

***Exploring Language Assessment and Testing:  
Language in Action***  
**Anthony Green (2021) (2<sup>nd</sup> edition)**

**New York: Routledge. Pp. 306  
ISBN: 9781138388772 (hbk)  
ISBN: 9781138388789 (pbk)  
ISBN: 9781003105794 (eBook)**

*Reviewed by Lynda Taylor*

‘The one thing that is worse than no test is a bad test.’<sup>1</sup> This remark was made in March 2020 by Professor Chris Whitty, then Chief Medical Officer (CMO) for England, the United Kingdom government’s Chief Medical Adviser, and head of the public health profession (Schraer 2020; BBC News 2020). He was speaking during the early weeks of the lockdown in the United Kingdom as the Covid-19 pandemic was beginning to take hold and as the media and general public were eagerly discussing a range of take-at-home antibody tests coming to market. Professor Whitty’s logic was clear: not all tests are reliable (i.e., sufficiently accurate), and not all tests are valid (i.e., appropriate for use). A decision based on faulty test evidence can lead to a worse outcome than if no test were available in the first place. Tests can sometimes grant a false sense of certainty, and false certainty in a high-stakes situation like medicine risks having life-threatening consequences.

Just a few months later, a ‘critical incident’ emerged involving testing and assessment in the UK education sector (BBC News 2020). In August 2020, a statistical algorithm was applied to school-based teacher assessments which had been gathered in England and Wales to determine A-Level results for 18-year-olds in the absence of final data from the usual nationally organised examinations, cancelled that summer due to Covid. The negative impact of the

algorithm's use met with public outcry, and the policy was swiftly reversed so that final results were instead based on the original teacher assessments. A similar problem with GCSE results for 16-year-olds was only narrowly avoided. At the same time, UK universities grappled with the decision of whether or not to hold in-person exams, and many professional tests moved to an online environment with test security concerns addressed through remote proctoring.

Extraordinary times, such as times of war or global pandemic, often call for extraordinary measures, and decisions made in haste or under pressure may not always lead to the best outcome. But how do policy makers, health professionals, educators, and the general public determine what is a good test and a fair assessment system in the first place? What specific knowledge and skills might be necessary to determine the quality and fairness of a test? People are comfortable with making broad judgements about fairness and justice in many areas in public life (consider the reaction to a football referee whose decisions are seen as biased towards one team or another). There is a curious lack of discussion around tests and testing processes, however. This is curious because we all encounter multiple testing regimes throughout our lives: school exams, driving tests, medical tests, job interviews, citizenship tests (for some), and professional certification exams. Testing is everywhere and yet, apart from isolated examples, the more technical intricacies of testing and assessment are rarely scrutinised by the news media and rarely the topic of public discussion. Moreover, despite the fact that testing and assessment play a central role in society and in our social interaction, we can be surprisingly lacking in knowledge and understanding of how testing and assessment work – for good and for harm.

In the specific field of educational assessment, recent research has sought to understand the nature of 'assessment literacy', defined broadly as the skills, knowledge, and abilities that different stakeholders in education and society (e.g., teachers, university admissions officers, policymakers) need in order to carry out assessment-related activities (Taylor, 2009; Taylor, 2013; Kremmel and Harding, 2019). It is self-evident that certain groups need to know a lot about testing and assessment: those who *create* the tests, those who *research* the tests, and those who *use* the tests. Recent years have seen significant growth in the research literature in this area, along with various attempts to communicate the theory and practice of assessment beyond just the research community. Much of the research concerns the testing of languages, mainly but not exclusively English for speakers of other languages (ESOL). Language teachers in particular are one constituency who need to develop sound understanding and practical skills in testing and assessment, partly for their own professional practice in the classroom context, but also in order to understand and evaluate the

nature and role of largescale examinations and their impact in a wider socio-educational context. Postgraduate students in applied linguistics and language education are another group who can benefit from improving their assessment literacy, for both their academic studies and their future career.

Taylor and Green (2020) describe the dissemination over the past two decades of language testing and assessment (LTA) knowledge and practice through a growing number of publications, including academic monographs, edited collections, technical handbooks and testing encyclopaedia. *Exploring language assessment and testing: Language in action* by Anthony Green, first published by Routledge in 2014, is a good example of just such a publication which aims to provide a practical introduction to testing and assessment for postgraduate students in language education and applied language studies.

