

Comparative Book Review

Language testing and assessment: Recent perspectives on research and practice

Christine A. Coombe Dubai Men's College, Higher Colleges of Technology,
United Arab Emirates
christine.coombe@hct.ac.ae

J. CHARLES ALDERSON, *Diagnosing foreign language proficiency*. London: Continuum, (2005). Pp. vi + 284. ISBN 0-8264-9388-2 (pbk).

GLENN FULCHER & FRED DAVIDSON, *Language testing and assessment: An advanced resource book*. New York: Routledge, (2007). Pp. vii + 403. ISBN 978-0-415-33947-6 (pbk).

MARGO GOTTLIEB & DIEP NGUYEN, *Assessment and accountability in language education programs*. Philadelphia, PA: Caslon Publishing, (2007). Pp. iii + 237. ISBN 0-9727507-7-0 (pbk).

ANTHONY GREEN, *IELTS Washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press, (2007). Pp. vii + 386. ISBN 978-0-521-69292-2 (pbk).

TIM MCNAMARA & CARSTEN ROEVER, *Language testing: The social dimension*. Malden, MA: Blackwell, (2006). Pp. xi + 291. ISBN 978-1-4051-5543-4 (pbk).

Few today would argue that tests and assessments are part of modern life. As the field increases in importance we are seeing more and more edited and authored volumes on this important topic. The books under review were chosen for four main reasons. First of all, they are perhaps the most current works to come out of the field of language testing and assessment. Second, the authors of these volumes are established authorities in the field and represent diverse global perspectives. Thirdly, these volumes all profess to deal in some way with the theory and research behind language assessment and the implications that this has on classroom practice. Finally, rather than being targeted towards those new to the field of language assessment, they are all intended for those with a more specialized interest in language testing and assessment.

Overview

In their volume, *Language testing and assessment: An advanced resource book*, Glenn Fulcher & Fred Davidson do just about everything they profess to do in the volume introduction. Content is organized in a unique template set out in the Routledge Applied Linguistics resource

book series, the purpose of which is to prepare readers for advanced study in core areas of the field. The volume is organized into three main sections, 'Introduction', 'Extension' and 'Exploration'.

The 'Introduction' sets out the key terms and concepts and extends the readers' knowledge through analysis and practical application. Some of the important content covered in this section of the volume, which consists of ten chapters, includes validity, test constructs, test specifications and design, scoring, fairness, ethics and standards. The 'Extension' section presents influential and seminal readings in testing and assessment by scholars such as Michael Canale, Merrill Swain, Lee Cronbach, Alan Davies, Liz Hamp-Lyons and Annie Brown, sets them in context and discusses their overall contribution to the field. Tasks are included after each excerpt encouraging the reader to apply the content to their own teaching context. The aim of the 'Exploration' section is for the reader to build on knowledge gained in the first two sections. The content of this section is varied and the activities are more open-ended and student-centered with tasks designed so that students can work individually or in groups.

The great strength of the Fulcher & Davidson volume is that it not only presents existing research and reflects on the state of the art in language testing and assessment but also presents a new approach that brings together testing practice, theory, ethics and philosophy. To this end, the authors contribute a new term to the jargon of the field, 'effect-driven testing', which they define as a view of test validity that is highly pragmatic. They claim to emphasize the outcome of testing activities and as such have let their own beliefs about the field guide the unique structure of the volume.

Another strength of this volume is its flexibility. Readers can progress through the sections in their entirety in a lockstep fashion starting with 'Introduction' and then proceeding to 'Extension' and finally 'Exploration'. Alternatively, readers can focus on individual sections (i.e., validity, writing items, test specifications, etc.) and then read corresponding content from the 'Introduction', 'Extension' and 'Exploration' sections.

The Margo Gottlieb & Diep Nguyen volume entitled *Assessment and accountability in language education programs* is situated firmly in the North American context and has as its aim to 'describe how we developed and implemented our assessment and accountability model with administrators and teachers in language education programs.' (p. x). The volume is divided into eight chapters which each begin with a chapter overview, some guiding questions and key concepts that are presented therein. At the end of each chapter the authors have included several 'Questions for reflection and action' for use with pre-service teachers, graduate students or individually on a self-access basis.

Early on in the volume Gottlieb & Nguyen present their BASIC (Balanced Assessment and Accountability System, Inclusive and Comprehensive) model. They then introduce five assessment principles that lead to strong language education programs and serve as the foundation for the BASIC model. They go on to address the components of the model and its distinguishing features. Throughout the volume, the authors guide readers through the multi-phase process of establishing a global framework for assessment, determining the specifics of their plan, implementing, reviewing, and revising it. The book concludes with the benefits of a comprehensive assessment framework. Chapter 4 was perhaps the most interesting as it introduced something new to me in the assessment literature: the pivotal portfolio.

