Review of Winke, P.; Brunfaut, T. (2021) Routledge handbook of second language acquisition and language testing

Jong, N.H. de

Book review: Routledge Handbook of Second Language Acquisition and Language Testing

P. Winke and T. Brunfaut. *Routledge Handbook of Second Language Acquisition and Language Testing*. New York, NY: Routledge, 2021. 542 pp. ISBN: 9781138490680 (hbk) £160; ISBN 9781351034784 (e-book) £31.99

Reviewed by: Nivja H. de Jong, *Leiden University, The Netherlands*

In linguistics, the big question "how are languages learned?" can be seen as an overarching question to which other fundamental questions, such as "how are languages used?," "how do languages differ?," "how do languages change?," and "how is language understood and produced in real time?," are related. The communicative, structural, sociocultural, and cognitive viewpoints engaged to answer these questions are likewise key when asking the question "how are *second* languages learned?" In other words, the field of second language acquisition (SLA) may perhaps be seen as the broadest subfield within linguistics, embracing all perspectives to understanding language learning.

A difference between more applied SLA research and more fundamental linguistic research is the attention for individual differences that is central in (applied) SLA but usually ignored in other fields. For instance, cognitive linguists have predominantly charted general processes of language users, treating individual differences as error variance (Kidd et al., 2018). SLA researchers, on the other hand, are inherently interested in the individual differences between language users because an understanding of how such differences can be explained may inform language teaching.

Because of this inherent need to chart (and explain) individual differences, it is no wonder that SLA researchers need to know how to measure such individual differences: individual differences with respect to all types of language performance but also measures of psychological and social characteristics underlying processes related to language development. Besides the already mentioned communicative, structural, cognitive, and sociocultural perspectives and approaches from the field of linguistics, another research field that is, therefore, typically relevant for SLA research is psychometrics.

The field of language testing (LT) has always busied itself with precisely this question: How can language knowledge and abilities be measured, combining insights from linguistics (for the *language* part in the term) with psychometrics (for the *testing* part)? Language testing experts and researchers in the field of SLA thus share a common interest: measuring individual differences that have to do with language knowledge and abilities.

The *Routledge Handbook of Second Language Acquisition and Language Testing* aims to cover these cross-roads between the research fields and, in doing so, to explicitly share expertise. Accordingly, the handbook is a comprehensive overview for researchers

interested in the measurement of language performance. As such, it is a valuable resource for researchers interested in SLA in general and for researchers involved in assessing language learners in particular. The volume successfully brings together 79 leading scholars from SLA and LT, an immense job for which the editors Paula Winke and Tineke Brunfaut must be applauded. They set the stage in the first chapter by problematizing measurement of language abilities by introducing various perspectives on what it means to "know" a language and chart the links between SLA and LT.

After this introductory chapter, Part 1 ("Assessment concepts for SLA researchers") introduces basic concepts that are deemed essential for researchers in SLA: validity in assessment, how to validate assessments for research, and how assessment literacy is necessary for SLA researchers. In addition, two chapters are informative for SLA and LT researchers alike: on frameworks and corpora in language assessment. Part 2, however, titled "Building instruments for SLA research," again gives the impression to be mostly written from the viewpoint of LT, giving advice to researchers in the field of SLA. Important topics in valid and reliable instrument building are discussed: from the test design cycle to the need for pilot testing and to rater behavior (among other topics).

The next two parts, Part 3 "Measuring individual differences" and Part 4 "Measuring Language Development," provide a state-of-the-art overview for research on measuring aspects that influence language development (Part 3) and measuring aspects of language performance (Part 4). Part 3 describes issues, recent research, and tips for the often-researched individual differences aptitude, motivation, working memory, anxiety, and willingness to communicate. These chapters are helpful for researchers investigating SLA in more general terms and for researchers investigating questions in the more specific field of LT. Part 4 is equally helpful for both research fields and encompasses issues and recommendations for a wide variety of language knowledge and abilities: from phonology, to lexicogrammar, pragmatics, the "traditional" skills (reading, speaking, reading, writing), and interactional competence, to name just a few. In addition, there are chapters on less-researched relevant abilities such as intercultural competence and literary competence. With its 18 chapters, it is the most voluminous part of the book; the other parts consisting of four to seven chapters each. Although the part is titled "Measuring Language Development," the chapters within the part mostly explicate on how to measure linguistic abilities at one time, thus measuring language performance as a static concept, although development may be individually shaped and nonlinear. Such an individual, dynamic approach would require other or additional techniques apart from those that are generally suggested in this part.

Part 5 considers who is measured and tested and introduces important considerations for assessing young learners, multilinguals and bilinguals, learners with specific learner difficulties (such as dyslexic students), heritage language learners, and sign language learners. Part 6, finally, delves into quantitative methods for SLA and LT research. Statistical modeling, constructing equivalent test forms, reliability, and generalizability are thoroughly discussed.

