

C. Coombe, P. Davidson, B. O'Sullivan & S. Stoyhoff. *The Cambridge guide to second language assessment*. Cambridge: Cambridge University Press. 2012. Pp. 321.

'Assessment', according to Davies et al. (1999), encompasses "an evaluation through assessment of learner(s), and assessment of the whole teaching operation including curriculum, methods, resources, plans and staffing" (p. 11). This term is likely to be used interchangeably with 'test(ing)', but it would seem, at least, that 'assessment' is used in a broader sense than 'test(ing)' (Bachman, 2004; Brown & Abeywickrama, 2010; Clapham, 2000). *Second Language Assessment* clearly reflects what should be considered in assessment with its comprehensive coverage of key issues beyond the scope of 'test(ing)'. This volume is designed mainly to enhance the understanding of assessment or "assessment literacy" (p. 20) of those involved in language teaching and assessment in the classroom, who spend one third to half of their time in assessment-related jobs, as mentioned in Chapter 2 by Coombe, Troudi and Al-Hamly. As such, assessment literacy is a crucial part of language teachers' professional work, which also leads directly to benefits for their students. The major contribution of this book is thus to enhance the benefits of these stakeholders in second language assessment by raising awareness of considerations in assessment for language teachers.

The 35 chapters of this collection are structured to introduce readers to a comprehensive range of issues related to second language assessment and to promote understanding of these issues. The topics covered in the book include the purposes and principles of several types of tests (e.g., proficiency, placement, classroom assessment, criterion-referenced tests, norm-referenced tests), historical backgrounds of each topic, their current considerations, and available existing resources for defining test constructs and developing instruments (e.g., CEFR). The editors have extended the scope to more practical aspects of assessment such as utilisation of electronic resources (e.g., online tests, software), supporting assessment technically, and administrative matters. The book also addresses several significant issues such as validity, reliability, and test consequence, which need to be considered in both test development and in its use for as long as the assessment instrument is operational. Throughout the book, these topics are arranged to introduce principal issues and theoretical aspects first, and then to promote awareness among readers about how to apply these key issues to actual administrations of assessment in their classrooms. The chapters are contributed by various authors; however, readers should be aware that the chapters all have the same objective:

highlighting engagement of both teachers and students in assessment; stressing the link between assessment and instruction and learning.

The comprehensive coverage of topics and the arrangement of the components of each chapter throughout the book allow readers to engage with the discussion of the topics from multiple angles. However, this may not happen if readers address each chapter independently, as at first glance of the table of contents, it may seem to be simply a readable collection of several topics contributed to by different authors. Each chapter indeed keeps its independence to some extent, and the previous chapters are not a pre-requisite for the subsequent chapters. However, it is more useful and interesting to read the prior chapters, as they foreshadow upcoming issues.

For instance, at first glance the chapter title, *Test-Taking Strategies* (Chapter 11 by Cohen), seems only to refer to the test-taker internal process and to be independent from the previous chapter, *Test Impact and Washback: Implications for Teaching and Learning* (Chapter 10 by Chen & Curtis). But these chapters are associated by illustrating the way that tests can affect test-taker strategies when working on a test. Addressing test-taker strategies from the viewpoint of potential effects of the testing, as well as from learners' cognitive processes, serves to deepen the discussion on why test-taker strategies need to be addressed and how they are characterised. Furthermore, test strategies can also be seen as linked to the issues addressed in another previous chapter, *Validity in Language Testing* (Chapter 3 by Akbari). Employing certain strategies can be a part of the test construct but at the same time, other strategies can be identified as external factors unnecessarily affecting assessee performances, namely, construct-irrelevance (Messick, 1989) especially with "test-wiseness strategies" (Chapter 11 by Cohen, p. 97). Thus the notions, key words, questions and challenges discussed in the previous chapters empower readers to expand their perspectives of the discussion in subsequent chapters.

This book is a guide for, most likely, teachers of second languages rather than experts or advanced level students working on research related to language testing and/or large-scale test developers, since the authors of each chapter avoid using jargon and introducing critical views of previous related research and debates originating from empirical studies about the topic. So then, how do the comprehensiveness of issues and the readability of the text benefit advanced readers? The issue explored in each chapter may not be deep enough to challenge more advanced readers in language testing. However, the value of this book is not lost for those more advanced readers. Another way to approach this book is for readers to relate each chapter and discussion to validity evidence (Messick, 1989), which is perhaps more of a concern for test

developers. *Validity in Language Testing* (Chapter 3 by Akbari) provides a picture of the historical development of the notion of and current approaches to validity. However, unlike *Language Testing and Assessment* (Fulcher & Davidson, 2007), not much space is given to validity throughout the book. Further, the term 'validation' is not used in the book. But in each chapter, the issues addressed can also be viewed as pieces of evidence to allow for partial or overall inferences about test score interpretation and test use. Considering the primary aims of this book and language teachers as the target readers, further discussion of validation may not be necessary. However, classroom teachers should at least pay attention to how students' task performances are scored and the impact and washback of the assessment on students' learning as they engage in developing and using assessment instruments. Tracing what aspects were involved in the assessment and investigating possible evidence in each phase of validation will serve to prevent over- or under-estimation of students' ability. Furthermore, it will identify more appropriate treatments for their students, in addition to justifying their interpretation of the test score.