This is a hybrid of both the working and the showcase portfolio. It is different from the traditional portfolio in that it follows the student for the length of the student's participation in the language education program and contains evidence gathered by the teacher of essential student learning and achievement. This chapter also includes components and characteristics of the pivotal portfolio as well as an explanation of how these assessments are used to measure students' language and academic growth. The ways in which educators can use the pivotal portfolio as instructional, evaluating and reporting tools are also discussed. Later content in the volume explores issues of assessment and accountability in two distinct US educational contexts, that of dual language and transitional bilingual programs. The authors follow two teachers, Mrs. Hernandez and Mrs. Meyer, a dual language teacher and a dual language program coordinator, respectively, as they recount their experiences using evidence from the pivotal portfolio to drive decision making at the classroom and program levels. A later chapter shares the perspectives of two transitional bilingual education program teachers and how they use information from the pivotal portfolio to help students adapt successfully to the English dominant school environment. Accountability is to the fore in chapter 7, which addresses the accountability that is internal to schools, districts, or programs. The authors highlight how information from evaluation can be useful in determining the effectiveness of language education programs. The final chapter of the volume reviews the critical elements of the BASIC model and shares lessons learned from the implementation of this model in a US school district.

A clear strength of this volume is the clarity in the presentation of the information and the reader-friendly graphs and charts. Despite the fact that it is based on a US context, I believe that educators in any country can benefit from this book. At a time when educators experience the competing demands of accountability, program improvement and appropriate classroom instruction, this volume presents complex assessment issues from a variety of diverse perspectives and offers solutions at all levels.

In *Language testing: The social dimension*, Tim McNamara & Carsten Roever attempt to reawaken the consciousness that language testing is a social practice and help readers come to grips with the theoretical and practical difficulties involved in testing the social side of language (p. 8). It is the fifth volume in the Language Learning Monograph Series, which reviews recent findings and current theoretical positions, presents new data and interpretations and sketches interdisciplinary research programs. In the introduction to this volume, McNamara & Roever set out a convincing case for the importance of the social dimension of tests and the effects of language testing on society.

The authors have framed their volume in relation to contemporary approaches to validity theory as first expounded by Messick (1989) and argue that the social role and function of tests is conceptualized within them. Chapter content is interesting and varied, focusing on such important issues as testing the social aspects of language competence and looking at how traditional psychometric approaches to testing have handled certain aspects of the social context of assessment. As a language testing professional, I found the most informative chapters to be chapters 5 and 6, which describe the social embeddedness of testing in the form of fairness reviews and ethics and language tests and their relationship to social identities. Particularly valuable was the insider perspective on how major educational testing boards such as Cambridge ESOL or ETS view

fairness. Chapter 6 examines the effects of language tests on individuals and societies and provides an informative historical perspective. The concluding chapters of the volume present information on the pervasive use of tests in educational settings and the focus on standards.

The focus on different and varied global testing systems is very much in evidence in this work. This practice is much appreciated and I feel sets this volume apart from much of the work done on language testing and assessment, which seems to focus on issues from one part of the world. A major strength of this volume is that it has the potential for opening up a dialog on the social and political consequences of language testing.

The last two volumes under review center around work done on two important language tests, the IELTS and DIALANG. Anthony Green's *IELTS Washback in context: Preparation for academic writing in higher education* presents findings of a study to investigate the washback effect of the IELTS writing subtest on English for Academic Purposes (EAP). The research reported in this volume examines IELTS preparation courses alongside broader programs designed to develop the academic literacy skills required for university study. Green's volume is the twenty-fifth in the Cambridge Studies in Language Testing series and the fourth in the series on the under-researched area of washback and impact. Despite its narrow focus, this volume adds much to the language testing and assessment literature. Results of Green's research have informed the subsequent development of the IELTS examination, most particularly that of the writing module. It has confirmed the need for greater accessibility to information on the meaning of band scores, resulting in the release of a DVD package 'IELTS Scores Explained' (Weir & Milanovic 2007: xi). The content is what might be expected of a volume based on a major research project: a literature review, pilot studies, instrumentation, results, analysis and implications for the various stakeholders. A highlight was the comprehensive review of the literature on washback and on academic writing. As a language teacher preparing students for the IELTS exam, the information on strategies for IELTS success and the time needed for score gain was particularly valuable.

