G. Fulcher & F. Davidson (Eds.) *Routledge Handbook of Language Testing*
New York: Routledge. 2012. Pp. 536.

Handbooks give us a means of grappling with shifting disciplinary boundaries, evolving key concepts and burgeoning research agendas. Importantly, authoritative states-of-the-art, as they are typically described, not only sanction boundary shifts but also set future direction. In this Routledge contribution to the growing number of handbooks on language testing and assessment, Fulcher and Davidson observe that young disciplines such as Language Testing have short collective memories; they lack replication and 'do not build on what is known' (p. 2). Their collection has the potential to contribute powerfully to collective memory by building a coherent story of evolving disciplinary facets.

The strength of this collection is the renegotiation of boundaries through a complicated interweaving of the practical and theoretical aspects of the field. This can be seen in nine broad themes which structure the 536-page volume: *Validity, Classroom assessment and washback, The social uses of language testing, Test specifications, Writing items and tasks, Prototyping and field tests, Measurement theory and practice, Administration and training, Ethics and language policy.* The majority of chapters within these themes follow the editorial blueprint, starting with a historical overview, including a critique of the key concepts, recommendations for practice, and finishing with a section on future directions and a brief annotated reading list.

The editors state that they have placed *Validity* at the beginning of the book because of its centrality to language testing design and research. Chapelle provides an excellent historical introduction to the prevailing metaphor of argument-based validity. The concept is then outlined by Kane, who finishes by pointing out that although the conceptualization of validity has come a long way, there is still a great deal to be done in terms of applying the abstract, general models to specific contexts. This is an important call, which resonates through other chapters in practical applications (e.g. 'Pre-operational testing' by Kenyon and MacGregor) and measurement concepts, (e.g. 'Reliability and dependability' by Jones). It is interesting that this theme is immediately followed by topics which do not readily fit within the rhetorical logic view of validity, such as classroom assessment and dynamic assessment. Indeed, Turner raises the question of whether validity requires reconceptualisation when it is applied to classroom-based assessment. The problem of how validity is understood is further elaborated by Brown in her later chapter on 'Interlocutor and rater training' where she observes that perspectives on

validity differ 'across the divide of the assessment context': small-scale teacher-led versus large scale, standardized assessments (p. 416). The various perspectives throughout the volume show that although validity is understood as central to the process and place of assessment in society, its application to specific contexts and types of assessment is very much work in progress.

Given the primacy of validity in the handbook, it may seem strange that the nature of language and communication appears to be absent from the contents. A significant editorial decision was to eschew the typical communication model/component breakdown (e.g. reading, writing, speaking, listening, vocabulary, grammar, etc.) that has become entrenched in large-scale language tests and test preparation coursebooks as well as academic books, series and courses about language testing. In fact, only one chapter, 'Piloting vocabulary tests' by Read, uses a particular language component as the focus and this is as much about developing item formats as it is about vocabulary. Surveying other handbooks and similar language testing/assessment reference works suggests that the field is at a crossroads in terms of the conceptualisation of construct. Contemporary volumes such as *The Cambridge Guide to Second Language Assessment* (Coombe, Stoynoff, O'Sullivan & Davidson, 2012) and the *Companion to Language Assessment* (Kunnan, 2013) maintain the separation of language components, although in the latter, the list is expanded to include literacy, pragmatics and integrated skills among others. In her introduction to the *Language Testing and Assessment* volume in the *Encyclopedia of Language and Education,* Shohamy (2008, p. xvi) explains that in seeking to represent the multiplicity of languages and the diversity of contexts, language is considered from multiple perspectives rather than clear-cut skills. Perhaps the Routledge handbook bears witness to the conflation of the 'what' of language testing (the construct) and the 'how' (the method), as Shohamy describes it. Certainly, the nature of test constructs is discussed usefully via various methodological routes, for example, defining the construct in Plakans' essay on 'Writing integrated items', the nature of test-elicited speaking in Ross's chapter on 'Claims, evidence, and inference in performance assessment', the effect of multimedia on the construct in Sawaki's chapter on 'Technology in language testing', Aviation English in Moder and Halleck's chapter on 'Designing language tests for specific social uses' and the question of construct in Hudson's chapter on 'Standards-based testing'. It could be argued that the spread of test construct across multiple facets of the language testing enterprise is a necessary effect of a unified concept of validity as outlined by Messick (1989), which includes social consequences in the construct validity purview.

For the test developer, there are valuable practical and technical summaries in a useful array of topics, with strong links between aspects of test design and

claims/inferences. The chapters on measurement do particularly well in balancing a satisfying breadth and depth with accessibility. Overall, the volume is well-proportioned in terms of attention to the theoretical, practical, social and technical aspects of language testing and the focus on future directions in each chapter should contribute constructively to the course of the field. As test developers and researchers, we think this excellent collection serves the collective memory well.

Review by Susy Macqueen and Rohan Chandran

The University of Melbourne

References

Coombe, C., Stoynoff, S., O'Sullivan, B. & Davidson, P. (Eds.). (2012). *The Cambridge guide to second language assessment*. Cambridge: Cambridge University Press.

Kunnan, A. J. (Ed.). (2013). *The companion to language assessment*. Chichester: John Wiley & Sons.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3 ed., pp. 13–103). New York: American Council on Education and Macmillan.

Shohamy, E. (2008). Introduction to Volume 7: Language testing and assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Language testing and assessment: Encyclopedia of language and education* (Vol. 7, pp. xiii–xxii). New York: Springer.