



## OPEN ACCESS

## EDITED BY

Itamar Lerner,  
University of Texas at San Antonio, United States

## REVIEWED BY

Paris Binos,  
Cyprus University of Technology, Cyprus  
Kee-Man Chuah,  
Universiti Malaysia Sarawak, Malaysia

## \*CORRESPONDENCE

Linyu Liao  
✉ christine\_liao@126.com  
Shelly Xueting Ye  
✉ yb87708@um.edu.mo

†These authors have contributed equally to this work and share first authorship

## SPECIALTY SECTION

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

RECEIVED 30 September 2022

ACCEPTED 15 March 2023

PUBLISHED 06 April 2023

## CITATION

Liao L, Ye SX and Yang J (2023) A mini review of  
communicative language testing.  
*Front. Psychol.* 14:1058411.  
doi: 10.3389/fpsyg.2023.1058411

## COPYRIGHT

© 2023 Liao, Ye and Yang. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](#). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted which  
does not comply with these terms.

# A mini review of communicative language testing

Linyu Liao<sup>1\*†</sup>, Shelly Xueting Ye<sup>2\*†</sup> and Jinsong Yang<sup>1</sup>

<sup>1</sup>School of Foreign Languages, Guangdong Medical University, Dongguan, China, <sup>2</sup>Department of English, University of Macau, Macau, China

Compared with traditional language testing, which heavily emphasizes psychometric reliability, communicative language testing (CLT), which uses authentic tasks to measure communicative abilities, has long been dominant in language assessment. Given its widely acknowledged advantages and widespread use, CLT has become less controversial in the language assessment field and thus is receiving decreased scholarly attention. However, real-world communication, in which CLT is grounded, evolves over time, suggesting the need to update our understanding of it. To address this need and facilitate the further development of CLT theories and practices, this paper offers an up-to-date review of CLT, including its various approaches, implementation challenges, and suggestions for future research.

## KEYWORDS

communicative language testing, approaches to CLT, challenges of CLT, language assessment, communicative abilities

## Introduction

As the communicative approach to language teaching gained popularity in the 1970s, communicative language testing (CLT) appeared alongside it and has gradually taken over language assessment in recent decades. Nowadays, most large-scale tests make the marketing claim that they assess test takers' communicative language ability either explicitly or implicitly (Fulcher, 2000). Contrary to conventional tests that rely heavily on achieving reliability through multiple choice items (see Lado, 1961), CLT stresses the test takers' ability to accomplish communicative purposes appropriately within authentic contexts (Morrow, 2018). Instead of assessing language knowledge (e.g., lexical, grammatical, or phonological knowledge) that test takers possess, CLT seeks to evaluate what they can do with the language (Morrow, 2012).

Because of these distinctive features, CLT has been widely used in test development practice and is thought to have become one of the important norms of test development (Fulcher, 2000). Given its prominent role and widespread use, CLT is generally considered an uncontroversial topic with little need for in-depth investigation or discussion when compared to other topics in language assessment. As a result, the considerable research attention that was given to CLT throughout the first decade after it was proposed has since decreased (Morrow, 2012). Meanwhile, it should be noted that CLT is grounded in real-world communication, which is constantly changing as time and technology progress (McNamara, 1996). These changes have facilitated the emergence of various CLT approaches as well as new challenges (Harding, 2014). Thus, the extant CLT research and discussions, most of which were undertaken decades ago (e.g., Canale and Swain, 1980; Canale, 1983), may not accurately reflect the current state of CLT and may fail to appropriately guide the development of language tests that can assess contemporary communicative language abilities. Therefore, to ensure the sustainable use of CLT, continuous and sufficient research should be conducted, particularly reviews that can present an accurate and up-to-date state of CLT. In response to this need, this paper offers a short and concise review of CLT that

includes an overview of the present CLT approaches, the as-yet-unsolved challenges to its implementation in large-scale assessments, and recommendations for future CLT research. Given the comprehensive updates provided in this review, it has significant implications for test developers and CLT researchers.

## Approaches to CLT

According to [Bachman's \(1990\)](#) and [McNamara's \(1996\)](#) introductions and [Harding's \(2014\)](#) summary of CLT approaches, the various CLT approaches have been classified into three categories: the theory-based approach, the real-life approach, and the integrated approach. The features of each of these are summarized below.