In 2021, Green published the second edition of his volume, updating his earlier work and supplementing it with a number of additional features of theoretical interest and practical value. These include hands-on activities for the reader within each chapter, comprising questions for personal reflection and numbered tasks for completion. Later in the book, a helpful section contains extensive author commentaries on the in-chapter tasks, providing detailed answers where appropriate and further insights into the issues under consideration. A particularly valuable new feature towards the end of the volume is the addition of a 15-page glossary of key terms. Each of the more than 150 items included in the glossary is highlighted in bold on first appearance in the main body of the book and can therefore be easily referred to for clarification or further explanation in the glossary section. Some measure of cross-referencing between similar or contrasting terms within the glossary is also included. A further enhancement to this new edition is the 6-page 'Further Reading' section with helpfully annotated references organised under a number of topical subheadings. The revised edition also provides updated coverage of some key theoretical and technological developments (e.g., the trend towards integrating assessment with teaching and learning, the steady growth of e-assessment).

In terms of its overall structure, *Exploring language assessment and testing* broadly subdivides into three main parts. Part I is intended as a practical resource to help teachers, students, trainees, and anyone interested in language education to understand more about the actual practice of language testing and assessment. The first two chapters set out the core functions of language assessment. Chapter 1 outlines what is meant by language assessment and its place in language education, while Chapter 2 briefly considers some of the reasons for assessing language abilities. Chapter 3 presents a practical and widely applicable approach to building quality assessments for use in the classroom or institutional testing

programmes, while Chapter 4 describes the qualities of useful assessments and how to evaluate them. A companion website for the book ([www.routledge.com/cw/rial](http://www.routledge.com/cw/rial)) provides supplementary resources to accompany Chapter 3, including documents on *Illustrative Assessment Cycles* and *Statistical Tools*. (NB: I was unable to access the website component for the volume, so it is not included in this book review).

Part II of the volume introduces the reader to some relevant theoretical insights and shows how the processes described in Part I are used in developing tests and assessments of different language skills. Chapters 5 and 6 offer accessible guidance on designing and developing effective assessments and preparing appropriate materials for both receptive and productive skills. Chapter 5 focuses on techniques for designing and scoring assessments of reading and listening, while Chapter 6 focuses on writing and speaking in both productive and interactive modes. The various techniques are presented and evaluated from a broadly sociocognitive standpoint which brings together a social view of language in use, recognising how the language we use is shaped by social expectations and conventions, and psychological insights into the mental processes involved in conveying ideas and making sense of what we see and hear. A sociocognitive approach to language test development and validation (Weir, 2005) has become widely established and accepted by the professional community over the past 10 years, and the approach has proved an effective framework within which language testing practitioners can develop their tests and make defensible claims about the quality of their assessments (Chalhoub-Deville and O'Sullivan, 2020). Supplementary materials on *Task Types* for both these chapters are available on the companion website.

Part III introduces the reader to the study of language assessment as a domain in its own right. It begins by tracing its emergence as a distinct field of academic inquiry before going on to outline some developing trends and highlight areas of controversy and debate. Chapter 7 closes with some thoughts on future trends and challenges in the field. The aim of this section is to contextualise and deepen understanding of the practices, techniques described in Part I and Part II, as well as to open up avenues for the reader to explore through further reading and study. A *Links* button on the companion website points the way to additional resources. The volume ends with a list of the references mentioned throughout the book and a final index combines entries for subject matter and authors.

*Exploring language assessment and testing: Language in action* is an accessible and engaging introductory text for anyone wishing to understand more about the field, but especially for those in language education contexts, such as classroom teachers, language teacher educators, and postgraduate students

in applied language studies. A distinctive, and particularly welcome, feature of the volume is the way it is structured to begin with the practical activity and experience of language assessment and moves on only later in the volume to the more theoretical, as this becomes relevant and necessary. This ‘back-to-front’ structure or ‘practice to theory’ approach, which is characteristic of the Routledge Introduction volumes, helps make what can sometimes appear to be quite a specialised and daunting academic domain into a much more accessible and navigable field. It makes the book highly suitable for use within a postgraduate education setting, such as a taught course in applied linguistics or a language testing and assessment module. Readers are encouraged from the outset to reflect on their own personal experiences of test-taking or test-making, and the beliefs and attitudes that are associated with those experiences. They are invited to review their own knowledge and understanding on a range of topics and to apply new information and insights offered by completing a variety of practical tasks and exercises. Some of these tasks and exercises lend themselves to use in class for pair or group work, while others could usefully serve as activities for homework or term assignments. One of my academic colleagues in the United States, for example, took the PRICE acronym (see Chapter 3, p. 21) and set her postgraduate students the task of producing a poster to communicate the key principles represented by PRICE (Planning, Reflection, Improvement, Cooperation and Evidence). She was delighted when some of her students unexpectedly produced a short but highly engaging video poster in response, cleverly integrating their learning about language testing with their creative skills and artistic talents!

