

***Ethics and context in second language testing: rethinking validity in theory and practice***

M. Rafael Salaberry, Albert Weideman, and Wei-Li Hsu

(2023). New York, Routledge, 248 pages

ISBN: 9781032471754

---

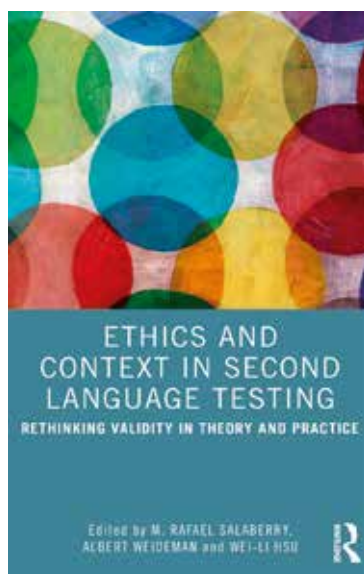
JING WANG

College of Foreign Studies, Liaoning University, China

DOI: <https://doi.org/10.30827/portalin.vi42.30001>

A resounding body of scholarship highlights the necessity of considering ethical concerns and social impacts of language tests and assessment practices in second language (L2) education (Brown & Harris, 2016; Davies, 2008; Fan et al., 2017). Without taking these issues into account, ensuring the right, reliable, and fair judgment is a tough task for L2 practitioners (Cohen, 2006; Gipps & Stobart, 2009). Inspired by the significance of ethics in L2 assessment, various studies have been done on the perceptions of stakeholders and common professional codes of ethics in the field of L2 education (Brown et al., 2020; Shohamy, 2020). Nevertheless, the situated nature of ethics across contextual particularities and in light of test validation theory has remained under-addressed. Urged to bridge this gap, Salaberry et al. (2023) compiled a handbook entitled “*Ethics and Context in Second Language Testing: Rethinking Validity in Theory and Practice*”, as a timely response to the need for a critical approach to L2 assessment. In so doing, they collected the voices of different stakeholders from different languages and contexts to comprehensively picture the status of social justice and equity in language education. The book, hence, provides an overview of key concepts, theories, and models of ethics and validity in language education and makes seminal suggestions for researching and practicing these constructs. This invaluable handbook is momentous for teachers, students, and researchers, especially those majoring in applied linguistics, sociolinguistics, and language policy and planning.

The book is divided into three parts covering three macro approaches to ethics and context in language testing. The first part entitled “*the ethical contextualization of validity*” includes two chapters that briefly introduce the volume and explicate some ethical challenges, contextualization concerns, and theoretical foundations of language testing and assessment. The first chapter provides a bird’s eye view of the whole volume and re-conceptualizes different constructs including language ability, validity and validation process, and ethical assessment design. They further elaborate on different perspectives of validity and ethics and touch the major turning points in L2 education, namely social, embodied, and multilingual turns. In the second chapter, Weideman and Deygers describe traditional conceptions of test validity,



goals of testing, and some test qualities such as usefulness, meaningfulness, impact, fairness, and beneficence. They also explain various conceptual complications in understanding and operationalizing test validity theory and its associated contested views and uncertainties. Particularly, they argue for the existence of various competing paradigms in validity research that raise objections against the linguistic turn in defining validity. To solve the problem, they also suggest a differentiation between subjective validation and objective validity that adds adequacy to the conceptualization of validity theory.

The second part of this book labelled “*agency and empowerment prompted by test adequacy*” deals with the ethical consequences of tests in relation to agency and empowerment. It comprises three chapters (3, 4, and 5). As a starting point, in Chapter 3, Richardson describes critical race theory (CRT) and raciolinguistic ideologies in language testing. She draws on Shohamy’s (1993) call for multilingual and multimodal tests that echo one’s linguistic background instead of imposing a monolingual ideology. She reports on a sample ideological imposition in Arizona, USA and urges educators to challenge ‘objective truths’ and move towards democratic assessment. Trying to present a contesting view of validity in Asia, in Chapter 4, West and Thiruchelvam depict the effect of a local English exit exam on both teachers and students at a university in South Korea. They used interviews to collect the data and analyzed them via positioning theory. The results highlighted the role of washback in validity theory and the need to go beyond psychometric models of testing and considering the consequential validity of test score interpretations. In Chapter 5, Suzuki takes a critical perspective to review empirical studies on empowerment and agency in L2 portfolio assessment. She argues that portfolio assessment is another form of imposed testing that does not necessarily develop L2 learners’ agency and empowerment. As a remedy, the author offers a pathway to implement L2 portfolio assessment, which is more ethically desirable and socially valid.