### The theory-based approach

This approach focuses on the measurement of underlying traits of communicative language ability (CLA). [Bachman's \(1990\)](#) model of CLA in use clarified that CLA includes language competences (e.g., vocabulary, syntax) and strategic competences (e.g., goal setting, planning, assessment). Accordingly, language tests developed using the theory-based approach and those specifically based on the [Bachman \(1990\)](#) model should contain tasks that can measure the model's stated competencies ([Fulcher and Davidson, 2007](#)). Among the illustrative projects provided by [Bachman and Palmer \(1996\)](#) to explain their use of this approach in test development is a test developed for a hotel seeking to hire employees to handle written English complaints. The linguistic and strategic competences necessary for this position were specified in detail and thus could be measured by the test. For instance, knowledge of English vocabulary for hotel facilities and services, as one of the language competencies required in the target domain, could be assessed by a task that asks test takers to provide the form and meaning of target words. Similarly, the strategic competency of assessment, which in this context refers to the evaluation of the appropriateness of responses to complaints, could be measured by a task requiring candidates to choose the best answers to a range of inquiries and complaints. [Bachman \(1990\)](#) and [McNamara \(1996\)](#) also referred to this approach as the interactional authenticity approach and the cognitive/psycholinguistic tradition, respectively. Given that these two approaches have the overlapping feature of measuring language skills based on theoretical models of CLA (e.g., [Canale and Swain, 1980](#); [Canale, 1983](#); [Bachman, 1990](#); [Bachman and Palmer, 1996, 2010](#); [Chapelle, 1998](#); [Douglas, 2000](#); [Purpura, 2004](#)), they are both regarded as falling under the category of the theory-based approach ([Harding, 2014](#)). According to these existing models, it is generally accepted that CLA is a multi-componential construct. However, consensus has not yet been reached on which components account for a comprehensive model of this construct (e.g., [McNamara, 1996](#); [Van Moere, 2012](#)). For example, it is uncertain whether the competences of managing emotional response ([Lindemann, 2002](#)) or comprehending different English varieties ([Canagarajah, 2006](#)) should be incorporated into the CLT construct.

Thus, the fundamental question of which specific components of CLA should be assessed still confronts the theory-based approach to CLT.

### The real-life approach

The real-life approach is atheoretical but authentic. It is also known as the real-life task-driven approach by [Bachman \(1990\)](#) and the work-sample tradition by [McNamara \(1996\)](#). This approach begins with the analysis of the target language use domain, which is then followed by the development of test tasks simulating real-life activities of language use. For example, language tests aimed at assessing prospective university students' English academic writing should include integrated listening-and-writing tasks that simulate the common university practice of completing a writing assignment after attending a lecture. It can be said that the real-life approach emphasizes simulation in test design ([McNamara, 1996](#)), and therefore prioritizes integrated tasks that represent language use in authentic contexts better than traditional independent tasks do. This approach has grown in popularity in recent years (see [Cumming, 2014](#)). In addition to being authentic, test tasks also need to be sufficiently diversified to elicit enough evidence of test takers' CLA. However, without practical guidance for using the real-life approach for CLT, the sampling and design of authentic and varied test tasks can be challenging due to the complexity of real-life communication.

### The integrated approach

The integrated approach is the combination of the theory-based approach and the real-life approach. On the one hand, it draws on existing theories of CLA to assess the underlying traits of communicative competence. On the other hand, it utilizes real-life tasks representative of target language use domain to ensure test authenticity. A clear example of this approach comes from a test designed to recruit administrative staff at an English-medium university based on [Bachman and Palmer's \(1996\)](#) illustrative projects. Since one of most frequent duties of this job is typing in handwritten documents, a test developed using the real-life approach might include tasks that require test takers to demonstrate their typing skills but not necessarily their English language competencies. However, applicants' subsequent performances on these tasks would probably fail to fully represent how well they will perform in this position. Given this, the theory-based approach, which requires the specification of linguistic and strategic competences that are necessary in the target context prior to test development, can be utilized in conjunction with the real-life approach to better serve the test purpose. This integrated approach could, for instance, develop a task requiring test takers to write English emails to announce department news, replicating the aforementioned real-life typing activities while also eliciting a demonstration of the test takers' English competencies. The integrated approach, which makes use of the strengths of both traditional approaches, is now regarded as the mainstream approach to CLT.

