

BOOK REVIEW

Sadeghi, K., & Douglas, D. (Eds.). (2023). *Fundamental Considerations in Technology Mediated Language Assessment (1st ed.)*. Routledge. 2023. Pp 302.

In response to the unprecedented situation that was caused by the COVID-19 pandemic, test administrations were required to take place online in many places over the world. Such a transition made it indispensable to incorporate technology into language assessment, beyond the notion that test administrations are facilitated, rather than mandated, by technology use. Consequently, many language testing programs were immediately urged to discuss whether and how to transform their pencil-and-paper tests into online tests, including those that have to be delivered remotely, by making use of technologies that they had not used for their tests. This dramatic shift in logistics naturally engendered testing programs' rethinking of their tests from existing and new perspectives that are fundamental to maintaining validity of score interpretation and use. In response to these immediate needs, *Fundamental considerations in technology mediated language assessment*, edited by two well-known language testing researchers—Karim Sadeghi and Dan Douglas—introduces theoretical and practical considerations surrounding technology-mediated language assessment (TMLA). The book is intended to fill the gap in previous literature by explicitly addressing the relationship between theoretical underpinnings of language testing and technology use. As such, this book introduces up-to-date issues around TMLA along with traditional frameworks, constituting “the current end point for TMLA” (p. xviii). The book consists of four major sections each of which includes three to five chapters along with the foreword given by Lyle Bachman and the introduction chapter authored by the two editors. This book review firstly summarizes each of the four major sections and then provides a review of this book.

Part I addresses validity issues in TMLA, consisting of five chapters. Chapter 2 introduces long-standing theoretical models of language ability, such as those proposed by Canale and Swain (1980) and Bachman (1990), and provides a comprehensive review of previous studies that compared conversations in face-to-face and online conditions. While

providing this review, the chapter further points out some potential shortcomings of the existing models to capture spoken interactions that are to occur in online environments. Chapter 3 introduces a digital assessment system called BalanceAI and reports an empirical study that investigated the relationship between self-regulated learning and writing ability while also reporting the effects of the system's feedback on writing ability. The findings suggested that the relationship between self-regulated learning and writing ability was not linear and that BalanceAI's tutoring sessions facilitated the writing ability of students, including those who had learning difficulties. Chapter 4 introduces the importance of striking a balance between measurement efficiency and assessment meaningfulness by providing a real-world example of the TOEFL Essentials test. While the chapter introduces many aspects of the TOEFL Essentials test, such as test design, scoring, and test security, the authors particularly direct the reader's attention to three major challenges of distinguishing many different levels of test takers' proficiency, establishing the relationship between test takers' performance and their performance in the real-world, and supporting decisions that are made based on test scores. Chapter 5 introduces a case study about a university placement test in Australia that was undertaken in response to their curriculum reform. Based on the data they collected from test takers and staff via a survey and focus groups, they found that test task instructions and technical instructions were often missed by test takers, and this inattentiveness rendered their test scores less representative of their language ability. Chapter 6 introduces a nation-wide test in Finland and provides empirical evidence that supports the comparability between the test administration in 2013 (i.e., paper-and-pencil test) and the test administration in 2020 (i.e., digital). Comparing scores between test takers who took the digital version of the test at home and those who took the test at school, the study reported that the former group received significantly higher scores, though the magnitude of the difference was not substantial.

Part II centers around reliability issues of TMLA and consists of three chapters. Chapter 7 reports on an empirical study that compared ratings that were assigned to handwritten essays and ratings that were assigned to typed essays. Their findings indicated that handwritten essays tended to receive higher scores in language use, potentially because raters took into consideration the fact that it is difficult to revise language errors when

test takers handwrite their essays. Furthermore, the results suggested that raters were able to evaluate essays while precisely recognizing typed words, which led to relatively harsher evaluations of test takers' language use. Chapter 8 introduces a validation study that explored an AI-driven scoring system that was used in the IELTS Smart Learning application. Employing a many-facets Rasch measurement approach, the study found that the scoring system provided ratings that were not biased towards the intended users of this application. Chapter 9 introduces fundamental considerations that have to be given when one designs an automated scoring system for short-answer tasks by the use of regular expressions. While this chapter focuses on the impact of authoring of regular expressions on reliability and construct validity, the chapter also introduces specifications of many online platforms that enable convenient test administrations with regular expressions.