Charles Alderson's *Diagnosing foreign language proficiency* addresses the need for tests that can diagnose the strengths and weaknesses in learners' developing foreign language proficiency. This volume presents the rationale for, and the research surrounding, the development of DIALANG, a suite of internet-delivered diagnostic foreign language tests funded by the European Commission. The author's use of the term 'diagnosing' in the title stems from the belief that despite the prevalent use of the word in discussions of language education and applied linguistics, very few truly diagnostic tests actually exist.

The book begins with a discussion of how diagnostic testing might most appropriately be developed and why it has thus far been neglected. The main focus of the volume is the presentation of DIALANG, and the author describes the diagnostic system developed for this project and explores the research results to date. Through DIALANG, Alderson helps us understand the notion of diagnosis, to explore the limits of what is known about this concept and to set an agenda for future research. With this volume, the author has contributed vastly to another under-researched area of the field – that of diagnostic testing – through an extremely thorough review of the literature on diagnosis and diagnostic testing in foreign/second language education as well as in other disciplines such as general education and reading. His introduction to DIALANG in chapter 3 is couched in several issues that are pertinent to diagnostic testing. Subsequent chapters center on the history of the test, pilot

studies, standard setting, sub-tests (i.e. listening, reading, writing, grammar, vocabulary) and item formats. A unique feature of this test is the role of self-assessment, and chapters 8 and 16 focus on its role in DIALANG and experimental findings.

Analysis and evaluation

Perspectives on research

A key issue in the area of language testing and assessment is the theoretical construct of assessment as well as the need for empirically-based assessment models. All five volumes score highly on these criteria. As Fulcher & Davidson put it in their volume description, ‘language testing is all about building better tests, researching how to build better tests and in so doing, understanding better the things that we test’ (p. xix). They make the point that although research might not be explicit in some parts of their volume, it is implicit throughout. Moreover, the creative drive of language testing makes it a research enterprise. Research is also at the fore of *Assessment and accountability in language education programs*, where Gottlieb & Nguyen draw on empirical evidence to promote their BASIC model. They describe research-based principles of assessment that help define the nature, role, and uses of different kinds of data that might be acquired through formal and informal assessments. McNamara & Roever provide a detailed account of the research done in the area of social dimensions of language testing and devote an entire chapter to future research initiatives they feel should be addressed. Similarly, Alderson and Green present a thorough examination of the research that has been conducted on two internationally-known language tests.

Perspectives on practice

Although they all claim to relate theory to practice in some way, some of the volumes do it more effectively than others. The two most teacher-friendly volumes are the Gottlieb & Nguyen and the Fulcher & Davidson volumes in that order. Even though the target audience for the Gottlieb & Nguyen book is administrators and teachers, I feel that it is very accessible for the average classroom teacher. This is helped by the initial features of the chapters ‘Chapter overview’, ‘Guiding questions’ and ‘Key concepts’, and those at the end, ‘Questions for reflection and action’. In *Language testing and assessment: An advanced resource book*, Fulcher & Davidson have essentially the same features interspersed throughout but they are not as easy to identify as those in the Gottlieb & Nguyen volume. That said, the Fulcher & Davidson book is perhaps the most versatile, as readers have the choice of how to proceed through its chapters. The format, in three sections (‘Introduction’, ‘Extension’ and ‘Exploration’), is a major asset but it will take readers a while to get used to this approach. Much practical information can be gleaned from the Alderson volume on DIALANG and the Green volume on the IELTS. The two most informative chapters in Alderson’s *Diagnosing foreign language proficiency* are chapters 8 and 14, which deal with the role of self assessment and the value of feedback and advice, respectively. Green’s volume on the IELTS academic writing test offers practical advice and success strategies to various exam stakeholders, namely administrators,

teachers and students. The final chapter of *Language testing: The social dimension* by McNamara & Roever is devoted to implications for training.

Volume titles

A crucial aspect of any professional volume is, of course, its title. This is probably the shortest sentence and/or phrase that any author will write and is perhaps the most difficult. For this reviewer, the title is a key feature of whether or not to buy (and subsequently to read) a book. For three of the volumes in this review I knew exactly what I was getting: the volume titles were descriptive, and indicative of, the volume contents. Two of the volume titles, however, did not meet my expectations.

In Fulcher & Davidson's *Language testing and assessment: An advanced resource book*, I was envisioning what some teachers might define as traditional 'resources' such as photocopiable grading scales, rubrics and the like. Upon receipt of the volume, I discovered that the nature of the 'resources' offered was very different from my expectations. This is not a criticism per se but when teachers search for materials on language testing, this volume is likely to come up on literature and internet searches and some teachers might have expectations similar to my own and could be disappointed. Alderson's title *Diagnosing foreign language proficiency* was perhaps the most misleading, as I was expecting a volume on diagnostic language assessment. In fact, only 27 of the 268 pages are devoted to this topic and the rest to the DIALANG project. Although the author eloquently explains his choice of title in the volume introduction, going as far as to call it 'over-ambitious' (p. 2), and effectively uses DIALANG to support his beliefs on diagnostic proficiency testing, I still feel that DIALANG should have at least appeared somewhere on the volume cover.