## Challenges of CLT

### The difficulty of construct operationalization

A long-standing CLT challenge is the difficulty of operationalizing CLA models for a specific context or test design. For example, when McNamara (1990) first utilized the innovative item response theory to validate language tests, he not only noted the general difficulty in implementing CLT in practice but also suggested that this challenge might be related to raters' varied perceptions of the importance of particular CLA components. Subsequent literature has also accepted this claim (Harding, 2014). McNamara (2003) in a later review observed that Bachman's model had seldom been used in test development projects. A similar observation was also made by Linda Taylor (as cited in Harding, 2014) while attempting to implement Bachman's (1990) CLA model in the International English Language Testing System (IELTS). She found it extremely challenging to turn CLT models into practical step-by-step instructions, which again emphasizes the difficulty in operationalizing the constructs of complex CLA frameworks. Some efforts have been made to solve this challenge. For instance, reports that describe the test development process of both the Common European Framework of Reference (CEFR) and the Educational Testing service's (ETS). Test of English as a Foreign Language (TOEFL) (e.g., Bejar et al., 2000; Butler et al., 2000) have been made public to assist test developers in better applying CLT to their assessment design. Additionally, recent research has explored people's perceptions of communicative ability from various perspectives (e.g., Plough et al., 2010; Abdul Raof, 2011; Kim and Elder, 2015; Pill, 2016), the findings of which have benefited the development of analytic assessment criteria. These criteria, which provide clearer descriptions of CLT constructs, might serve as practical references for test development based on sophisticated CLT frameworks. For example, parts of the rating criteria for evaluating oral communicative competence, which were initially identified by Davies et al. (1999), have been used by Pearson Education (2022) for its test of general English. In spite of the attempts that have been made during the past three decades, the notable challenge of construct operationalization has not yet been fully resolved and should continue to receive attention in the 21st century. As Harding (2014) comments, "[i]ncreasingly complex frameworks of communicative language ability may hinder as much as enable good test design" (p. 191) because they are too complex to be applied.

### The inadequacy of construct definitions

There is also an issue with the definitions of the constructs of CLA. On the one hand, exiting models tend to be too complex to be applied in practice and thus can only partially be validated (e.g., Harley et al., 1990; Phakiti, 2008). On the other hand, they are not yet complex or rich enough to provide a comprehensive definition of CLA (Purpura, 2008; Harding, 2014). Some alternative CLA models have been proposed to try and cover CLA components overlooked in older models. For

example, Purpura (2008) suggests that existing models should be supplemented with theories of Second Language (L2) learning and development. L2 is defined as the language that people use other than their mother tongue (Lado, 1957). However, a consensus on exactly which components constitute a comprehensive CLA model has not been reached. Moreover, CLA is not a static construct. No matter how comprehensive a CLA model is, it cannot satisfy the needs of each subsequent generation without constant evolution. According to Harding (2014), the theoretical constructs of CLA should incorporate changes to communicative practices brought about by technology such as those of online and mobile communication. He also highlighted a number of abilities that were neglected in the past but have eventually come to be recognized as CLA components as technology has advanced and communicative practices changed. Typical examples include (1) the ability to interact with interlocutors in paired or group conversations (Taylor and Wigglesworth, 2009), (2) the ability to adapt to various language varieties (Leung, 2005), and the ability to employ socio-pragmatic knowledge while communicating (Timpe, 2012). Additionally, it should be noted that test developers and experts in task content have different opinions about which CLA constructs are important, with the former focusing primarily on linguistic features and the latter emphasizing non-verbal elements and behaviors that aid in learning (Sato, 2018). Given this, if the CLT constructs were specified by test developers alone, they might fail to accurately reflect some specific context's communicative needs (Wette, 2011). Therefore, it is crucial to combine both perspectives in order to elicit an adequate definition of the CLT constructs.

