

A Book Review of "Practical Language Testing" by Glenn Fulcher (2010): From the Stage of Test Design to Score Delivery

Sayed Hadi Sadeghi^{1*}

Received: 23 July 2022

Accepted: 20 November 2022

Abstract

The book *Practical Language Testing*, authored by Glenn Fulcher (2010), is an epitome of comprehensive and detailed documentation on the relevance of the appropriate techniques, knowledge, and conceptual frameworks required to interpret and develop language assessments. It is also a systematic, helpful reference for researchers and educators that address assessment selection and design concerns, offering instructors guidance on picking the proper test for a particular educational setting and assisting in interpreting the scores generated by such tests. The book lays the groundwork for instructors to establish a thoughtful awareness of test scenarios to furnish the most efficient language assessment for their learners. The book's language is highly lucid and user-friendly to both experts and novices. The book finds an appropriate amalgamation between practical applications and theoretical underpinnings to empower practitioners in comprehending the objectives of effective assessment and its function in facilitating good learning and teaching. On account of the fact that the referenced book has the potential to make a substantial contribution and have significant implications for teachers, test designers, and other stakeholders, the current article attempts to provide a review of the book and emphasize the critical concepts in each chapter.

Keywords: practical language testing; score interpretation; test administration; test construction; test purpose

1. Introduction

Teacher education programs depend partly on language testing and assessment textbooks to provide teachers and administrators with a body of knowledge on testing methodologies and test plans informed by research and practitioner experience. Multiple language testing and assessment textbooks are available on the market, and teachers need to determine which book would greatly facilitate testing purposes in their settings. Once educators, adjuncts, and instructional assistants are familiarized with the features of a particular textbook, they will become more cognizant of the philosophies and advantages of the book to discern whether it meets their respective needs and interests. One such consideration is reviewing the important books in the realm of language testing and assessment.

¹ Assistant Professor at Supreme National Defense University, Tehran, Iran. Research Assistant at the University of Sydney, Australia. (*corresponding author*, ssad2473@alumni.sydney.edu.au)

Practical Language Testing is a well-known textbook that needs to be given sufficient prominence. *Practical Language Testing* finds an appropriate amalgamation between practical applications and theoretical underpinnings to empower practitioners in comprehending the objectives of effective assessment and its function in facilitating good learning and teaching.

This functional manual may address numerous intersecting communities of experienced language test writers, executives, scorers, and instructors. It is widely obtainable and well-organized for its teacher audience. *Practical Language Testing* describes the testing procedure in 10 chapters, from the earliest phases of constructing an exam to the last processes of sending the result to the test-taker. The book covers considerations of reliability, scorability, and dependability, which are essential characteristics of high-quality testing. To aid the reader's comprehension, specialized and general topics are demonstrated using pertinent literature, graphs, figures, images, and colorful illustrations. Each chapter concludes with an activity component that encourages users to employ the new information they have received from the readings. This book's glossary of fundamental language testing terminology is an excellent hallmark.

The first chapter, *Testing, and Assessment in Context*, sets out to outline the essential information with regard to test purposes, tests in educational systems, testing rituals, unintended consequences of testing on society (e.g., washback, migration, health issues), the historical interlude of tests, the politics of language testing, professionalizing language education and testing, and the validity of tests. In this light, the book strikes sufficient insight into the contemporary contexts of pedagogical assessment, analyzing its motives, procedures, and ramifications, such as the issues regarding internally/externally required examinations and the frequent negative washback. In this chapter, it has been signposted that formative tests are in lockstep with internal mandate tests, and summative tests are in line with external mandate tests as they are aimed at measuring proficiency at the end of the courses. In addition, the notion of distributive justice was discussed, which "deals with the appropriateness of access to the conditions and goods that affect individual well-being, which is broadly conceived to include psychological, physiological, economic and social aspects" (Fulcher, 2010, p. 21).