Cohesion and readability

As befitting the academic standing of these authors, the volumes are all well written and pedagogically and empirically sound. With the exceptions of Fulcher & Davidson and McNamara & Roever, all contain graphical data in the form of charts and tables to explain concepts or to report research findings. One major criticism that practitioners make about assessment books is the difficulty they have in interpreting the data presented. I found the graphical data presented here to be very understandable. The Gottlieb & Nguyen volume not only makes use of clear tables and survey forms in a practical 12 point font but also gives a photocopying release to the purchaser of the book.

Use as a course textbook

ELT professionals purchase professional volumes for a number of reasons. Some simply want to supplement their own professional libraries while others look to the purchase of a new volume as a potential course text. With that in mind I examined the volumes under review for their potential as graduate or teacher training course textbooks. The two volumes in this

review that are perhaps the most suitable as course textbooks are the Fulcher & Davidson and the Gottlieb & Nguyen volumes. In the former, the authors provide very well constructed 'Extension' and 'Exploration' tasks that would be ideal for use with pre-service and in-service teachers. In the latter, the authors employ 'Guiding questions' to set the chapter content into context and 'Questions for reflection and action' as a review and extension of chapter information. The other three volumes would also make valuable complementary texts in a language testing course although comprehension questions and extension activities for these volumes would need to be devised.

Ancillary materials

Some of the most valuable information in books nowadays is not to be found solely within the chapters themselves but also in the appendices. For the books under review here, this is no exception. Fulcher & Davidson's book stands out in this respect. In addition to the article and book excerpts found in the 'Extension' section, the authors provide an extensive glossary of terms as well as a comprehensive list of resources. Another valuable resource found in this volume is the companion website, which provides extensive additional readings, activities and links to relevant websites. Gottlieb & Nguyen's book contains a wealth of photocopiable charts, surveys and timelines adaptable for use in diverse contexts as well as a short glossary of terms. Green's book provides 40 pages of appendices which include the instruments used in his study as well as the IELTS writing band descriptors for Tasks 1 and 2.

Future directions

An important section in all professional volumes is a 'where do we go from here' section. These authors have tackled this in very different ways. Both Alderson and McNamara & Roever have devoted the final chapters of their volumes to looking to the future in their respective areas and suggesting implications for research and training. Green very effectively weaves this information into his concluding chapter, entitled 'Research questions revisited', in which he revisits his research questions individually, provides succinct summary answers to these questions and presents his findings in an 'Implications' section. Gottlieb & Nguyen accomplish this through their concluding chapter, wherein they make a plea to educators to systematically use relevant and purposeful language assessment practices in education programs. Fulcher & Davidson emphasize the importance of keeping stress at bay and having fun in the journey of learning about language assessment. They go on to encourage readers to keep a 'testing diary' of important assessment issues that they encounter along the way and invite readers to contact them with ideas for a future second edition.

Concluding remarks

It is evident that the volumes under study are complementary to one another with some limited but important overlap. The five volumes under review are similar in that they each

are heavily based on empirical findings within the field of language testing and assessment. They also focus on issues within the field that are deemed critical in the profession. The authors do not shy away from voicing their opinions on these thought-provoking issues. For these reasons I feel that the volumes in this review will make significant contributions to the growing body of literature on language testing and assessment.

Acknowledgements

I wish to express thanks to Ewa Jaworska and Graeme Porte, and two anonymous reviewers for their support and helpful feedback on earlier drafts of this review. Gratitude is also expressed to Stephen Stoyloff who assisted in the initial framing of this review.

References

- Messick, S. (1989). 'Validity.' In R. L. Linn (ed.), *Educational measurement*. New York: Macmillan & American Council on Education, 13–103.
- Weir, C. & M. Milanovic (2007). Series editors' notes. In A. Green, *IELTS Washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press, ix–xii.

CHRISTINE A. COOMBE is a faculty member of Dubai Men's College and an Assessment Leader for the Higher Colleges of Technology, Dubai, United Arab Emirates. Her research interests include testing, assessment, teacher evaluation, teacher research and leadership skills in ELT. She is co-author of *A practical guide to assessing English language learners* (University of Michigan Press, 2007) and co-editor of *Evaluating teacher effectiveness in EF/SL Contexts* (University of Michigan Press, 2007), *Language teacher research in the Middle East* (TESOL Publications, 2007) and *Leadership in English language teaching and learning* (University of Michigan Press, 2008).