### Feasibility in large-scale tests

CLT requires that test tasks be as naturalistic and authentic as possible, replicating real-life language use activities. Despite that this contextualized approach is beneficial and feasible in classroom settings, it can complicate assessment by introducing a variety of task demands and contextual factors, especially for large-scale tests (Morrow, 2018). Some tasks such as presentations, although authentic and practical, are difficult to incorporate into large-scale assessment. Since tasks of this type require particular knowledge and skills, test takers may not be able to accomplish them even in their first language without purposeful training and practice. Another example of this problem is the utilization of paired and group assessments to measure test takers' interaction skills in real-world situations (Taylor and Wigglesworth, 2009). Test takers' performances on these tasks involve multiple individuals, and therefore a single performance might not accurately represent that test taker's ability since they might be disadvantaged by being unlucky enough to be assigned to less-engaged or skilled partners (Isaacs, 2013). Furthermore, the ability to understand diverse language varieties, which is becoming increasingly important in communication today, is one of the CLA constructs that some assert should be measured in language assessments (e.g., Taylor and Geranpayeh, 2011; Ockey et al., 2016). However, test developers are responding cautiously to this suggestion (Elder and Harding, 2008) given their concern over the potential issue of

test bias when various varieties are added. That is, test takers tend to differ in their familiarity with language varieties, which likely affects their test performance and thus leads to unfairness (Ockey and Wagner, 2018). As a result, various compromises for feasibility must be made when designing authentic tasks for large-scale tests.

## The tension between face validity and reliability

CLT emphasizes face validity by requiring test takers to produce actual language in interactive tasks that resemble real-life communication. This, however, tends to reduce test reliability when compared with conventional discrete-point test items that focus on specific elements of language knowledge or skill because language performance involving qualitative rating is notoriously difficult to score reliably. In addition, since interactive tasks require both receptive and productive skills, tests may not be able to reliably assess test takers' real CLA if they are not equally capable in both types of skills. This is particularly true for test takers with a lower language proficiency. As Morrow (2018) explains, since beginners are usually better at comprehending language than producing it, their limited language performance cannot represent or capture the development of their receptive skills. Additionally, it is said that assessments based on CLT theories should measure not only linguistic knowledge but also the topic knowledge that is required in communicative contexts (Douglas, 2013) since communication in the real world often involves content exchange (Basturkmen and Elder, 2004). On the one hand, the evaluation of both knowledge types in a test has the potential to better represent the target language domain and thereby improve test validity. On the other hand, however, it may increase the difficulty of rating and score interpretation, thus having a negative impact on test reliability. It is probably because of this tension, the issue of whether and how linguistic and content knowledge should be assessed in language tests, that CLT test validity and reliability has attracted some attention in the field of English for specific research (Johns, 2012). Yet despite some research having been conducted, no conclusions have yet been reached. These illustrations demonstrate that maintaining a balance between face validity and reliability in CLT remains a challenging issue.

## Directions for future research

Based on the above descriptions of CLT features, approaches, and challenges, this section proposes some potential directions for future research.

- *Improve theoretical models of CLA.* Existing CLA models can be improved in two ways. In terms of content, they can be complemented with theories of L2 learning and development (Purpura, 2008) and incorporate new construct components brought about by the technological revolution

(Coiro et al., 2008), such as digital literacies (Gillen and Barton, 2010). In terms of practicality, efforts should be made to simplify complex CLA models into practical step-by-step guidelines so that the constructs of these models can be better operationalized (Harding, 2014).

- *Investigate new tasks that reflect modern digital communication.* Different forms of digital communication, whether they be written or verbal, are widely used in today's technologically advanced world. Language assessment needs to account for these changes that have been brought on by technology (e.g., Teasdale and Leung, 2000; Godwin-Jones, 2008).
- *Explore test takers' adaptability.* Since real-world communication often involves diverse and unfamiliar language varieties (Harding and McNamara, 2017; Harding, 2018), research should investigate test takers' ability "to deal with different varieties of English, to use and understand appropriate pragmatics, to cope with the fluid communication practices of digital environments, and to notice and adapt to the formulaic linguistic patterns associated with different domains of language use" (Harding, 2014, p. 194).
- *Utilize both independent and integrated tasks.* According to the principles of authenticity and interactivity (Bachman, 1990), integrated tasks are preferred in CLT since they can demonstrate test takers' competence in using multiple skills for effective communication (Cumming, 2013; Plakans et al., 2019; Rukthong and Brunfaut, 2020). Independent tasks, on the other hand, provide more precise information regarding test takers' various language skills, which is beneficial for both teaching and learning. Further research should be conducted to explore how the combination of the two task types can elicit information that is more useful for assessment and pedagogical purposes (Cumming et al., 2005; Plakans, 2010).