The second chapter, *Standardised Testing*, points out the two paradigms of tests, test score analysis, introducing and calculating reliability, and the relationship of reliability with test length and other measures. The chapter elucidates the distinction between norm-referenced tests and criterion-referenced tests, and it is highlighted that these two are different in terms of the type of interpretation and measurement, the purpose of testing, distribution of scores, test structure, and knowledge of questions. Moreover, the conceptualization of trait theory of validity was sophisticated, signaling whether the test measures the intended construct under investigation or not. Further, the concept of validity, like other books on language testing and assessment, focuses on traditional views, including the definition proposed by Messick (1989). The chapter also reviews a more unfamiliar concept such as scorability, which is defined as the simplicity with which a test question or assignment may be scored; therefore, a set of multiple items is more easily scored than an open-ended response to a writing task.

The third chapter, *Classroom Assessment*, is subsumed under the components of the assessment for learning, assessment of learning, self and peer assessment, assessment and second language acquisition, and the dependability of classroom tests. In the spirit of the focal

point of this chapter, it is maintained that assessment for learning is more compatible with the learners' needs and interests than the assessment of learning because the assessment for learning procedures, such as dynamic assessment, will provide the learners with the supplementary mediation in the form of clues and hints that can activate their zone of proximal development (ZPD) to perform the assessment task successfully (e.g., Ashraf, Tabatabaee-Yazdi & Samir, 2016; Estaji & Farahanynia, 2019; Lantolf & Poehner, 2010; Momeni, 2022; Poehner, 2007; Safdari & Fathi, 2020; Samir & Tabatabaee-Yazdi, 2020; Sadeghi, 2022). Note should be taken that the first three chapters discuss the importance or role of testing within society and the education system, although they are implied in the first chapter.

The fourth chapter, *Deciding What to Test*, attends, in detail, to the test design cycle, constructs definitions and pertinent issues, and models of communicative competence. The central tenet of this chapter lies in the interface of the recursive/spiral process of test design, which will go through the process of test purpose, test criterion, construct definition, task/item design (e.g., specification writing), evaluating prototyping and piloting, field testing (e.g., assembling into final format), inferences, and decisions (Fulcher, 2006; Fulcher & Davidson, 2009). According to the author, the construct constitutes "the abilities of the learner that we believe underlies their test performance, but which we cannot directly observe. These begin as 'concepts', and we can identify them because they are usually abstract nouns" (Fulcher, 2010, p. 113).

The fifth chapter, *Designing Test Specifications*, is concerned with the matter of test specification, the concern of specifications for testing and teaching, granularity, and target language use domain analysis. In this respect, the test specification is conceptualized as the blueprint/design documents of the test through which we can develop our intended tests. In light of the report by Mislavy, Almond, and Lukas (2003), the possible specifications include items/task specifications (e.g., the format of instructions), evidence specification (e.g., response attribute), test assembly specification (e.g., specifications of items for each category), presentation specification (e.g., presentation of items), and delivery specification (e.g., details of test administration, test security, and timing). Furthermore, it is contended that item-objective congruence, also called item-specification congruence, tries to make an association between the task/item and the educational learning purposes to pave the way for designing items/tasks that coincide with the learning objectives of various courses. As a critical point raised against test specifications, Popham (1992, 1994) recommended that test specifications be restricted to a basic description and a sample item since they can limit a teacher's creativity if they are too detailed, restrictive, and comprehensive.

The sixth chapter, *Evaluating Prototyping and Piloting*, tackles significant considerations such as investigation of usefulness and usability of the test, evaluation of items, tasks, and specifications, guidelines for multiple-choice items, prototyping, piloting, field testing, item shells, and operational item review and pre-testing. This chapter touches upon the parameter of critical reverse engineering, in which test constructors critically select a sample item and examine it to determine what it is testing, if it is a meaningful item, and analyzes what challenges they would have if they utilize the item. It is asserted that prototyping involves obtaining the commentary of a testing expert on the items (e.g., alpha prototyping) and then trialing them with a small group of students. The difference between prototyping and piloting

rests in the number of learners who are trialed by the items; in prototyping, the number of learners is relatively small, whereas, in piloting, it is quite large. However, according to the author, if test constructors intend to undertake complex statistical analysis on the pre-test findings, they may choose field testing, which employs a significantly larger number of students and exposes test constructors to more consistent estimates of item statistics.

