Book Review:

The Influence of Technology on Language Testing[1]

Reviewed by Qianru Li[2] & Kan Kan Chan[3], University of Macau, Macao, China

[Fulcher, G. & Harding, L. (Eds.) (2022). *The Routledge handbook of language testing* (2nd ed.), Routledge]

Since the handbook *The Routledge handbook of language testing* is in its second edition, Glenn Fulcher and Luke Harding not only added a new section – *Assessing the language skills*, but also spent more effort discussing the changes in language assessment with a computerized delivery mode due to technological innovations and recent global health pandemic. The new edition of this handbook remains useful for language test researchers to adopt as a reference book and is inspiring for doctoral students who consider carrying out language assessment research with computer-based test mode. The handbook consists of ten sections, including: validity, the use of language testing, classroom assessment and washback, assessing the language skills, test design and administration, writing items and tasks, prototyping and field tests, measurement theory in language testing, technology in language testing, and ethics, fairness, and policy. The ten sections are divided into thirty-five chapters and one summative chapter with an informative index.

To begin with, Carol Chapelle and Hye-won Lee introduce the history of validity in the first chapter *Conceptions of validity*. This is followed by a description of the current views towards validity and its uses in language constructs, language development and language testing. Finally, the authors identify the importance of validity to test developers, testing researchers, and external evaluators. In Chapter 2, *Articulating a validity argument*, the author Michael Kane highlights that the validity argument is to make an interpretation of tests. Interpretation/use arguments and validity arguments are discussed in terms of development with examples of different assessments. In the last chapter of this section, *Inference and prediction in language testing*, Steven Ross mainly presents the recent validity prediction of evaluating language learners' affective and cognitive factors which have influenced in their second language acquisition outcome with an instance of a longitudinal method.

The first chapter of the second section, *Social dimensions of language testing*, which is almost identical to the first edition of this handbook. Richard Young illustrates two social dimensions of language testing on construction and consequences on individuals and languages. Carol Moder and Gene Halleck focus on the work-related language for specific purposes testing with the example of Aviation English. Understanding the socio-political context, communication with stakeholders, and consideration in the following test design and validation are of importance to language testers. Antony Kunnan compares present-day voluntary immigrants with those in the past. Although language policy and language tests have been used to address the concerns related to immigration and citizenship in the United States, they have been found to be a barrier to citizenship in the last chapter of this section, *Revisiting language assessment for immigration and citizenship*.

Janna Fox, Nwara Abdulhamid, and Carolyn Turner introduce *Classroom-based assessment* at the beginning of the third section. They underline how the rise of technology might influence the traditional format of classroom-based assessment, an updated trend that the authors newly include in this chapter. In the next chapter, *Washback*, Liying Cheng and Nasreen Sultana, who present a review of recent washback literature, find that the current meaning of "washback" is based on innovative research questions and solid conceptual and theoretical frameworks. Yuko Butler, in the next chapter, *Assessing young learners,* suggests that the way young language learners master a language is different from that of adults. Age-appropriate language assessment is recommended to help individual students learn a language. With the development of digital technology, the author of this chapter finally predicts that Artificial Intelligence (AI) might alter the way children learn a language. In the next chapter, *Dynamic assessment*, Marta Antón and Próspero García discuss the theory and development of dynamic assessment to assess second language abilities. The authors mainly focus on presenting the use of dynamic assessment in different language assessment contexts. In the final chapter of this section, *Diagnostic assessment in language classrooms*, Eunice Jang and Jeanne

[2] yc07101@connect.um.edu.mo
[3] kankchan@um.edu.mo

Sinclair introduce the features and issues in diagnostic assessment. The way technology significantly improves diagnostic assessment of speaking proficiency and reading comprehension are discussed by the authors.

As a new section of the second edition of this handbook, it is thoughtful that authors divided *Assessing the language skills* into four chapters, including speaking, listening, writing, and reading. Since these four-skill tests are essential to providing the most accurate picture of a test-taker's language proficiency (Powers, 2010), it is important to understand the advancement of the assessment of the four skills. Fumiyo Nakatsuhara, Nahal Khabbazbashi, and Chihiro Inoue start by using a quote from Lado (1961) to exemplify the difficulty of *speaking assessment*, followed by a presentation of four speaking tasks and three methods of analysis. The authors also mention that technology has been found to influence the construct of speaking test and the test delivery mode. Regarding *Listening assessment*, Elvis Wagner highlights that listening ability involves the ability to understand spoken input and suggests that technology could be utilised to provide test-takers with more real-life context. In the next chapter, *Assessing writing*, Ute Knoch points out the complexity of second language writing, which makes writing assessment very difficult. Based on the concerns related to the evaluation of writing, the scoring reliability in both human and machine raters offering student feedback is discussed. *Assessing reading*, in the last chapter of this section, Tineke Brunfaut presents several standard test formats for assessing reading, such as multiple-choice, gap-fill tasks, matching, true-false-not given, information transfer, and short-answer questions. To find individual differences, it is recommended that research in reading assessment could be expanded to use eye-tracking technology to analyze test-takers' reading fluency and reading rate through eye movement in a reading test.