## Conclusion

Even though CLT is not new and has become a widely accepted approach for language assessment, test developers still find it challenging to apply CLT models to their test designs (Weir, 2005) due to several unresolved issues. Additionally, CLT is an approach that reflects real communicative activities. Given the rapid advancement of technology, real-world communication has evolved significantly since CLT was first proposed and is still changing. Therefore, our understanding of CLT must keep up with these ongoing changes. In other words, we should not let the discussions of CLT die (Harding, 2014). Motivated by this, this review has gone beyond prior CLT conversations (e.g., Bachman, 1990; McNamara, 1996), and described the dominant CLT approaches and unresolved challenges based on today's communicative practices. It not only provides readers with up-to-date information on communicative competence but also offers guidance for the future development of language tests. We encourage that dynamic, ongoing reviews of CLT be

conducted after accounting for the developments of modern real-world communication.

## Author contributions

LL: conceptualization and writing. SY: writing, revising, and editing. JY: revising and visualization. All authors contributed to the article and approved the submitted version.

## Funding

The work was supported by Education Department of Guangdong (project number 4SG22287G) and Guangdong Medical University (project number 4SG22027G), China.

## References

- Abdul Raof, A. H. (2011). "An alternative approach to rating scale development," in O'Sullivan, B. ed. *Language Testing: Theories and Practices* (Hampshire: Palgrave Macmillan), p. 151–63.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford, UK: Oxford university press.
- Bachman, L. F., and Palmer, A. S. (1996). *Language Testing in Practice*. Oxford, UK: Oxford University Press.
- Bachman, L. F., and Palmer, A. S. (2010). *Language Testing in Practice*. Oxford, UK: Oxford University Press.
- Basturkmen, H., and Elder, C. (2004). "The practice of LSP" in *The Handbook of Applied Linguistics*, eds A. Davies and C. Elder (John Wiley & Sons), 672–694.
- Bejar, I., Douglas, D., Jamieson, J., Nissan, S., and Turner, J. (2000). *TOEFL® 2000 listening framework: A working paper (TOEFL Monograph No. MS-19)*. Princeton, NJ: Educational Testing Service.
- Butler, F. A., Eignor, D., Jones, S., McNamara, T., and Suomi, B. K. (2000). *TOEFL® 2000 Speaking Framework: A Working Paper (TOEFL Monograph No. MS-20)*. Princeton, NJ: Educational Testing Service.
- Canagarajah, A. S. (2006). Changing communicative needs, revised assessment objectives: testing english as an international language. *Language Assess Quarterly* 3, 229–242. doi: 10.1207/s15434311laq0303\_1
- Canale, M. (1983). "From communicative competence to communicative language pedagogy," in Richards, J. C., and Schmidt, R. W., eds *Language and Communication* (London, UK: Longman), p. 2–27.
- Canale, M., and Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Appl. Linguist.* 1, 1–47. doi: 10.1093/applin/1.1.1
- Chapelle, C. A. (1998). "Construct definition and validity inquiry in SLA research," in Bachman, L. F. and Cohen, A. D., eds *Interfaces between second language acquisition and language testing* (New York: Cambridge University Press), p. 32–70. doi: 10.1017/CBO9781139524711.004
- Coiro, J., Knobel, M., Lankshear, C., and Leu, D. (2008). "Central issues in new literacies and new literacies research," in Coiro, J., Knobel, M., Lankshear, C., and Leu, D. eds *Handbook of research on new literacies* (New York, NY: Erlbaum), p. 1–21.
- Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly*, 10, 1–8. doi: 10.1080/15434303.2011.622016
- Cumming, A. (2014). "Assessing integrated skills," in Kunnan, A., ed. *The Companion to Language Assessment* (Hoboken: John Wiley and Sons), p. 216–29. doi: 10.1002/9781118411360.wbcla131
- Cumming, A., Kantor, R., Baba, K., Eouanzoui, K., Erdosy, U., and Jamse, M. (2005). Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL®. *ETS Res Rep Ser.* 1–77. doi: 10.1002/j.2333-8504.2005.tb01990.x
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. and McNamara, T. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Douglas, D. (2000). *Assessing Languages for Specific Purposes*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511732911
- Douglas, D. (2013). "ESP and assessment," in *The Handbook of English for Specific Purposes*, eds B. Paltridge and S. Starfield (Boston, MA: Wiley-Blackwell), 367–383.
- Elder, C., and Harding, L. (2008). Language testing and English as an international language: constraints and contributions. *Aust. Rev. Appl. Linguist.* 31, 34.1–34.11. doi: 10.1075/ara1.31.3.07eld
- Fulcher, G. (2000). The "communicative" legacy in language testing. *System* 28, 483–497. doi: 10.1016/S0346-251X(00)00033-6
- Fulcher, G., and Davidson, F. (2007). *Language Testing and Assessment*. Milton Park: Routledge.
- Gillen, J., and Barton, D. (2010). *Digital Literacies: A Research Briefing by the Technology Enhanced Learning Phase of the Teaching and Learning Research Programme*. London, UK: London Knowledge Lab, University of London.
- Godwin-Jones, R. (2008). Emerging technologies web-writing 2.0: enabling, documenting, and assessing writing online. *Language Learn. Technol.* 12, 7–13. doi: 10.10125/44138
- Harding, L. (2014). Communicative language testing: current issues and future research. *Language Assess. Quarterly* 11, 186–197. doi: 10.1080/15434303.2014.895829
- Harding, L. (2018). "Listening to an unfamiliar accent: Exploring difficulty, strategy use, and evidence of adaptation on listening assessment tasks," in Ockey, G., and Wanger, E., eds *Assessing L2 listening. Language Learning and Language Teaching* (John Benjamins), p. 98–112. doi: 10.1075/llt.50.07har
- Harding, L., and McNamara, T. (2017). "Language assessment: the challenge of ELF" in Jenkins, J., Baker, W. and Dewey, M., eds *The Routledge handbook of English as a lingua franca* (Routledge), p. 570–82. doi: 10.4324/9781315717173-46
- Harley, B., Allen, P., Cummins, J., and Swain, M. (1990). *The Development of Second Language Proficiency*. New York, USA: Cambridge University Press. doi: 10.1017/CBO9781139524568
- Isaacs, T. (2013). "International engineering graduate students' interactional patterns on a paired speaking test: Interlocutors' perspectives," in *Second language interaction in diverse educational settings* McDonough, K., and Mackey, A., eds, (Amsterdam: John Benjamins), p. 227–246.
- Johns, A. M. (2012). "The history of English for specific purposes research," in *The Handbook of English for Specific Purposes*, eds B. Paltridge and S. Starfield (Boston, MA: Wiley-Blackwell), 5–30.
- Kim, H., and Elder, C. (2015). Interrogating the construct of aviation english: feedback from test takers in Korea. *Language Test.* 32, 129–149. doi: 10.1177/0265532214544394
- Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press.
- Lado, R. (1961). *Language Testing*. New York, USA: McGraw Hill.
- Leung, C. (2005). Convivial communication: recontextualizing communicative competence. *Int. J. Appl. Linguist.* 15, 119–144. doi: 10.1111/j.1473-4192.2005.00084.x

