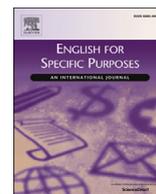


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

English for Specific Purposes

journal homepage: <http://ees.elsevier.com/esp/default.asp>

Review

Second Language Testing for Student Evaluation and Classroom Research, Greta Gorsuch, Dale T. Griffee. Information Age Publishing, Charlotte NC (2018). vii + 371 pp., US\$73.09 Hardcover, ISBN: 9781641130127

In recent years, there has been an increasing need for language teaching professionals to understand the principles and practices of language testing, as their testing-related responsibilities have grown. Among the factors that have been cited for this increase are the growing use of tests and assessments for accountability (i.e., the effects of the No Child Left Behind legislation in the United States), a rise in the use of language tests as a component of immigration policy, and an increased interest in the role that assessment can play in language programs in enhancing learning (Fulcher, 2012). These developments have prompted much discussion about the “assessment literacy” of language teachers, or the range of skills and knowledge they possess, and how to best ensure that it is adequate to engaging with issues of language testing both within the classroom and when engaging with the assessment-related policy and program administration decisions that directly impact both teachers and students. One aspect of this discussion has focused on the textbooks that are used to teach language testing, evaluating their accessibility and content (e.g., Davies, 2008). While the conclusions of such investigations are varied, it has been repeatedly noted that the training materials available on language testing are often aimed at “technicians” (psychometricians, statisticians) and inaccessible to classroom teachers, either due to high technical specification or a lack of explicit connection to teachers’ experiences with real language production in the classroom (Davies, 2008; Taylor, 2009).

It is into this landscape of changing testing-related demands, increased interest in assessment literacy among language teachers, and recognition of the need for accessible training materials that Greta Gorsuch and Dale T. Griffee have published *Second language testing for student evaluation and classroom research*. Their aim is to provide a balanced introduction to testing theory and practice that acknowledges the multiple roles that language educators often hold (administrative, research-related, and as classroom teachers).

The book succeeds in this goal largely because it is highly accessible. The authors explicitly state that they wish to make testing knowledge and skills available to teachers and graduate students, those interested parties “who do not like testing, especially the idea of statistical analysis” (p. xii), and the non-native English speakers among the previous groups. To serve this audience, in addition to explaining concepts clearly and providing examples from and links to language classrooms, statistical formulae are supplemented by detailed instructions on how to use them, and the authors have included a robust glossary featuring not just the technical but also the idiomatic terms used throughout the book.

The book is accompanied by a student workbook, which contains supplemental material for each of the book’s twelve chapters. Exercises in the student workbook fall into three categories: “Test Yourself”, “Discussion Questions”, and “Application Tasks”. The “Test Yourself” questions prompt readers to review key terms and concepts from the chapter. “Discussion Questions” involve less information recall and more application, asking readers to draw on their own experiences or providing hypothetical scenarios and asking readers to make and defend choices. Finally, “Application Tasks” involve applying concepts from the chapter to sample data sets. The activities presented in the student workbook are thorough, providing opportunity for both review and practice with both major concepts covered at the broadest level and nuances.

The textbook itself contains an introduction with a brief overview of testing as we know it, 12 chapters, each focusing on a different aspect of testing theory and practice, and a robust glossary. The introduction provides a brief history of testing, starting with IQ tests, and an explanation of how the ideas of IQ testing proliferated, leading to a variety of large-scale proficiency tests. After this brief overview, the introduction goes on to mount an argument for why testing as we know it may be over, citing reasons such as expense, and ethical concerns regarding fairness, power, and equality. Although these points are well-made, a fuller discussion of their ramifications for language teachers would be useful here—in other words, readers are left to wonder how these speculative major and foundational changes to large-scale testing might inform our understanding of the rest of the book.

The book contains 12 chapters, which can be broadly separated into three themes. Chapters 1 through 5 present the information readers need to understand test types and create their own testing instruments. Chapters 6 through 10 focus on

analytics, including interpreting test results and evaluations of instrument reliability and validity. Chapters 11 and 12 describe approaches to applying the information in previous chapters, broadly, to teaching and classroom research.

Chapters 1, 3, and 5 each introduce a different type of test (norm-referenced, teacher-made, and performance, respectively). Each chapter clearly presents the defining characteristics of its featured test type, along with how the test type works, its advantages and disadvantages, and considerations for approaching its design. Concepts are clearly explained throughout, well-cited, and illustrated with both general examples (drawing on familiar language classroom experiences) and examples from specific research studies. One particularly useful, repeated element throughout these three chapters is a chart presenting the characteristics of each test type. The chart gains a column per chapter as a new test type is introduced, allowing for quick comparison between the three test types, and contains information such as the test type's function, aim, reliability concerns, and characteristics of a good item for that test type.

