)**

Vous entendez:	À quelle heure la prochaine réunion aura-t-elle lieu?
Vous entendez ensuite:	(A) Elle viendra jeudi. (B) C'est un lieu charmant. (C) Juste après la pause.

**Figure 2. Example From Section II (Conversations)**

Vous entendez: (HA) Hôtel Saint Germain, bonjour.  
 (FA) Bonjour. Je voudrais réserver une chambre pour deux avec un grand lit et salle de bains.  
 (HA) Oui, Madame, pour quel jour?

La question et les quatre choix sont:

Que désire la cliente?

(A) Réserver une chambre simple.  
 (B) Réserver une chambre double.  
 (C) Réserver une suite.  
 (D) Réserver deux chambres simples.

**Figure 3. Example From Section III (Talks)**

Vous entendez: **Les questions 1 à 3** portent sur cette publicité.

Quelle que soit la passion qui vous guide, le Québec possède ce qu'il faut pour satisfaire tous les goûts: activités culturelles, romantisme et découvertes historiques, proximité de la nature et joies du plein air. Tout y est! Vous pouvez être certain de passer des vacances inoubliables. Venez vous dépaysier et vous amuser chez nous.

Les questions et les quatre choix sont:

1. À quoi cette publicité sur le Québec vous invite-t-elle?

- (A) À vous y installer.
- (B) À y trouver l'âme soeur.
- (C) À y prendre des vacances.
- (D) À y pratiquer des sports.

2. Selon cette publicité, qu'est-on certain de pouvoir faire au Québec?

- (A) Bien manger.
- (B) Visiter des parcs d'amusement.
- (C) Faire des rencontres romantiques.
- (D) Profiter du grand air.

3. Qui pourrait être l'auteur de ce commentaire?

- (A) Un office du tourisme.
- (B) Une association culturelle.
- (C) Une agence matrimoniale.
- (D) Une organisation pour l'emploi.

In terms of reading comprehension, both Sections IV (Error Recognition) and V (Incomplete Sentences) are intended to measure the ability to recognize language forms in context. In the Error Recognition section, items are presented in sentences with four words or phrases underlined, with the answer key being the one option that contains a specific

grammatical error. The errors cover a range of linguistic knowledge needed to comprehend written French and include agreement, confusion, incorrect word or word form, omission, and inclusion. The Incomplete Sentences section of the test is composed of gap-filling questions, requiring test takers to identify either an appropriate lexical unit or proper grammar structure to complete a given sentence. In both of these reading sections, test takers make selections of the correct language forms by considering the contexts provided in the prompt sentences. The contexts depicted in these sentences are related to daily life and the workplace, representative of semiformal spoken and written French. Section VI (Reading Comprehension) consists of extended written texts (e.g., prose passages) each followed by two to four comprehension questions. Similar to Section III (Talks) for Listening Comprehension, the questions in this section target three types of strategic purposes for understanding a written text: understanding specific details/information, understanding the main idea, and understanding the implied meaning. Figures 4 to 6 present examples from Section IV (Error Recognition), Section V (Incomplete Sentences), and Section VI (Reading Comprehension). See ETS (2022a) for more examples.

#### Figure 4. Example From Section IV (Error Recognition)

Les joueurs de football est de grands athlètes.  
A            B            C            D

#### Figure 5. Example From Section V (Incomplete Sentences)

Quand il fait froid, il faut mettre -----.  
(A) des allumettes  
(B) un manteau  
(C) une pelle  
(D) du feu

**Figure 6. Example From Section VI (Reading Comprehension)**

Les questions 1 et 2 portent sur cette nouvelle.

Lisez ce texte: C'est le grand match en coulisses "de la Coupe du monde que tous les économistes attendent. La société Sporton, leader planétaire de l'équipement sportif, pourra-t-il détrôner son rival commercial, Kiufu qui règne, lui, sur le marché du foot? D'ici au prochain Mondial, nous dominerons le domaine du ballon rond, affirme un responsable de la firme américaine qui a frappé un grand coup en enrôlant une des meilleures équipes latino-américaines pour un contrat astronomique d'un milliard de francs sur dix ans.

1. À qui s'adresse essentiellement cette nouvelle?
  - (A) Aux joueurs de football
  - (B) Aux amateurs de sport
  - (C) Aux hommes d'affaires
  - (D) Aux astronomes
  
2. En mettant une grande équipe sous contrat, de quoi s'est assuré Sporton?
  - (A) De gagner le Mondial
  - (B) De prendre contrôle du marché du football
  - (C) De dominer le sport planétaire
  - (D) De gagner un milliard de francs

Finally, the TFI test content and items are designed in a way that reflects French as it is used in everyday and workplace language use situations. The listening part samples varieties of spoken texts by adapting authentic sources of spoken texts (e.g., conversational and extended monologic speech), interactant relationships, and accents of international French (e.g., L1 speakers of French from France, as well as speakers from Canada), reflecting the various demands and contextual factors relevant to listening tasks carried out in workplace contexts (Taillefer, 2007). Similarly, the reading part items are adapted from authentic sources of texts (e.g., news articles, charts, advertisements, workplace correspondence) to represent a variety of real-life topics (e.g., personnel, general office affairs) and text types (e.g., telegraphic, prose passage) that professionals are likely to encounter in the global workplace.