The seventh chapter, *Scoring Language Tests*, puts emphasis on the facets of scoring items, scorability, scoring constructed response tasks, automated scoring, and the role of guessing in testing. It examines the issues linked with measurement variables, including the interpretation of a score and the lengthy time required for scoring. In the chapter, scorability is described as a pivotal factor influencing the effectiveness of a test, and numerous indications of early and ongoing methodologies for scoring close-response items are provided. The chapter further delves into the various approaches to rating scales, including holistic scales (e.g., overall quality of performance), primary trait scales (e.g., specific response for each prompt or question), and multiple traits scoring (e.g., two or more scores for different prompts or traits).

The eighth chapter, *Aligning Tests to Standard*, elaborates on the conceptions of the standards and the definition of standard-based testing, using standards for harmonization and identity, performance level descriptors, and standard-setting methodologies. This chapter addresses the complication resulting from adopting standards and standard-based testing, as well as the issues connected with this complication. To this end, two prevalent methods for standard-setting procedures were mentioned, namely the test-centered method (e.g., Angof method, Ebel method, Nedelsky method, and Bookmark method) and the examinee-centered method (e.g., borderline group method, and contrasting group method). Furthermore, the author has stated three sources of proof to assess a standard-setting study, including procedural, internal, and external evidence (Kane, 1994).

The ninth chapter, *Test Administration*, comprehensively discusses the issues of controlling extraneous variables in testing, the role of planned and unplanned variations in testing, scoring and moderation, and reporting outcomes to the stakeholders. The chapter illuminates the principle that the interpretability and usefulness of test scores have a direct, and positive relationship with the test administration status quo and that the test's fairness will be strengthened in many ways if the test administration circumstances are adequately observed. In compliance with the authors' remarks, cooperation is essential to guarantee test security of score interpretation, test validity, and test result accuracy in order to administer a test effectively. A competent test administration, he argues, demands offering all test-takers the same opportunity to succeed under uniform settings. In addition, the chapter conceptualizes test administration as a function of managing extraneous variables (e.g., air conditioning breakdown), implying that the quality of test administration processes increases as the number of extraneous variables decreases.

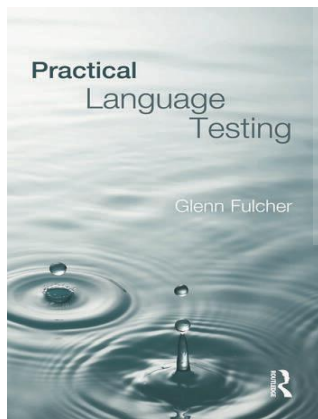
The tenth chapter, *Testing and Teaching*, allocates some space to dwell on the prominent account of washback, the interface of washback and content alignment, learners' preparation for tests, and the selection and utilization of tests. In the chapter, the concept of washback is defined as "the extent to which the introduction and use of a test influences language teachers and learners to do things that they would not otherwise do that promote or inhibit language learning" (Fulcher, 2010, p. 113). Furthermore, the washback hypothesis is

argued, claiming that the test is affected by what teachers teach, how teachers teach, what learners learn, how learners learn, and the rate and sequence of teaching and learning.