Section Five relates to test design and administration. The first chapter, *Test specifications*, Yan Jin highlights the importance of specifications on high-stakes language testing, classroom-based language assessment, and test validity. Chengbin Yin and Robert Mislevy start the next chapter, *Evidence-centered design in language testing*, by identifying two frameworks. The first is a comprehensive framework of evidence-centered design, which has been adopted in the context of language testing. The second is conceptual assessment framework, including an evaluation component and measurement model. In the next chapter, *Accommodation and universal design*, Jamal Abedi presents the effectiveness on examining accommodations and its validity for English language learners with different types of dictionaries and language tests. In the final chapter of this section, Larry Davis emphasizes the necessary training and validity frameworks for both rater and interlocutor, who are considered to have potential impacts on test-takers' speaking performance.

The next section, *Writing items and tasks*, is mainly about writing tasks and the task design. Dongil Shin summarizes four approaches towards *item writing and item writers* and reviews its critical issues. In the following chapter, *Writing integrated tasks*, Lia Plakans initially defines the meaning of integrated assessment, followed by comparing the integrated and independent skills tasks with unique task and performance features. To comprehend test results beyond statistics, Andrew Cohen adopts a non-psychometric approach in the chapter *Test-taking strategies and task design*. Construct validity has been underlined between the strategies that test-takers use and the purpose of the test in this chapter.

Prototyping, piloting, and field testing are three important stages on test design with an example from the Educational Testing Service illustrated by Susan Nissan and Elizabeth Park in the first chapter of Section Seven, *Prototyping new item types*. Benjamin Kremmel, Kathrin Eberharter, and Franz Holzknecht demonstrate that the role of *Pre-operational testing* in two argument-based validation models in the following chapter. These three authors also provide suggestions for practice from planning, administration, data retrieval, analysis, and storage to documentation and dissemination. John Read finally exemplifies the process of *piloting vocabulary tests* in the last chapter.

James Browns primarily defines *classical test theory* and states the history of it in education and psychology, and language testing, followed by a discussion about the significance of statistical research methods towards this theory in Chapter 26. In the next chapter, Gary Ockey describes *Item response theory and many-facet Rasch measurement* to show the association between individuals' observed performances and latent variables. In the *Reliability and dependability* chapter, Xun Yan and Jason Fan stress reliability in language testing since it continuously affects the validity and fairness of a test. They further present factors that might impact on test reliability. Evelina Galaczi and Gad Lim show their considerations on human and machine raters, and rubric in *Scoring performance tests*, in the three stages including design, development, and operation.

The chapters In Section Nine, *Technology in language testing*, have been redone into a new section to highlight technical assistance in language assessment. Xiaoming Xi shows the development of automated scoring on computerized test and expresses concerns about the issue of validity in *Validity and the automated scoring of performance tests* chapter. In the next chapter, *Computer-based testing*, Yasuyo Sawaki summarizes three phases of the use of a computerized language test along with a brief comparison between a paper-based test and a computer-based test. In the final chapter of *Corpus linguistics and language testing*, Sara Cushing presented the significance of corpus linguistics that boosts the advancement of language assessment with explanations on the validity of a test, language proficiency, and automated scoring and feedback tools.

Ethnics, fairness, and policy are essential to language assessment and are discussed in the final section of this Handbook. Francis Walters begins by defining the important roles of ethics and fairness on language testing in the first chapter *Ethics and fairness*. In the next chapter, *Standards in language proficiency measurement*, Bart Deygers presents different kinds of language proficiency standards and critically examines them for not reaching the standard criteria of specificity, reliability, and universal recognition. In *Quality management in test production and administration*, Nick Saville and Sarah McElwee reveal that quality management is essential for enhancing assessments and guaranteeing tests reach professional standards by applying quality management to the assessment cycle. In the last chapter, *Language testing: where are we heading*, is written by Luke Harding and Glenn Fulcher who are also the authors of this Handbook. They summarize their main findings by recognizing the emerging role of technology in language assessment, especially the use of digital technology in both language learning and assessment related to the global pandemic. In addition, they suggest technology helps to reduce the complexity of language assessment by using corpus. They finally remind the reader that language testing scholars need to have a critical perspective towards language testing based on the features of socio-political forces.

To conclude, this Handbook is updated to allow readers who are trying to use digital technology to think about how to maximize the effectiveness of it into teaching and language testing. In addition, the historical background, theories and frameworks, challenges and limitations are well discussed. The inclusion of technology is the biggest difference in comparison with the contents of the first edition as it has been applied into every aspect of language testing. As for the format of the Handbook, Fulcher and Harding were thoughtful enough to provide brief introductions to the contributors of this Handbook at the beginning to allow readers to meet the experts in this research field at the early stage and this can help readers to easily follow up on those specialists' latest research. Regarding possible improvement of this second edition, the first suggestion is is to a new chapter related to integrated tasks. Although speaking, listening, writing, and reading skill assessments have been presented, integrated tasks have been mentioned across four chapters. In this case, a new chapter entitled *Integrated tasks in four language skills tests* could be added to allow readers to include more integration in language assessment, as it is represented in daily-life and academic interaction. The second suggestion is related to the use of abbreviations. In Chapter 18, the concept of ELL students is needed to be written as English Language Learner for the first time of this chapter to help readers have a clear idea about what ELL indicates. Apart from the two suggestions, this Handbook serves as a useful foundation for the use of technology in language testing.

References

Lado, R. (1961). Language testing: The construction and use of foreign language tests (A teacher's book). Longman.

Powers, D. E. (2010). The case for a comprehensive, four-skills assessment of English language proficiency. *R & D Connections, 14*, 1-12. https://www.ets.org/Media/Research/pdf/RD_Connections14.pdf