### ***TFI Test Scores***

TFI test scores are determined by the number of correct answers on the Listening and Reading Comprehension parts of the test. The number of correct answers within each part is converted to a scaled score ranging from 5 to 495. As briefly described in the Target Population section of the current paper, TFI test scores are mapped onto CEFR proficiency levels A2 (elementary) to C1 (advanced; Baron & Tannenbaum, 2010). Table 2 presents the minimum scaled score recommendations for each CEFR level.

**Table 2. TFI Test Scores and CEFR Levels**

CEFR levels	Listening	Reading
Below A2	5–80	5–100
A2	85–155	105–180
B1	160–295	185–300
B2	300–390	305–425
C1	395–495	430–495

*Note.* The TOEIC Listening and Reading test sections also report scaled scores using a range from 5 to 495, but these scores are interpreted differently and should not be considered interchangeable with TFI test scores in terms of language proficiency. CEFR = Common European Framework of Reference for Languages.

### **Conclusion**

This paper presented the theoretical and empirical foundations supporting the construct definition and the design of the TFI test, which is intended to measure the ability of French language learners' listening and reading comprehension relevant to everyday and workplace settings. The underlying construct of the TFI test incorporates both foundational knowledge of French as it applies to listening and reading comprehension, relevant strategies to support performance, and in some cases, higher order communication goals relevant to completing tasks representative of the TLU domain. A description of the test sections was provided to demonstrate how this construct is operationalized in the TFI test. Overall, this paper is expected to provide a reference that supports claims about the appropriateness of the TFI test as an assessment of French language proficiency and to provide reference for interpretations and justification of the TFI score use.

## References

- Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511732935>
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. Continuum.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning*, 62(Suppl. 2), 49–78. <https://doi.org/10.1111/j.1467-9922.2012.00706.x>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. 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. E. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 41–72). University of Ottawa Press. <https://doi.org/10.2307/j.ctt1ckpccf.9>
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
- Baron, P. A., & Tannenbaum, R. J. (2010). *Mapping the Test de Français international onto the Common European Framework of Reference* (Research Memorandum No. RM-10-12). ETS. <https://www.ets.org/Media/Research/pdf/RM-10-12.pdf>
- Batty, A. O. (2015). A comparison of video- and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing*, 32(1), 3–20. <https://doi.org/10.1177/0265532214531254>
- Batty, A. O. (2021). Measuring L2 listening. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 275–284). Routledge. <https://doi.org/10.4324/9781351034784-30>



- Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL® 2000 listening framework: A working paper* (TOEFL Monograph Series No. MS-19). ETS.  
<https://www.ets.org/Media/Research/pdf/RM-00-07.pdf>
- Bernhardt, E. B. (2010). *Understanding advanced second-language reading*. Routledge.  
<https://doi.org/10.4324/9780203852408>
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171–191. <https://doi.org/10.1017/S0267190500003536>
- Brisbois, J. E. (1995). Connections between first- and second-language reading. *Journal of Literacy Research*, 27(4), 565–584. <https://doi.org/10.1080/10862969509547899>
- Brunfaut, T. (2016). Assessing listening. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 97–112). Mouton de Gruyter.  
<https://doi.org/10.1515/9781614513827-009>
- Brunfaut, T. (2022). Assessing reading. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed., pp. 254–268). Routledge.  
<https://doi.org/10.4324/9781003220756-20>
- Buck, G. (2001). *Assessing listening*. Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511732959>
- Carrell, P. L., & Grabe, W. (2002). Reading. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 233–250). Arnold.
- Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). Cambridge University Press.  
<https://doi.org/10.1017/CBO9781139524711.004>
- Chapelle, C. A., Grabe, W., & Berns, M. (1997). *Communicative language proficiency: Definition and implications for TOEFL 2000* (TOEFL Monograph Series No. MS-10). ETS.  
<https://www.ets.org/Media/Research/pdf/RM-97-03.pdf>

