Book Review

AUTHOR(S)   : Greta Gorsuch and Gale T. Griffee

TITLE   : Second language testing for student evaluation and classroom
        research

PUBLISHER   : Information Age Publishing, Inc (IAP)

YEAR   : 2018

PAGES   : xix, 371

REVIEWED BY   : Ferit Kılıçkaya

Summary

Language assessment and evaluation has been one of the main concerns for many teachers of English due to the demands of creating and using appropriate assessment items and content. As Gorsuch and Griffee (2018) clearly indicate, "Testing would not and will not go away or decrease in importance; in fact, it is likely to increase. Then and now, the only reasonable response for a language teacher is to learn as much as possible about tests and testing" (p. x). Therefore, language assessment literacy requires language teachers not only to know and use several item formats but also benefit from statistical tools appropriately (Bailey & Curtis, 2015). There are quite a few books that aim to provide theory and practice regarding assessment and testing. Some of these books take a generic approach to introduce essential elements of measurement and assessment in the classroom without targeting any specific subject (e.g. Miller, Linn, & Gronlund, 2013), some focus only on language testing and assessment (e.g. Heaton, 1988; Hughes, 2003; Brown & Abeywickrama, 2010).

The current book, written by Greta Gorsuch and Dale T. Griffee, aims at investigating student evaluation and research in the language classroom, with the main focus on designing tasks, items, and key concepts such as test reliability and validity. The book includes 12 chapters, except the introduction. The introduction section of the book discusses tests and testing in various aspects and their functions and roles in classrooms as well as the society while touching upon important issues such as (un)fair testing and tests as instruments of power and control.

In Chapter 1, 'Norm-referenced Tests', the authors explain what norm-referenced tests mean, how teachers can use them, and the possible problems that are likely to emerge during application. In addition to the discussion on the characteristics of norm-referenced tests, this chapter also introduces the theory and the practice of 'item analysis' while trying to answer the question of what is a good norm-referenced item.

Chapter 2, 'Test Item Formats', deals with test contents: test items, which several other books on language testing and assessment name 'techniques for testing'. The chapter briefly introduces these items formats by categorizing them into two: receptive item

formats such as multiple-choice item and true/false item formats and productive formats such as short response and fill in the blank and discusses the main characteristics and the possible problems of each item format.

In Chapter 3, 'Teacher-mate Tests', the authors focus on criterion-referenced tests, which are also known as teacher-made tests. As indicated in the glossary of the book, a criterion-referenced test 'measures students' knowledge of a specified domain or content or skills area. Criterion in this case means the level of minimum desired performance to pass a criterion-referenced test' (p. 317). Unlike norm-referenced tests, criterion-referenced ones do not consider other learners' performance, but the individual performance compared to a minimum expected performance. This chapter also introduces item facility (IF) and difference index (DI) used in examining individual items.

Chapter 4, entitled 'The Role of Theory in Second Language Testing', discusses the relationship between theories and language testing and underscores that the knowledge of theory enables test writers and teachers to clearly examine and determine what is tested, which is the very first step of validation. In this regard, the chapter proposes a new model called HML theory (High-level, middle-level, and low-level theory), which includes grand theories (validity, communicative competence, and language proficiency), domain theories (theories from related areas such as education, linguistics, and psychology, and teacher theories (practical questions such as what works for teachers and why and how learners learn) are included.

In Chapter 5, 'Performance Tests'', the authors introduce performance tests that aim to assess learners' ability to use language during a task such as role-plays, class presentations, and final projects. The chapter discusses how human raters score the performance of learners in these tasks using analytic and holistic approaches to assign scores and presents clearly several advantages and disadvantages of performance tests.

Chapter 6 entitled 'Scales, Distributions, and Descriptive Statistics' describes the most commonly used scales and statistics used in the classroom. After the introduction of the types of scales (nominal, ordinal, interval, and ratio), the authors explain the concept of distribution, which is followed by the discussion on descriptive statistics including mean, median, mode, and kurtosis in addition to standard deviation, which provides useful information about test results. This chapter aims to provide the basic statistical information to language teachers willing to go beyond test scores.