Although the title of the handbook rightfully suggests that sharing expertise between SLA and LT research is worthwhile and helps researchers in both fields, many chapters in the book are more about how SLA research can benefit from LT expertise than the other way around. This is also apparent from the titles of Parts 1 and 2 (concepts and

instruments for SLA) and the recommended readings given at the end of many chapters. One small example of how expertise from LT is presented as helpful for SLA is the suggestion that SLA doctoral students would benefit from learning about Rasch analysis (in the chapter by Harding and Kremmel, titled "SLA researcher assessment literacy"). Indeed, in LT, Rasch analysis has been a prominent technique to measure latent traits and capture item difficulties. Of course, it is useful to learn about this technique, but perhaps it is just as worthwhile to learn about logistic regression in mixed models where persons can be added as random effects, thus capturing latent traits. It will be even more worth the while, in my opinion, to understand how the Rasch model is actually a type of generalized linear mixed model (Agresti, 2013, pp. 492–493) and why the Rasch model is not the best option in all circumstances (Dixon, 2008; Doran et al., 2007). G theory, the topic of the last chapter of Part 6 and better-known in LT than in the broader field of SLA, can likewise be seen as a special case of linear mixed-effects modeling (Jiang, 2018); the variance components in G theory being the variances from random effects in a (random intercepts) linear mixed model. It should be noted that the chapter on statistical modeling by In'nami and Murakami in Part 6 does helpfully describe possibilities of mixed-effects modeling in SLA research but without relating it to the special cases as they have been used in LT.

This small example shows how actually more ground is shared between the research fields than now sometimes may be perceived by researchers from each field. In addition, although the handbook (especially in Parts 3 and 4) will be helpful for SLA and LT researchers alike, the bidirectionality could also have been more pronounced if another shared interest would have been included, namely the interest to investigate and measure processes that occur during task performance, such as cognitive, interactional, and sociocultural processes. From a more fundamental viewpoint in SLA research, this type of research helps to better understand the processes leading to development; from a more applied viewpoint in LT research, these investigations are necessary to validate tasks and items within tasks (e.g., Embretson & Gorin, 2001). To mention just a few shared techniques, conversation analysis is used in both fields for understanding mechanisms during speaking tasks and interaction, eye-tracking is used for investigating reading and speaking processes, and stimulated recall or thinking-aloud protocols are used to gauge individual's cognitive processes during different types of task performance.

To conclude, the handbook is a valuable resource for researchers seeking up-to-date information on how to measure competencies, abilities, and traits that have to do with language development. Such measurements are important for both researchers from the broad field of SLA and those from the more specific field of LT. The handbook can be perceived as somewhat imbalanced, with more chapters written from an LT perspective for SLA researchers than the other way around, or than by considering both fields of expertise equally in mind. Specifically, the handbook does not describe how to investigate processes *during* task performance, which constitutes another cross-roads of the two research fields and may have led to a more balanced overview of shared interests. However, this would also have led to an even heavier handbook, and with its current 500+ pages weighing over a kilogram, the comment may be better seen as a call for a second handbook, rather than criticism on the present volume.

ORCID iD

Nivja H. de Jong https://orcid.org/0000-0002-3680-3820

References

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Wiley.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, *59*(4), 447–456. https://doi.org/10.1016/j.jml.2007.11.004

Doran, H., Bates, D., Bliese, P., & Dowling, M. (2007). Estimating the multilevel Rasch model: With the lme4 package. *Journal of Statistical Software*, *20*(2), 1–18. https://doi.org/10.18637/jss.v020.i02

Embretson, S., & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement*, *38*(4), 343–368. https://doi.org/10.1111/j.1745-3984.2001.tb01131.x

Jiang, Z. (2018). Using the linear mixed-effect model framework to estimate generalizability variance components in R. *Methodology*, *14*(3), 133–142. https://doi.org/10.1027/1614-2241/a000149

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Book review: Technology-Assisted Language Assessment in Diverse Contexts: Lessons from the Transition to Online Testing During Covid-19

K. Sadeghi (Ed.), *Technology-Assisted Language Assessment in Diverse Contexts: Lessons from the Transition to Online Testing During Covid-19*. New York: Routledge, 2022. 274 pp. ISBN 9781003221463

Reviewed by: Tomohito Hiromori, *Meiji University, Japan* and Hassan Mohebbi, *European Knowledge Development Institute, Turkey*

The COVID-19 pandemic has had a profound impact on people worldwide on an unprecedented scale, leading to significant changes in national economies, social life, and education (for recent systematic reviews, see Betthäuser et al., 2023 and Tang, 2023). As many schools closed and transitioned to online education, there has been a rapid increase in demand for technology-assisted language assessment, with various forms of online language assessment now available. However, these new language assessments also present radical challenges. Accordingly, education stakeholders require evidence-based recommendations for practice, informed by research conducted across diverse educational