In between these test type chapters, readers encounter test item formats (Chapter 2) and the role of theory in second language testing (Chapter 4). At first glance, the order of the topics covered in the first five chapters may seem counterintuitive, given that the three test types are perhaps best understood in contrast to each other. However, when read as presented, the order of information makes sense: norm-referenced tests are a reasonable starting place, given that high-stakes, standardized testing is familiar to graduate students and teachers, and given that few classroom teachers will be involved in the writing or creation of such tests. The chapter on test item formats, which thoroughly covers item types, their appropriate use, and tips for their creation, provides foundational knowledge that informs the discussion of teacher-made tests. Similarly, the introduction of the role of theory in testing provides information that can be drawn upon in the following chapter (performance tests).

Chapters 6 through 10 focus on the knowledge and skills needed to evaluate tests and interpret test results, covering Scales, Distributions, and Descriptive Statistics (Ch. 6), Correlation (Ch. 7), Reliability (Ch. 8), Test Validation and Validity (Ch. 9), and Standard Setting and Cut Scores (Ch. 10). The authors take particular care to acknowledge that even teachers who have been trained in statistics often find the analysis aspects of testing at best annoying and at worst daunting. In addition, the usefulness of each concept in language classrooms is carefully highlighted. Again, this is an area where the book stands out as exceptionally accessible. Formulae are well-explained, with detailed step-by-step instructions, and numerous opportunities to work with sample data sets are available in both the book and the workbook.

Perhaps the most exciting chapters in the book are 11 and 12, which encourage readers to take all that they learned and consider how it can be applied to their own classrooms, both through teaching (Ch. 11) and research (Ch. 12). Chapter 11 (Tests and Teaching) advocates for a view of testing as a tool that can have instructive purpose, rather than as a tool for assigning grades. To this end, the authors present what they describe as “two emerging ways of thinking about second-language tests and classroom teaching” (p. 267), discussing Roediger and his colleagues' work on test effect (e.g., [Butler & Roediger, 2007](#)), and [Grigorenko & Sternberg's \(1998\)](#) work on dynamic assessment. The explanation of test effect and its possible incorporations into a classroom teachers' repertoire is well-explained and supported. The authors cite promising research, summarizing the methods and results, and end this section with a list of concrete suggestions for how classroom teachers might implement these ideas themselves. In comparison, the description of dynamic assessment is sadly somewhat lacking. The concept of dynamic assessment remains somewhat difficult to grasp, and the authors acknowledge that “it may not be clear how to apply dynamic assessment to classes of learners” (p. 281). They make a valiant effort at addressing this by proposing an experimental, five-step model for the application of dynamic assessment. While the model is well-presented, the underlying concepts remain nebulous, with the result that this section of the text feels significantly less concrete and applicable than the rest of the book. Nevertheless, the authors' enthusiasm for emerging ideas about testing and its place in teaching is contagious, and for readers who are interested in learning more, the authors' experimental model and thorough citations provide a sound starting point.

The final chapter addresses testing and classroom research, describing three research types (confirmatory, descriptive, and evaluative) and the role that tests can play in each. Course evaluation is given more time than the other two research types, with a focus on how tests can help teachers make decisions that inform policy. Again, the authors are enthusiastic in their descriptions of how such research can benefit classroom teachers and students. A table on pages 290–292 consolidates mentions of tests and research from the rest of the book into a list of topics, by chapter, that may be of interest to readers who wish to conduct further inquiry. However, given the inclusion of “classroom research” in the title of the book, and despite the table, this chapter feels somewhat perfunctory. Although the table is useful, it is limited, and a more explicit elaboration of the research applications of testing would benefit readers with little research experience.

Overall, Gorsuch and Griffiee have written a strong introduction to testing knowledge and skills. The book will be useful both to true novices, including graduate students, and to classroom teachers who desire a targeted, easily understood refresher on language testing principles and practices. The strength of this volume is largely in its accessibility: the writing is clear, theoretical concepts are well and thoroughly explained, statistical procedures are approachably presented, and ample concrete examples are given from both research and classrooms. Given the growing testing responsibilities of language teachers, this book offers a timely resource for those wishing to improve their assessment literacy.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.esp.2018.06.001>.

References

- Butler, A. C., & Roediger, H. L., III (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19(4–5), 514–527.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327–347.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132.
- Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, 124(1), 75.
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36.

Erin Schnur
The University of Utah, USA
E-mail address: erin.schnur@utah.edu

Available online 30 August 2018