Practical Language Testing is a valuable reference for practitioners who are considering which examinations are most appropriate. By focusing largely on the successful design and administration of assessments, the book assists educators in developing high-stakes tests that are consistent with their curriculum and benchmarks. In addition, the book informs instructors on how to improve learners' test scores by incorporating all required modifications to ensure authenticity and the validity of the test. The book's language is highly lucid and user-friendly to both experts and novices. The author has offered an extensive, systematic, and understandable overview of essential themes and concerns in language testing. A further promising characteristic of the book is the activity section at the end of each chapter that will allow the users to check their understanding of the presented concepts of the chapters. However, a subsequent book edition might benefit from slight modifications to some sections. It is recommended that the beginning section of each chapter encompasses abstract-focused content to introduce to the reader what will be highlighted in the chapter. Additionally, the author can take some space at the end of each chapter to introduce further research works and recommendations for practice. The author is also encouraged to put together high-quality papers that address the most crucial concerns in language assessment and to offer additional information and documentation for testing language skills and sub-skills. Moreover, the author did not pay much attention to the discussion of the key terms and models related to item response theory, comprising the modern test theory. Lastly, chapters seven and nine are out of order because scoring procedures should be determined before test administration and this should be taken into consideration in future revisions of the book.

Appendix

Important Information about the Book



Author	Glenn Fulcher
Publisher	Hodder Education
Pages	352 Pages
ISBN-13	9780340984482 0340984481
Price	\$44.95 USD
Edition	1st Edition
First Published	2010
E-book Published	25 March 2010
Publication	London
Location	
Imprint	Routledge
DOI	https://doi.org/10.4324/980203767399
Subjects	Language and Literature

References

- Ashraf, H., Tabatabaee-Yazdi, M., & Samir, A. (2016). An in-depth insight into EFL university students' cognitive processes of C-test and X-test: A case of comparison. *International Journal of Language Testing*, 6(2), 101-112.
- Estaji, M., & Farahanynia, M. (2019). The immediate and delayed effect of dynamic assessment approaches on EFL learners' oral narrative performance and anxiety.

-
- Educational Assessment*, 24(2), 1–20.
<https://doi.org/10.1080/10627197.2019.1578169>
- Fulcher, G. (2006). Test architecture. *Foreign Language Education Research*, 9, 1–22.
<https://rb.gy/153irg>
- Fulcher, G. (2010). *Practical language testing* (1st ed.). London: Routledge.
<https://doi.org/10.4324/980203767399>
- Fulcher, G. & Davidson, F. (2009). Test architecture, test retrofit. *Language Testing*, 26(1), 123–144. <https://doi.org/10.1177%2F0265532208097339>
- Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64(3), 425–461.
<https://doi.org/10.3102%2F00346543064003425>
- Lantolf, J. P., & Poehner, M. E. (2010). Vygotsky's teaching- assessment dialectic and L2 education: The case for dynamic assessment. *Mind, Culture, and Activity*, 17(4), 312–330. <https://doi.org/10.1080/10749030903338509>
- Mislevy, R. J., Almond, R. G. & Lukas, J. F. (2003). *A brief introduction to evidence-centred design*. Research Report RR-03–16. Princeton, NJ: Educational Testing Service.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11.
- Momeni, A. (2022). Online assessment in times of COVID-19 lockdown: Iranian EFL teachers' perceptions. *International Journal of Language Testing*, 12(2), 1–24.
- Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *Modern Language Journal*, 91(3), 323–340.
<https://doi.org/10.1111/j.1540-4781.2007.00583.x>
- Popham, W. J. (1992). A tale of two test-specification strategies. *Educational Measurement: Issues and Practice*, 11(2), 16–17. <https://doi.org/10.1111/j.1745-3992.1992.tb00235.x>
- Popham, W. J. (1994). The instructional consequences of criterion referenced clarity. *Educational Measurement: Issues and Practice*, 13(4), 15–18.
<https://doi.org/10.1111/j.1745-3992.1994.tb00565.x>
- Sadeghi, S. H. (2022). *Smart Campus E-Readiness: A Framework for Cyberspace Learning Strategic Management*. Routledge. <https://doi.org/10.4324/9781003198826>
- Safdari, M., & Fathi, J. (2020). Investigating the role of dynamic assessment on speaking accuracy and fluency of pre-intermediate EFL learners. *Cogent Education*, 7(1), 1–19.
<https://doi.org/10.1080/2331186X.2020.1818924>
- Samir, A., & Tabatabaee-Yazdi, M. (2020). Translation quality assessment rubric: A rasch model-based validation. *International Journal of Language Testing*, 10(2), 101–128.