



Content list available at <http://ijltr.urmia.ac.ir>

*Iranian Journal
of
Language Teaching Research*

BOOK REVIEW



Urmia University

***Language Testing and Assessment: A Comprehensive Guide
by Masoomeh Estaji and Farhad Ghasvand (2022), 252 pp.
ISBN: 978-964-2175-17-8***

Ali Momeni ^{a,*}

^a *Allameh Tabataba'i University, Iran*

Language Testing and Assessment: A Comprehensive Guide by Estaji and Ghasvand (2022) constitutes a quintessential example of an instructive and informative textbook for English as a Foreign Language (EFL) students eager to navigate the subtleties of assessment in language education. The book opens new horizons for students and reinforces their assessment insights by unraveling the models, theories, notions, and practical considerations of language testing and assessment. The book is divided into 14 chapters and features critical and research-oriented content that could lay the groundwork for avid researchers to satiate their theoretical and practical knowledge of testing and assessment.


Chapter 1 casts light on the fundamental concepts of test, measurement, evaluation, assessment, modes and types of testing, scoring procedures, ipsative assessment, and interim assessment. These conceptualizations have been offered through simple and germane explications, and it is opined that these concepts might have conceptual overlaps. The chapter concludes with a discussion of four major paradigms in language testing, including *pre-scientific*, *psychometric-structuralist*, *psychometric-sociolinguistic*, and *communicative* paradigms. Chapter 2 points out the most prevalent functions of language tests and their uses in teaching through feedback that they provide for teachers and institutions concerning their effectiveness. The authors hold that practitioners can offer stakeholders helpful information based on the function a test fulfills. The chapter also provides a critical account of various functions of assessment grounded in instructional, administrative, and research classifications.

Chapter 3 affords a detailed account of various principles of language testing, namely validity, reliability, practicality, washback effect, and interactiveness, in light of Bachman and Palmer's (1996) test usefulness model. In the spirit of the frameworks offered by leading figures (e.g., Messick, 1989), the progressive matrix of validity and the significance of construct validity have also been explored in this chapter. Through a comparative lens, this chapter elucidates the

* Corresponding author: Allameh Tabataba'i University, Tehran, Iran

Email address: momeni_ali@atu.ac.ir

© Urmia University Press

 10.30466/ijltr.2023.121408

distinctions between test authenticity and interactiveness, drawing upon Widdowson's conceptualization of genuineness and authenticity. Chapter 4 underscores three notions of model, framework, and construct to establish the foundation for discussing prominent models of communicative competence, such as Bachman's (1990) model of communicative language ability and Canale and Swain's (1980) model of communicative competence. The authors outline the primary steps involved in constructing items and tasks, taking advantage of reverse engineering and prototyping concepts.

Chapter 5 is at the interface of three major statistical theories common in language testing, namely classical true score theory (CTS), generalizability theory (G-theory), and item response theory (IRT). To this end, the premises, merits, and demerits underlying each theory have been approached critically throughout the chapter. The authors have also incorporated various diagrams and visual representations to enhance the comprehensibility and clarity of comparing such theories. Chapter 6 unravels different conceptualizations of validity, ranging from the question of "Does the test measure what it claims to measure?" (p. 57) to test usefulness and validity argument. In addition, the genesis of validity, including gestational, crystallization, fragmentation, (re)unification, and deconstruction period, have been discussed in sufficient detail. In the chapter's concluding section, a comparative analysis is undertaken regarding the conceptualization of test use argument and Toulimin's model, which encompasses six fundamental components, namely *claim, grounds, warrant, backing, modal terms, and rebuttal*.

Chapter 7 deconstructs various assessment approaches and techniques concerning the assessment of the four major language skills (e.g., speaking, listening, reading, and writing). Furthermore, the skills are described with their respective typologies, enabling a more specific and accurate assessment of their nature and complexities. Assuming a critical stance, the last part of the chapter deals with the assessment of L2 pragmatics and English as Lingua Franca (ELF), which have been novel lines of inquiry within the last couple of years. Chapter 8 presents an overview of statistics used in language assessment and its fundamental techniques. For this purpose, descriptive and inferential statistics, accompanied by numerous schematic representations of data, such as bar graphs, line graphs, polygon, scatter plots, and box plots, are replete with sample figures and appropriate exemplifications. The chapter also provides detailed explanations of the concepts of t-score, z-score, and CEEB-score, establishing their relationship with hypothesis testing.