Part III consists of three chapters that address fairness, security, and ethical considerations. Chapter 10 introduces a real-world effort that was made at WIDA in response to the surging needs to transform their large-scale tests to online, remotely delivered tests due to the pandemic. The chapter firstly provides a brief introduction of two large-scale tests (i.e., WIDA Screener, ACCESS) and their decision-making processes during their discussions. It was only possible to develop an online, remotely delivered version of WIDA Screener, but not for ACCESS, due to the difference in stakes between those two tests. Chapter 11 addresses the impact of technology on diagnostic language assessment that aims to provide feedback for test takers and thus enhance their learning experience. After introducing two major diagnostic language assessment systems (i.e., DIALANG, DELTA) and their limitations, the chapter discusses major limitations of the existing systems in general and proposes potential solutions. Chapter 12 discusses the impact of remote proctoring on test security while also introducing a variety of remote proctoring systems that are used in the existing testing programs. Furthermore, the chapter explicates key issues in test security, such as authentication, data security, cheating, and technical knowledge while also providing a framework that can be used to guide the best implementation of remote proctoring.

Consisting of five chapters, Part IV primarily addresses classroom assessment that is supported by technology use. Chapter 13 introduces interdisciplinary research collaborations that are undertaken at the Cambridge Institute of Automated Language Teaching and Assessment (ALTA) in order to improve learning-oriented assessment harnessing technology. The chapter firstly provides a brief introduction to learning-oriented assessment and the ALTA and then underscores the central role of collaboration among different fields, such as machine learning, natural language processing, and psychology. Chapter 14 introduces an online speaking assessment that is intended to measure interactional competence at a Vietnamese university and evaluates the speaking assessment in terms of five criteria—authenticity and interactions, practicality and impact, validity, reliability, and fairness. Although some support was obtained for the online speaking assessment, the chapter raises serious concerns regarding test takers' accessibility to the high-speed internet infrastructure and their familiarity with technology. Chapter 15 reports on a study that reviewed 11 studies on test taking strategies. The review found that test takers tend to use a wide variety of strategies—language learner strategies, test-management strategies, and test-deviuousness strategies—while also proposing future research directions regarding test taking strategies that are specific to TMLA. Chapter 16 discusses the impact of the pandemic on research methodology, focusing on the transition from in-person to online data collection procedures. By reflecting on a study where one of the authors was required to collect data from in-service teachers online, the chapter provides important considerations that should be made when researchers plan their online data collection.

This book has two major advantages. First, as the book contains many real-world efforts to transform existing pencil-and-paper tests to online tests, readers will quickly notice practical constraints that are inherent in TMLA. In particular, many chapters stressed that test takers' accessibility and familiarity to technology, the high-speed internet in particular, cannot be taken for granted, indicating the utmost importance of considering fairness as a repeatedly emerging theme in this book. Depending on the contextual factors, many aspects of the test that were initially deemed ideal and viable for pencil-and-paper tests have to be compromised simply because they cannot be realized in a way that is most appropriate in an online environment. It should be noted that this real-world dilemma

was faced not only by testing programs that offer large-scale tests, such as the TOEFL Essentials test (Chapter 4) and ACCESS (Chapter 10), but also by local test providers whose resources are relatively limited, including an Australian university in Chapter 5 and a Vietnamese university in Chapter 14. Those chapters effectively introduce the reader to their own unique context, thereby making it easy for the reader to digest their programmatic local needs (Dimova et al., 2022). While those chapters introduce context-specific dilemmas, some chapters introduce practical constraints that seem to be context general, such as Chapter 9 on authoring answer keys by the use of regular expressions, and Chapter 12 on online remote proctoring that is to guard against test takers' cheating. Those technology-specific chapters will be an informative reference for readers who are interested in those particular aspects of TMLA, not necessarily in a specific context that harnesses TMLA. By presenting real-world efforts that were made to create online tests and specific technologies that facilitate language assessment, the book will undoubtedly benefit a wide range of readers who are interested in TMLA as an academic exploration as well as those who are currently faced with real-world challenges that pertain to TMLA.

The second major advantage is that many chapters introduce theoretical frameworks at the beginning, facilitating readers' understandings of fundamental considerations regarding TMLA in a specific research context. For instance, Chapter 2 provides a comprehensive overview of language ability models while explicitly motivating the explorations into the dimensions of language ability in online speaking assessments that are difficult to cover with language ability models that were intended for conversations in in-person environments. Chapter 15 also provides a research background of test-taking strategies by introducing traditional and revised classifications of test-taking strategies along with empirical studies that addressed main themes in this research area. Indeed, technology does not stand alone; rather, technology is used for certain purposes and thereby has to be facilitative of aspects of language assessment within a particular theoretical framework. According to the editors, it was their felt scarcity of this strong connection between theoretical frameworks and technology uses that in part motivated them to edit this book. In this view, the book certainly fills the gap in the literature by providing "theoretical considerations in language assessment as a consequence of technology integration...by revisiting key traditional assessment concepts" (p. 7).