- Chaudron, C., & Richards, J. C. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics*, 7(2), 113–127. <https://doi.org/10.1093/applin/7.2.113>
- Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, 35(1), 3–25. <http://doi.org/10.1177/0265532216676851>
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. SAGE Publications.
- Cohen, A. D., & Upton, T. A. (2006). *Strategies in responding to the New TOEFL reading tasks* (TOEFL Monograph Series No. MS-33). ETS. <https://doi.org/10.1002/j.2333-8504.2006.tb02012.x>
- Council of Europe. (2001). *The Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). A Manual*.
- Crystal, D. (2012). *English as a global Language*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139196970>
- Dalman, M., & Plonsky, L. (2022). The effectiveness of second-language listening strategy instruction: A meta-analysis. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688211072981>
- Ellis, R. (1989). *Understanding second language acquisition*. Oxford University Press.
- Enright, M. K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. A. (2000). *TOEFL 2000 Reading Framework: A working paper* (TOEFL Monograph Series No. MS-17). ETS. <https://www.ets.org/Media/Research/pdf/RM-00-04.pdf>
- ETS. (2022a). *TFI test examinee handbook*. <https://www.ets.org/pdfs/tfi/tfi-examinee-handbook.pdf>
- ETS. (2022b). *TFI test user guide*. <https://www.ets.org/pdfs/tfi/tfi-user-guide.pdf>

- Fecteau, M. L. (1999). First- and second-language reading comprehension of literary texts. *The Modern Language Journal*, 83(4), 475–493. <https://doi.org/10.1111/0026-7902.00036>
- Flowerdew, J., & Miller, L. (2010). Listening in a second language. In A. D. Wolvin (Ed.), *Listening and human communication in the 21st century* (pp. 158–177). Wiley Blackwell. <https://doi.org/10.1002/9781444314908.ch7>
- Freedle, R., & Kostin, I. (1993). *The prediction of TOEFL reading comprehension item difficulty for expository prose passages for three item types: Main idea, inference, and supporting idea items* (TOEFL Research Report No. 44). ETS. <https://doi.org/10.1002/j.2333-8504.1993.tb01524.x>
- Ginsburgh, V., Melitz, J., & Toubal, F. (2017). Foreign language learning and trade. *Review of International Economics*, 25(2), 320–361. <https://doi.org/10.1111/roie.12268>
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139150484>
- He, L., & Jiang, Z. (2020). Assessing second language listening over the past twenty years: A review within the socio-cognitive framework. *Frontiers in Psychology*, 11, Article 2123. <https://doi.org/10.3389/fpsyg.2020.02123>
- Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis*. Routledge.
- Hubley, N. J. (2012). Assessing reading. In C. Coombe, P. Davidson, B. O’Sullivan, & S. Stoyloff (Eds.), *The Cambridge guide to second language assessment* (pp. 211–217). Cambridge University Press.
- In’nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, 26(2), 219–244. <https://doi.org/10.1177/0265532208101006>
- 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>
- Johnston, P. H. (1984). Assessment in reading. In P. D. Pearson, R. Barr, M. L. Kamil, & P. B. Mosenthal (Eds.), *Handbook of reading research* (Vol. 1, pp. 147–182). Longman.

- Kang, O., Thomson, R., & Moran, M. (2019). The effects of international accents and shared first language on listening comprehension tests. *TESOL Quarterly*, *53*(1), 56–81.  
<https://doi.org/10.1002/tesq.463>
- Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge University Press.
- Koda, K. (2013). Assessment of reading. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley-Blackwell. <https://doi.org/10.1002/9781405198431.wbeal0051>
- Kök, İ. (2018). Relationship between listening comprehension strategy use and listening comprehension proficiency. *International Journal of Listening*, *32*(3), 163–179.  
<https://doi.org/10.1080/10904018.2016.1276457>
- Lefrançois, P., & Armand, F. (2003). The role of phonological and syntactic awareness in second-language reading: The case of Spanish-speaking learners of French. *Reading and Writing: An Interdisciplinary Journal*, *16*(3), 219–246.  
<https://link.springer.com/content/pdf/10.1023/a:1022874425314.pdf>
- Löwenadler, J. (2019). Patterns of variation in the interplay of language ability and general reading comprehension ability in L2 reading. *Language Testing*, *36*(3), 369–390.  
<https://doi.org/10.1177/0265532219826379>
- McCray, G. (2014). *Statistical modelling of cognitive processing in reading comprehension in the context of language testing* [Unpublished doctoral dissertation]. Lancaster University.
- Mecartty, F. H. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning*, *11*(2), 323–348.
- Morrison, L. (2004). Comprehension monitoring in first and second language reading. *The Canadian Modern Language Review*, *61*(1), 77–107.  
<https://doi.org/10.1353/cml.2004.0005>