Chapter 9 features a preliminary account of traditional testing techniques and approaches (e.g., discrete point and integrative tests, norm-referenced and criterion-referenced tests) regarding their drawbacks and theoretical underpinnings in teaching. The authors also define the concept of "*bias for best*," which implies that the language test should aim to measure the test-taker's performance optimally (Van Moere & Hanlon, 2020). Furthermore, the authors call for alternative assessment techniques that can meet the communicative needs of language learners in real-life settings. Chapter 10 emphasizes several alternative assessments and the advantages of each for assessment and educational purposes. As a case in point, portfolio and e-folio have been proposed to demonstrate the trajectory of language development in learners instead of employing one-shot assessments. The conclusion of the chapter delves into a research-focused discussion on *cognitive diagnostic assessment*, which investigates different theories and assumptions regarding how attributes interact, and influence observed item response patterns (Ravand & Baghaei, 2020).

Chapter 11 stresses the significance of classroom assessment practice and, by extension, distinguishes formative assessment from assessment *for* learning, assessment *of* learning, and assessment *as* learning. A critical review of classroom assessment literacy in the context of language education, along with a lucid specification of knowledge in assessment principles and theories, test development and interpretations, and scoring, brings the chapter to a close. Chapter 12 mainly underlines the social dimensions of language testing and assessment. The chapter covers considerations of test bias and fairness, as well as explanations of the consequences of the tests

regarding students' personal, intellectual, and social lives. In particular, the ethics and standards governing language tests' development, administration, and scoring are explored. Additionally, the chapter emphasizes the importance of prioritizing *ethical codes* within educational contexts, as this serves to clearly articulate the established rules of conduct that have been mutually agreed upon by the professional community.

Chapter 13 covers the conception, genesis, and research themes of critical language testing as a social component that concentrates on the hidden agendas of language tests involving power relations and ideologies forced on the test-takers. In this chapter, the authors maintain that no test is devoid of power, ideology, imposition, and hierarchical organizational structure. The chapter's final suggestion is that educational settings should shift their focus toward adopting a more *democratic* assessment framework. Chapter 14 addresses the adoption of technology in language testing and assessment. The chapter highlights significant advantages that the integration of technology may provide to language testing, particularly in test design, item generation, item tryout, test management, scoring, and communicating the findings through electronic applications. The chapter also discusses the need to redefine the construct being measured when incorporating technology into language testing and assessment.

This illuminating book boasts a myriad of suits for students in the area of testing and assessment in language education. The first noteworthy characteristic is its brevity, which communicates crucial concepts with the minimum number of pages possible (e.g., 10 to 15 pages for each chapter). The second positive feature is that the authors have backed up their assertions with recent references to prominent figures, each an expert on the chapter's topic. A further positive feature of this book is the utilization of simple-to-understand language to communicate the otherwise complex testing and assessment notions. The authors also skillfully elaborate on the complexities of assessment and testing, rendering them more comprehensible to readers of varying levels of educational knowledge. Additionally, the provision of a series of multiple-choice tests along with a glossary of language testing and assessment at the end of the book renders it a more practical textbook.

Despite the merits of the book, there appears to be some room left for improvement. The chapter-bound presentation of the multiple-choice questions would be far preferable to including some progress tests at the end of the book. The authors are encouraged to include activities at the end of each chapter to enable the readers to reflect on the chapter's content. Better proofreading will remove the misspellings and APA violations in the manuscript and reference lists. The chapter highlights at the beginning of each chapter could be updated to correspond more closely with the chapter's content. The information repeated in chapter 8 (e.g., CRT versus NRT) has been elaborated upon in subsequent chapters (e.g., chapter 9) of the book. As such, the authors might revise this chapter's content to make it more concise. Additionally, the authors could have supplied some practical and instructive guidelines on integrating technology in language testing in chapter 14, such as a list of online testing applications, websites, and programs. It is also crucial to acknowledge that including a single introduction and conclusion chapter by the authors can substantially enhance the systematicity and overall coherence of the chapters. However, these minor shortcomings could not diminish the overall excellence of this course book on language testing and assessment.

References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47. <http://dx.doi.org/10.1093/applin/I.1.1>
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp. 13-103). Macmillan.
- Ravand, H., & Baghaei, P. (2020). Diagnostic classification models: Recent developments, practical issues, and prospects. *International Journal of Testing*, 20(1), 24-56. <https://doi.org/10.1080/15305058.2019.1588278>
- Van Moere, A., & Hanlon, S. (2020). A Bayesian approach to improving measurement precision over multiple test occasions. *Language Testing*, 37(4), 482-502.