

Second Language Testing for Student Evaluation and Classroom

Research. Greta Gorsuch and Dale T. Griffee. Charlotte, NC:

Information Age Publishing, 2018. xix + 371 pp.

Reviewed by

Brandon Kramer

Osaka Jogakuin University & Junior College

Language testing is an inescapable part of almost any language curriculum.

Through its near ubiquitous influence on most courses of study, and even as the entire purpose for many students' English studies, language testing has, deservedly at times, earned somewhat of a bad reputation. In *Second Language Testing for Student Evaluation and Classroom Research*, much thought was given to addressing these concerns head on, encouraging teachers to learn more so that they can enact change and make better decisions using test results. Greta Gorsuch and Dale Griffee both earned their EdD degrees from Temple University Japan, and many of the points they make seem rooted in their experience teaching English as a Foreign Language. The book is very accessible. The authors try to position testing and the ideas that come with it, including discussions of validity and test statistics, in a way that not only allows teachers to understand the concepts but also see the relevance to their own contexts.

Although the book begins with a standard introduction to the history of testing, it is apparent from the start that this is a book made with a conscience, with the authors immediately arguing that we should use tests to help those in need rather than to find and enable those who started with an advantage. Chapter 1 contains the basics of norm-referenced testing, which refer largely to the standardized tests EIKEN, TOEIC, TOEFL, and IELTS, which are popular with learners in Japan. These tests are designed to separate students by ability or serve as a gatekeeping metric to identify students

judged to have a sufficient level of proficiency. It was particularly refreshing to see a discussion of test validation included in this chapter as well, with the salient disclaimer that tests themselves are never “valid,” but “rather, the question of validity is focused on the interpretation that test consumers make based on the scores [emphasis in the original]” (p. 12).

In Chapter 2, the authors’ efforts to explain ideas from the ground up with this book become clear as they move from defining basic terms such as test item to explaining the different item formats along with their unique considerations. An example of this is on page 26, where the authors advise ensuring item distractors (answer choices) for multiple choice questions are approximately the same length and recommend piloting test items with similar classes before their use. Although some of the rules of thumb provided here seem to lack explanation or justification, the practical focus of the book becomes clear and both novice and experienced test creators would likely find some useful advice within. In Chapter 3, the reader is guided through the process of developing one’s own tests while considering curriculum and course goals. This serves as an introduction to criterion-based testing and its role within a curriculum. In what is perhaps the most useful and important chapter for teachers and researchers new to testing, the authors outline the stages of test creation from construct operationalization to reporting on the test after administration. In Chapter 4, the authors seek to bridge the gap between theory and practice—or rather, the perceived gap between theory and practice. Positioning this book as one aimed for teachers who are not necessarily testing experts, they take great care in this chapter to explain why teachers should indeed care about theory in their daily teaching and testing. Chapter 5 is devoted to performance testing, which they define as a way for learners to demonstrate their language ability through a task or activity that they will likely need to perform in the future. These are, therefore, primarily measuring productive skills, and the authors discuss the creation

of holistic and analytic scales to measure student success in addition to rater training, which is an important but easy to neglect part of language assessment.

In Chapters 6, 7, and 8 the authors start to dive into the measurement and statistical aspects of testing, beginning with scales, distributions, and descriptive statistics, moving to a chapter on correlations and their uses, and then concluding with an introduction to reliability and dependability measures for norm- and criterion-referenced tests. Although whole books have been written on test validity (e.g., Markus & Borsboom, 2013), Gorsuch and Griffie manage to squeeze in a good deal of important basics into Chapter 9, discussing the most common subcomponents of test validity without getting too philosophical. Washback from testing and its influence on teaching for the test, as well as test use, are included in this discussion, which finishes with a short description of what the test validation process looks like. Chapter 10 is focused on score cutoffs such as the assignment of letter grades based on specific score ranges (e.g., 91%-100% is often an A) or the decision to admit or reject students who are above or below a decided score, as is often the case with admissions tests in Japan. The ideas presented in this chapter are very important given the weight that these score cutoffs often have on our students' futures, either through entrance to university or as records of grades that stay with students throughout their schooling. In Chapter 11, the authors challenge the reader by pointing out that there is often a mismatch between the assumptions teachers make about learning and the ways that they test knowledge. They suggest ideas to make tests more useful to learners by focusing on two topics, test effect and dynamic assessment. Test effect is the learning that takes place from the taking of tests and quizzes, and dynamic assessment is a growing field of testing which tries to measure not only what students have learned, but also what they "can do with assistance" (p. 275-276). In effect, both ideas stress that tests can be more than simply measurement tools—they can also aid in the

learning itself. The authors conclude the book with Chapter 12, a shallow dip into how tests can be used in research, including some example studies.

One of the more unique features of this book is the glossary, which is full of not only statistics and testing vocabulary, but also common idioms and phrases used throughout which may otherwise be difficult for a learner of English reading the book. In addition, this book has an accompanying workbook available with review questions and activities that encourage those using this textbook to bridge the gap between theory and practice in their own contexts. Together, these books would serve as suitable class materials for graduate courses in a master's program and as resource materials for doctoral students looking specifically into testing and assessment.

The strength of *Second Language Testing for Student Evaluation and Classroom Research* is its frequent references to real teacher experiences, and its aim to make language testing principles accessible. The authors wrote that they started their careers not particularly fond of tests, but over time begrudgingly acknowledged their necessity within language programs. It is easy to imagine this book was written to be the book they wish they had had when taking their testing classes early in their careers—indeed, in some ways, it is the book I wish I had had. The book is very teacher-centered, and it is apparent on every page that these authors still see themselves as teachers first and foremost. In most chapters of the book, they take the time to lay out step-by-step instructions on how to implement the ideas they introduce, helping those who may otherwise feel too overwhelmed to get started.

A weakness of this book could be that it is not comprehensive enough on its own to support the creation and maintenance of a language program.

That is not the aim, however, as it seems much more appropriate as an introduction to language testing for graduate students or teachers in training.

Readers looking to go to the next step may be interested in Carr's (2011) hands-on introduction to test design and analysis (including practice data

sets) or Fulcher's (2010) more advanced introduction with in-depth discussions of topics such as item specifications and washback. Overall, however, in this book, Gorsuch and Griffee provide the necessary explanations and practical advice for teachers to get started and talk about testing principles using accurate and accepted vocabulary. It provides a modest degree of depth for those curious to learn more while focusing more directly on the immediate steps teachers need to take, and the issues they need to think about now in order to improve their approach to language testing.

References

Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford, England: Oxford University Press.

Fulcher, G. (2010). *Practical language testing*. London, England: Routledge.

<https://doi.org/10.4324/980203767399>

Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning*. New York, NY: Routledge.

<https://doi.org/10.4324/9780203501207>