Based on the above suggestion, the contents of the book can be reframed from the view of validity as evidence for argument-based approaches to validation (e.g. Chapelle et al., 2008; Kane, 2006, 2012). Needs analysis, mentioned in Chapter 21, *ESL Needs Analysis and Assessment in the Workplace* (Greenberg) and other chapters addressing assessment of the four skills of second language, grammar, vocabulary and L2 pragmatics, is concerned with defining the target test construct depending on the test purposes. This corresponds to the inferences of evaluation, explanation and extrapolation (Xi, 2008), which are related to the test construct. Major advantages of existing electronic resources for language assessment are also addressed in this book, but as pointed out in Chapter 30, *Computerized Language Assessment* (Davidson & Coombe), there are drawbacks associated with these resources, which may lead to possible construct-irrelevant factors (Messick, 1989). The scope of the book also expands to administrative issues, which the editors refer to as an "often-neglected aspect of assessment" (p. 289) but which should be considered to maximise the benefit and minimise the burden of teachers in assessment administration.

Marking Assessment: Rating Scales and Rubrics (Chapter 34 by Green & Hawkey) can be seen as approaching the inferences of evaluation and generalisation. These inferences are addressed in other chapters, for instance, *Assessing Writing* (Chapter 25 by Weigle) and *Assessing Speaking* (Chapter 27 by O'Sullivan), that provide considerations for designing or selecting rating scales or scoring rubrics and for rater selection and training raters. The generalisation inference is also addressed more explicitly with the score reliability issues summarised in *Principal of Language Assessment* (Chapter 4 by Farhady) and in *What Teachers*

Need to Know About Test Analysis (Chapter 12 by Brown). Several other chapters deal with defining target test constructs, test taking strategies, needs analysis, potential construct irrelevant factors, engagement of peer students and teachers in the assessment the process; investigations of which can allow for explanation and extrapolation inferences. In particular, application of assessment by stakeholders themselves (teachers, peers, the students themselves) as proposed in several chapters, and utilisation of portfolio assessment introduced in *Alternative Assessment in Language Education* (Chapter 17 by Coombe, Purmensity & Davidson) are expected to provide stronger evidence to enhance the extrapolation inference. The consequences of the test addressed in *Test Impact and Washback: Implication for Teaching and Learning* (Chapter 10 by Chen & Curtis) are concerned with the utilisation inference.

This book is aimed at the enhancement of assessment literacy of teachers; it is not explicitly a guide for validation of their assessment. The comprehensive coverage of topics in a useful array and readable descriptions are strengths of this book, but it will not fully challenge readers who seek more thought-provoking discussions. Readers will have their own way of maximising the value of this book from the perspective of their own profession. I attempted to do so by making my own connections across chapters, which strengthens the discussion by associating the notions, considerations, questions and so forth addressed in other parts of the book. Each independent chapter completes the discussion by itself, but readers can also broaden their appreciation of the issues by seeking these stronger associations with the issues presented in other chapters. Although not mentioned by the editors, I also reframed the contents of the book as sources for seeking pieces of evidence for the inferences in an argument-based approach to validity. This book is by no means simply a collection of several related topics, nor is it a glossary of language assessment. It provides readers with various ways to enrich and refine their knowledge and perspectives in second language assessment and language testing, thus achieving the ultimate goal of the book, which is to raise awareness of considerations in assessment for language teachers.

Review by Naoki Ikeda

The University of Melbourne

References

- Bachman, L. F. (2004). *Statistical analysis for language assessment*. Cambridge: Cambridge University Press.
- Brown, H. D. & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). White Plains, NY: Pearson Education.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York, NY: Routledge.
- Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics*, 20, 147-161.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London: Routledge.
- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Westport, CT: Greenwood Publishing.
- Kane, M.T. (2012). Articulating a validity argument. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 34-47). New York: Routledge.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan.
- Xi, X. (2008). Methods of test validation. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.) (Vol. 7: Language Testing and Assessment). New York: Springer.