As well as being directly relevant as part of a taught course in language education or applied linguistics within a postgraduate teaching and learning context, the volume is a valuable introductory resource for other constituencies. The book’s content and approach are equally accessible to groups such as language teacher educators and in-service language teachers who may wish or need to update their knowledge of assessment as part of their ongoing professional development or improve their skills and confidence when faced with designing assessments for their own context. Another group of individuals who might benefit from this volume are novice test writers working for testing organisations. It would not be difficult to imagine the book being used as the basis for individual self-study, or even for a departmental or group-based self-study initiative by any of these constituencies.

Following a secure grounding in the essential principles and practices of language testing and assessment in Parts I and II of the book, readers are invited in Part III to broaden their theoretical horizons and to consider how what they

have learned thus far sits within a much larger historical, linguistic, educational, and sociopolitical context. This section is both interesting and informative as it helpfully overviews some of the issues that have occupied language assessment researchers and shaped the practice of language assessment professionals over more than a century. They include differing views on the nature of language proficiency; the evolving relationship between teaching, learning, and assessment; the growing role of technology; and the increased attention given to ethical considerations. Although Parts I, II, and III of the book are coherently linked in terms of a logical progression from practice to theory, they also stand discretely to some extent and can be used flexibly. The final Part III could even be viewed as follow-up (rather than essential) reading for those who are interested.

The main virtue of this volume, in my opinion, is its accessibility and relevance to a wide range of stakeholders – postgraduate students, teachers, teacher educators, and others – who are introduced to the field of language testing and assessment using a practical, useful, and positive approach which encourages them to think about assessment issues in a personally relevant way for their own context of operation. This resonates with a growing trend nowadays towards acknowledging and valuing local factors in assessment contexts and initiatives, which should help to shape decisions and outcomes (Weir, 2020). The author's writing style is clear and encouraging, and a wealth of analogies and illustrations sprinkled throughout the book helps make sometimes difficult concepts more readily understandable (e.g., discussion of the quality of reliability in Chapter 4). One small exception is the author's use of the word 'caboose' on page 44 as an image or metaphor, without providing any gloss or explanation for those for whom this may be an unfamiliar term!

Despite its many positive features, one or two shortcomings in the volume are worth noting in case these can be addressed in any future revision. In terms of content coverage, there are a couple of areas that I feel could have been fleshed out rather more fully than was the case in the book. One is the area of 'specific purpose testing' which is mentioned in passing at several points in the volume but might have merited a fuller discussion in some sections? Green is open and transparent at the outset that his primary focus is on language assessment in educational contexts, so it is perhaps not surprising that the role of testing for professional or other specific purposes should receive relatively little attention. Nevertheless, given the important role nowadays of language testing for professional certification / registration and in other workplace contexts (Knoch and Macqueen 2019), whether for internationally trained healthcare workers, for the travel aviation industry (pilots and air traffic controllers) or for teachers and teaching assistants, it seems a pity that these areas were not more fully explored,

especially in relation to the development and use of indigenous scoring criteria (which is not mentioned as a term or included in the glossary). Furthermore, given the controversial role of language testing in the face of growing global migration and matters of citizenship, this area would seem to be particularly relevant and timely for postgraduate students and early career professionals who may be looking to engage in language testing and assessment activities in the future.

A second content area which seems somewhat neglected in the volume is any discussion of assessment accommodations for test takers (or assesseees to use Green's term) who have special needs or requirements, either due to short-term or permanent disability (or even for religious or cultural reasons). This area is referred to in passing in no more than a couple of sentences, despite the fact that such test takers are now known to account for a significant proportion of the testing population. According to the International Dyslexia Association (2012), for example, the number of students living with dyslexia and coping with its implications for their learning and assessment experience may be as high as 15%. In terms of their experience of teaching and learning, learners with disabilities for the most part undertake their primary, secondary, and tertiary education in the mainstream educational sector, often completing their studies through routine school-based testing and national examination programmes. Despite its practical focus, the volume offers readers (such as teachers and novice test developers) little practical guidance on the sorts of special arrangements or adaptations to test materials that are possible and desirable for such students. Chapter 5 and Chapter 6 might have been an ideal place in which to include short additional subsections addressing this for the different language skills covered. Taking a broader view, this seems to be an area which is routinely neglected or marginalised in much of the language testing and assessment literature, with relatively few books and even fewer journals in the field giving the topic the attention it merits. This seems curious in an age which is increasingly concerned with matters of equity, diversity, inclusion, and access, and it will hopefully be more equitably addressed in language testing books and journals in the future. Interestingly, a discussion around current trends in translanguaging, multilingualism, and assessment does find a place in Green's volume in Chapter 7, and this is to be welcomed.