Despite these reader-friendly features, the book seems to have some room for improvement. Echoing the existing reviews of this book (Wang & Wang, 2024; Yu & Lin, 2024), it was felt that chapters could be grouped into sections in a way that is more intuitive. For instance, as Wang and Wang (2024) rightly pointed out, though Part IV is titled *Options and issues in technology mediated classroom assessment*, not all chapters address classroom assessment practices. Chapter 13 appeals the need for interdisciplinary collaborations among experts in different fields mainly by introducing the ALTA's internal research program, and Chapter 15 provides a review study on test-taking strategies that do not necessarily pertain to classroom assessment given its inclusion of studies on large-scale tests, such as TOEFL iBT (Yang & Plakans, 2012) and TOEFL Primary (Lee & Winke, 2018). In addition, the title of Part I, *Validity concerns in technology mediated language assessment*, may sound too general for some readers, potentially not giving them a clear idea at first glance about the specific content discussed in this section. In this view, it would not have hurt the quality of this book if the editors had arranged the chapters in a different way. Perhaps they could have aggregated chapters that reflected on real-world consequences that were caused by the transition from the pencil-and-paper mode to the online mode (Chapters 4, 5, 6, 10, 16), chapters reporting empirical studies on certain aspects of TMLA (Chapters 3, 7, 8, 14, 15), and chapters reviewing existing theoretical frameworks and introduce considerations regarding particular technological tools (Chapters 2, 9, 11, 12, 13). That being said, as Yu and Lin (2024) mentioned, this blurred demarcation among chapters can be ascribed to the intricate nature of issues surrounding TMLA. It is not an easy undertaking to have a consistent and reasonable theme across different TMLA research contexts, especially for a book that consists of chapters contributed by different authors. While some background knowledge in language assessment may be required for readers, the book will be a valuable reference for any individuals who are interested in utilizing technology for language assessment purposes. In accordance with its title, the book certainly provides considerations that are fundamental to appropriate uses of technology in language assessment.

In conclusion, this book introduces many real-world undertakings that vividly embody knowledge construction from the pragmatism perspective, which is explicated by Poehner

and Inbar-Lourie in their introduction to this special issue. In other words, those real-world efforts were driven by immediate needs caused by the worldwide pandemic that had to be satisfied for the sake of best possible assessment practices at the time. Such needs will have to be satisfied many years ahead as well, potentially in ways that are more ethical than those used thus far. In this sense, chapters focusing on particular technological affordances for stakeholders who need to operate their language tests in the midst of ongoing and future emergency conditions, such as a pandemic or war, are particularly relevant. Accordingly, many readers of this special issue, which also describes stakeholders' innovations and intelligent actions in response to the pandemic, will benefit from reflecting on the problems and potential solutions described in this book. This special issue and the reviewed book will be a good companion to each other, collectively offering fundamental considerations in language assessment enhanced by the use of technology.

Reviewed by Taichi Yamashita

Kansai University, Japan

Author disclosures

The author declares no conflicts of interest.

ORCID iD

Taichi Yamashita  <https://orcid.org/0000-0001-8531-7266>

References

- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
<https://doi.org/10.1093/applin/I.1.1>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Dimova, S., Yan, X., & Ginther, A. (2022). Local tests, local contexts. *Language Testing*, 39(3), 341–354. <https://doi.org/10.1177/02655322221092392>

- Lee, S., & Winke, P. (2018). Young learners' response processes when taking computerized tasks for speaking assessment. *Language Testing*, 35(2), 239–269. <https://doi.org/10.1177/0265532217704009>
- Wang, X., & Wang, J. (2024). Review of the book *Fundamental Considerations in Technology Mediated Language Assessment*, edited by K. Sadeghi & D. Douglas. *Language Assessment Quarterly*, 21(3), 292–296. <https://doi.org/10.1080/15434303.2024.2375656>
- Yang, H. C., & Plakans, L. (2012). Second language writers' strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly*, 46(1), 80–103. <https://doi.org/10.1002/tesq.6>
- Yu, N., & Lin, D. (2024). Review of the book *Fundamental Considerations in Technology Mediated Language Assessment*, edited by K. Sadeghi & D. Douglas. *Asia Pacific Journal of Education*. <https://doi.org/10.1080/02188791.2024.2337466>