- Noonan, B., Colleaux, J., & Yackulic, R. A. (1997). Two approaches to beginning reading in early French immersion. *The Canadian Modern Language Review*, 53(4), 729–742.  
<https://doi.org/10.3138/cmlr.53.4.729>
- Ockey, G. J., & French, R. (2016). From one to multiple accents on a test of L2 listening comprehension. *Applied Linguistics*, 37(5), 693–715.  
<https://doi.org/10.1093/applin/amu060>
- O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524490>
- Oxford, R. L. (1999). Relationships between second language learning strategies and language proficiency in the context of learner autonomy and self-regulation. *Revista Canaria de Estudios Ingleses*, 38(1), 108–126.
- Papageorgiou, S., Stevens, R., & Goodwin, S. (2012). The relative difficulty of dialogic and monologic input in a second-language listening comprehension test. *Language Assessment Quarterly*, 9(4), 375–397. <https://doi.org/10.1080/15434303.2012.721425>
- Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp., 609–634). Erlbaum.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26–56.  
<https://doi.org/10.1191/0265532203lt243oa>
- Purpura, J. E. (1997). An analysis of the relationship between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, 47(2), 289–325. <https://doi.org/10.1111/0023-8333.91997009>
- Purpura, J. E. (1998). Investigating the effects of strategy use and second language test performance with high- and low-ability test takers: A structural equation modelling approach. *Language Testing*, 15(3), 333–379.  
<https://doi.org/10.1177/026553229801500303>

- Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A SEM approach*. Cambridge University Press.
- Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, 35(1), 31–65. <https://doi.org/10.1017/S0272263112000678>
- Rogerson-Revell, P. (2007). Using English for international business: A European case study. *English for Specific Purposes*, 26(1), 103–120. <https://doi.org/10.1016/j.esp.2005.12.004>
- Rost, M. (2005). L2 listening. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 527–552). Routledge. <https://doi.org/10.4324/9781410612700-36>
- Schedl, M. (2010). Background and goals of the TOEIC® Listening and Reading test redesign project. In D. Powers (Ed.), *The research foundation for the TOEIC tests: A compendium of studies* (pp. 2.1–2.18). ETS. <https://www.ets.org/Media/Research/pdf/TC-10-02.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>
- Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*, 18, 268–286. <https://doi.org/10.1017/S0267190500003585>
- Suvorov, R. (2015). The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos. *Language Testing*, 32(4), 463–483. <https://doi.org/10.1177/0265532214562099>
- Taillefer, G. F. (2007). The professional language needs of economics graduates: Assessment and perspectives in the French context. *English for Specific Purposes*, 26(2), 135–155. <https://doi.org/10.1016/j.esp.2006.06.003>

- Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, *65*(2), 390–416. <https://doi.org/10.1111/lang.12105>
- Vandergrift, L., & Baker, S. C. (2018). Learner variables important for success in L2 listening comprehension in French immersion classrooms. *The Canadian Modern Language Review*, *74*(1), 79–100. <https://doi.org/10.3138/cmlr.3906>
- Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.
- Vandergrift, L., Goh, C. C. M., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning*, *56*(3), 431–462. <https://doi.org/10.1111/j.1467-9922.2006.00373.x>
- Wagner, E. (2010). The effect of the use of video texts on ESL listening test-taker performance. *Language Testing*, *27*(4), 493–513. <https://doi.org/10.1177/0265532209355668>
- Wagner, E. (2014). Assessing listening. In A. J. Kunnan (Ed.), *The companion to language assessment: Vol. 1. Abilities, contexts, and learners* (pp. 47–63). John Wiley & Sons. <https://doi.org/10.1002/9781118411360.wbcla094>
- Wagner, E., & Wagner, S. (2016). Scripted and unscripted spoken texts used in listening tasks on high stakes tests in China, Japan, and Taiwan. In V. Aryadoust & J. Fox (Eds.), *Current trends in language testing in the Pacific Rim and the Middle East: Policies, analyses, and diagnoses* (pp. 438–463). Cambridge Scholars Publishing.
- Weir, C. (2005). *Language testing and validation: An evidence-based approach*. Palgrave Macmillan. <https://link.springer.com/book/10.1057/9780230514577>
- Zhang, L., Goh, C. C. M., & Kunnan, A. J. (2014). Analysis of test takers' metacognitive and cognitive strategy use and EFL reading test performance: A multi-sample SEM approach. *Language Assessment Quarterly*, *11*(1), 76–102. <https://doi.org/10.1080/15434303.2013.853770>

Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, 26(4), 696–725. <https://doi.org/10.1177/1362168820913998>



### Notes

- <sup>1</sup> The TFI test was modeled after the TOEIC Listening and Reading test before the latter was redesigned. For a description of the TOEIC Listening and Reading test redesign project, see Schedl (2010).
- <sup>2</sup> Standard setting refers to the process by which a panel of domain experts provide recommendations of minimum scores (cut scores) that correspond to the descriptions of specific proficiency levels (Cizek & Bunch, 2007). The identified correspondence between test scores and the proficiency levels of an external language framework is intended to enhance score interpretability, offering a useful indication of what an examinee at a specific score level can (and possibly cannot) do with the target language.