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Lindemann, S. (2002). Listening with an attitude: a model of native-speaker comprehension of non-native speakers in the United States. *Language Soc.* 31, 419–441. doi: 10.1017/S0047404502020286
- McNamara, T. (1996). *Measuring Second Language Performance*. London, UK: Longman.
- McNamara, T. F. (1990). Item response theory and the validation of an ESP test for health professionals. *Language Test.* 7, 52–76. doi: 10.1177/026553229000700105
- McNamara, T. F. (2003). Looking back, looking forward: rethinking Bachman. *Language Test.* 20, 466–473. doi: 10.1191/0265532203lt268xx
- Morrow, C. K. (2018). “Communicative language testing,” in Liontas, J. L. ed. *The TESOL encyclopedia of English language teaching* (Hoboken: John Wiley and Sons), p 1–7. doi: 10.1002/9781118784235.eelt0383
- Morrow, K. (2012). “Communicative language testing,” in Coombe, C., Davidson, P., OSullivan, B. and Stoyoff, S., eds *The Cambridge guide to second language assessment* (Cambridge, England: Cambridge University Press).
- Ockey, G. J., Papageorgiou, S., and French, R. (2016). Effects of strength of accent on an L2 interactive lecture listening comprehension test. *Int. J. Listen.* 30, 84–98. doi: 10.1080/10904018.2015.1056877
- Ockey, G. J., and Wagner, E. (2018). *Assessing L2 Listening: Moving Towards Authenticity*. John Benjamins Publishing Company.
- Pearson Education (2022). *PTE Academic Test Takers Score Guide*. [https://assets.ctfassets.net/yqwtwibiobs4/3Bm0RMkKoNV0oOxUe38mg4/6a438b3a161b5cd313ddda516f2f500b/PTE\\_Academic\\_Score\\_Guide\\_for\\_Test\\_Takers\\_-\\_Nov\\_2022.pdf](https://assets.ctfassets.net/yqwtwibiobs4/3Bm0RMkKoNV0oOxUe38mg4/6a438b3a161b5cd313ddda516f2f500b/PTE_Academic_Score_Guide_for_Test_Takers_-_Nov_2022.pdf) (accessed August 21, 2022).
- Phakiti, A. (2008). Construct validation of Bachman and Palmers (1996) strategic competence model over time in EFL reading tests. *Language Test.* 25, 237–272. doi: 10.1177/0265532207086783
- Pill, J. (2016). Drawing on indigenous criteria for more authentic assessment in a specific-purpose language test: health professionals interacting with patients. *Language Test.* 33, 175–193. doi: 10.1177/0265532215607400
- Plakans, L. (2010). Independent vs. integrated writing tasks: a comparison of task representation. *Tesol Quarterly* 44, 185–194. doi: 10.5054/tq.2010.215251
- Plakans, L., Liao, J. T., and Wang, F. (2019). “I should summarize this whole paragraph”: Shared processes of reading and writing in iterative integrated assessment tasks. *Assess. Writ.*, 40, 14–26. doi: 10.1016/j.asw.2019.03.003
- Plough, I. C., Briggs, S. L., and Van Bonn, S. (2010). A multi-method analysis of evaluation criteria used to assess the speaking proficiency of graduate student instructors. *Language Test.* 27, 235–260. doi: 10.1177/0265532209349469
- Purpura, J. (2004). *Assessing Grammar*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511733086
- Purpura, J. E. (2008). “Assessing communicative language ability: Models and their components,” in Duff, P. A., and Hornberger, N. H., eds *Encyclopedia of language and education* (New York, USA: Springer), p. 2198–213. doi: 10.1007/978-0-387-30424-3\_167
- Rukthong, A., and Brunfaut, T. (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Lang. Test.* 37, 31–53. doi: 10.1177/0265532219871470
- Sato, T. (2018). The gap between communicative ability measurements: general-purpose English speaking tests and linguistic laypersons judgments. *Papers Language Test. Assess.* 7, 1–31.
- Taylor, L., and Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalizing the test construct. *J. English Acad. Purpose.* 10, 89–101. doi: 10.1016/j.jeap.2011.03.002
- Taylor, L., and Wigglesworth, G. (2009). Are two heads better than one? Pair work in L2 assessment contexts. *Language Test.* 26, 325–339. doi: 10.1177/0265532209104665
- Teasdale, A., and Leung, C. (2000). Teacher assessment and psychometric theory: a case of paradigm crossing?. *Language Test.* 17, 163–184. doi: 10.1177/026553220001700204
- Timpe, V. (2012). Strategic decoding of sociopragmatic assessment tasks—an exploratory think-aloud validation study. *Second Language Stud.* 30, 109–246.
- Van Moere, A. (2012). A psycholinguistic approach to oral language assessment. *Language Test.* 29, 325–344. doi: 10.1177/0265532211424478
- Weir, C. J. (2005). *Language Testing and Validation*. Hampshire: Palgrave Macmillan.
- Wette, R. (2011). English proficiency tests and communication skills training for overseas-qualified health professionals in Australia and New Zealand. *Language Assess. Quarterly* 8, 200–210. doi: 10.1080/15434303.2011.565439