The third section, which is the lengthiest part of this resource, encompasses four chapters on “*sociointeractional perspectives on assessment*” (6, 7, 8, and 9). In Chapter 6, Räsänen and Kivik examine the facilitating role of portfolio assessment in language learning and how it can produce a positive washback effect. They used a portfolio task in the Finnish and Estonian language courses at a North American university and found it beneficial for learners’ independent language use and created a positive washback effect. Their study underscores the importance of authentic and situated language use in L2 education. Moving to classroom-based assessment tests, in Chapter 7, Kley (2024) describes how language competence can be assessed through a multimodal approach to language use. She used two speaking test tasks (paired and group) to explore German learners’ topic initiations and corresponding scores. Specifically, she used an open-topic task and a group discussion. Using conversation analysis (CA), the author found the topic initiation task balanced among test-takers, yet the discussion was misleading and unbalanced due to the dominance of some test-takers to start the conversation. To solve this issue, Kley suggests a change in the rubric for the discussion task or the task design. The affordances of L1-L2 speaker interaction for assessment are explained in Chapter 8 by Kley, Kunitz, and Yeh. They report the influence of the interlocutor’s native/non-native-speakerness on 28 Chinese as a foreign language (CFL) test-taker’s production of repair practices at a US university. The authors conclude that the native speaker interaction provides more affordances for initiating repairs and a lin-

guistically asymmetric test setting is more beneficial. In the last Chapter, Weideman (2024) proposes a yardstick for conceptualizing measurement and directing the future of language assessment. He presents a test designer's reflection on facing juridical and ethical issues in language testing. Weideman opposes the cynical view that no development is on the way of testing as power governs the field. He illustrates a progression in teachers' assessment literacy despite throwbacks but expects more evolution. The author calls for a criterion to start knowing and gauging development and "a theory of disclosure" to understand the meaning of language test design.

Overall, this volume is momentous for L2 education practitioners and researchers given its use of a contextualized and critical approach to L2 assessment ethics. The book is also advantageous for integrating different voices regarding ethical concerns across languages and contexts. Another merit of this resource is the smooth transition from traditional conceptualizations of ethics and validation toward the latest turning points (critical and multilingual turns). The use of empirical data to evidence arguments is also a strong point in this book. Methodological and design diversity of sample studies effectively illuminates concerns about social justice and equity in language education. Finally, the presentation of further readings and future directions after each chapter is praiseworthy.

Despite these benefits, this volume has some weaknesses, too. First, the distribution and length of chapters were not consistent across the sections. It would have been better to use an equal number of chapters under each part. Second, it is unfortunate that the book does not have a glossary of terms to foster the understanding of key concepts of this line of research. Another drawback is that the book presents different theories and models within chapters whose origin and full-description are missing. This makes ideas disconnected and dispatched rather than fully fitted. Finally, it would have been more insightful if the voices of researchers, test developers, administrators, and students from other countries had been included in the volume. A critical issue as ethics required a voluminous book with more chapters. Regardless of these weaknesses, this book is significant for its provision of a contextualized and critical approach to ethics in L2 assessment, which is absent in existing resources in this area. It helps test developers by understanding the role of ethical issues in testing and how they can employ specific test tasks that observe such concerns. Many books on language testing ethics revolve around fairness, justice, and equity in a theoretical fashion, while this volume is a combination of theory, practice, and research on ethical concerns. Therefore, it is a fruitful resource of L2 teachers, learners, trainers, and policy-makers in that they can use the global and contextualized ideas to observe and implement ethics in language testing and modify misleading practices into helpful ones.

## FUNDING INFORMATION

Basic Research Projects for Universities (Humanities and Social Sciences) of the Educational Department of Liaoning Province in 2022, A Comparative Study of Conceptual Metaphors in Chinese and British Environmental News Reports (LJKMR20221388).

**REFERENCES**

Brown, G. T., & Harris, L. R. (Eds.). (2016). *Handbook of human and social conditions in assessment*. Routledge.

Cohen, J. (2006). Social, emotional, ethical and academic education: Creating a climate for learning, participation in democracy and well-being. *Harvard Educational Review*, 76(2), 201–237. <https://doi.org/10.17763/haer.76.2.j44854x1524644vn>

Davies, A. (2008). *Assessing academic English testing English proficiency*. Cambridge University Press.

Fan, X., Johnson, R., & Liu, X. (2017). Chinese university professors' perceptions about ethical issues in classroom assessment practices. *New Waves Educational Research & Development*, 20(2), 1–19.

Gipps, C., & Stobart, G. (2009). Fairness in assessment. In C. Wyatt-Smith, & J. Joy Cumming (Eds.). *Educational assessment in the 21st century* (pp. 105–118). Springer.

Shohamy, E. (2020). *The power of tests: A critical perspective on the uses of language tests*. Routledge.