This brings me to a final set of comments on certain presentational features of the volume. For the most part the text is set out in a clear and accessible format, with transparent signposting and guidance for the reader throughout. Tables and figures are generally easy to read and to understand, with only the occasional instance when the typeface is slightly too small or faint for comfort (e.g., pp.

84, 128, 174). Questions for personal reflection are clearly presented in plain, unshaded boxes, while numbered tasks are easy to see in shaded boxes, though it took me a while to realise that the TC within a small circle in the margin beside each task box signified ‘Task Commentary’ and relates to the set of task commentaries contained in Chapter 8! It is also not entirely clear why some of the discussion material in Chapter 7 is set within a box, namely, the lengthy discussion of two approaches to score interpretation on page 199, and the piece on translanguaging and assessment on page 223. The formatting seemed to suggest a particular status for these two pieces, but it was not clear what that was or why the material followed a different format to the other subsections in the chapter. Conversely, the use of this ‘boxed’ approach in Chapters 5 and 6 to summarise and highlight practical guidelines for assessment task design presented no such confusion. The References section at the end of the volume seemed comprehensive and complete, though Green (2013), which is mentioned a couple of times in the main volume, does not appear in the ‘References’.

Ideally, testing and assessment contribute to a society where quality data is gathered to inform sound decision making for the common good and where opportunities are awarded on the basis of talent, effort, and achievement rather than wealth or social class. They appeal to a common-sense approval of rigour and high standards. But tests, including language tests, can also be used as a convenient political tool to show that something is being done or to achieve policy objectives beyond what a test was originally designed to assess, so the sociopolitical role of tests within a sociopolitical context also needs to be understood (McNamara 2005). A broadly knowledgeable public is the best way to hold those who use tests for high-stakes decisions accountable to the principles of fairness and justice. Widening assessment literacy will only serve as a force for social good if those who create assessments and those use assessments are able to question and think critically about the connection between the evidence collected by a test, the result, and the ultimate decision. Anthony Green’s *Exploring language assessment and testing: Language in action* represents an important and timely contribution in the ongoing endeavour to promote language assessment literacy to a wider audience.

## Notes

1. The opening section of this book review was adapted from an online article co-authored by Professor Lynda Taylor and Professor Luke Harding in November 2020 for the United Kingdom Academy of Social Sciences. *The article was originally commissioned and published by the Campaign for Social Science as*



*part of its Covid-19 programme and can be freely accessed at: A testing time for testing: Assessment literacy as a force for social good in the time of coronavirus – Academy of Social Sciences (acss.org.uk).*

## References

- BBC News (August 2020) A-levels and GCSEs: How did the exam algorithm work? Retrieved on 6 September 2021 from: <https://www.bbc.com/news/explainers-53807730>.
- Chalhoub-Deville, M. and O’Sullivan, B. (2020) *Validity: Theoretical development and integrated arguments*. Sheffield: Equinox.
- International Dyslexia Association (2012) *Dyslexia Basics*. <https://dyslexiaida.org/dyslexia-basics-2/>
- Knoch, U. and Macqueen, S. (2019) *Assessing English for professional purposes: Language and the workplace*. London: Routledge.
- Kremmel, B. and Harding, L. (2019) Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly* 17(1): 100–120. Doi: <https://doi.org/10.1080/15434303.2019.1674855>.
- McNamara, T. (2005) 21<sup>st</sup> century Shibboleth: Language tests, identity and inter-group conflict. *Language Policy* 4(4): 351–370. Doi: <https://doi.org/10.1007/s10993-005-2886-0>.
- Schraer, R. (March 2020) Coronavirus: home tests not available ‘next week’. BBC News. Retrieved on 6 September 2021 from: <https://www.bbc.com/news/health-52035615>.
- Taylor, L. (2009) Developing assessment literacy. *Annual Review of Applied Linguistics* 29: 21–36. Doi: <https://doi.org/10.1017/S0267190509090035>.
- Taylor, L. (2013) Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing* 30(3): 403–412. Doi: <https://doi.org/10.1177/0265532213480338>.
- Taylor, L. and Green, A. (2020) The role of academic institutions in language testing research and consultancy. In L. Taylor and N. Saville (eds) *Lessons and legacy: A tribute to professor Cyril J Weir (1950–2018)* 175–206. Cambridge: Cambridge University Press.
- Weir, C. J. (2005) *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.
- Weir, C. J. (2020) Global, local or ‘glocal’: Alternative pathways in English language test provision. In L. I-W. Su, C. J. Weir, and J. R. W. Wu (eds) *English language proficiency testing in Asia: A new paradigm bridging global and local contexts* 192–225. Abingdon, Oxon: Routledge.

(Received and accepted 6<sup>th</sup> September 2021)