In Chapter 7, 'Correlation', the authors introduce another common statistical procedure that can be considered as a basic statistical tool like the ones in Chapter 6. Correlation is a statistical procedure used to determine the relationship between variables such as exam scores and hours of study. This chapter focuses on different types of correlation calculated to investigate relationships such as Pearson product-moment correlation and Spearman's rank-order correlation, aims to explain why a specific type of correlation is selected and provides insights into the interpretation of negative and positive correlation between variables.

In Chapter 8, 'Reliability', the discussion is moved towards a crucial element of constructing well-written tests and of interpretation of the scores obtained on a test. This chapter first defines what reliability is in terms of theory and practice and discusses

different types of reliability such as test-retest reliability and parallel forms reliability. The chapter also focuses on test score reliability and presents various strategies that can be utilized to ensure a reliable test.

Chapter 9, 'Test Validity and Validation', deals with another important element of a good test, Validity, which requires determining whether the test assesses what it intends to assess. Like reliability, ensuring a valid test is one of the most important aims of a language test since language teachers and test writers wish to measure what they want to measure. Presenting several types of validity such as content and construct validity and the chronological steps to validate a test, the authors present various validation strategies to the readers and inform them of other issues such as test washback and its consequences.

Chapter 10, 'Standard Setting and Cut Scores', discusses several other important terms such as standard setting and cut scores in various methods such as norm-referenced method and the criterion-referenced cut score method.

In Chapter 11, 'Tests and Teaching', the authors reconsider the relationship between tests and teaching while discussing test effect and dynamic assessment, which might appear as a new term for many teachers around the world. Testing and teaching seem to be intertwined as any changes in one will lead to changes or at least might lead to several effects on the other. Therefore, the current chapter tries to increase awareness on the information provided back to them regarding learners' performance. Moreover, dynamic assessment is discussed and defined as 'a test combined with intervention to quantify developed abilities as well as learning potential' (p. 274), with a clear picture of how dynamic assessment works by explaining the model that this type of assessment uses: zone of proximal development.

The last chapter, Chapter 12, 'Tests and Classroom Research' discusses how tests can be used in classroom research as data-collection instruments. With this aim in mind, the authors present three types of research (confirmatory, descriptive, and evaluation) together with examples studies and the test functions and provide in a table a brief review of research conducted using tests in the previous chapters of the book. The chapter also touches upon course evaluation and the distinction between the types of evaluation: formative and summative evaluation.

Evaluation

The major strength of the book lies in the detailed and organized discussion of the test item formats, critical elements of tests such as validity and reliability, and the basic statistical tools and procedures. The discussions present not only the theory behind each issue presented in the chapters but also the practice with clear examples. Moreover, the chapter on the role of tests in classroom research provides an invaluable summary of the current research dealing with tests as data collection tools to graduate students as well as researchers who are willing to infuse tests into their research but do not know where to start. It is also due to note that any of the terms introduced in the chapters are defined with brief but clear explanations in the glossary.

One topic that is relevant and important but received very little attention and devotion in the book is the assessment of the receptive (listening and reading) and

productive skills (speaking and writing). While several other include assessment of the four skills (e.g. Hughes, 2003; Coombe, Folse & Hubley, 2007; Brown & Abeywickrama, 2010) and language components such as grammar and vocabulary in addition to these four main skills (Heaton, 1988), the current book does not include a separate chapter on this. One might also expect to a chapter that might focus on the use of technological tools in language testing in addition to another one that includes discussion on young learner assessment (e.g. Hughes, 2003). Many undergraduate programs in English language teaching include one and/or two-semester courses on language testing, and teacher trainers and undergraduate students might like to benefit from more information on these issues. Inclusion of cloze test formats in the second chapter would also be highly beneficial (see Bailey & Curtis, 2015 for this format).

In conclusion, regardless of the several shortcomings, the book effectively addresses the graduate students and scholar in Applied Linguistics and Language teaching programs. They will appreciate the detailed and organized discussion of the test item formats and the detailed discussion on the basic but powerful statistical tools, not to mention the summary of research presented in Chapter 12. Thus, I believe that the book will be of utmost value for those especially interested in test item formats and the use of tests as data collection instruments.

References

Bailey, K. M., & Curtis, A. (2015). *Learning about language assessment: Dilemmas, decisions, and directions* (2nd ed.). Boston, MA: National Geographic Learning.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). New York: Pearson Education.

Coombe, C., Folse, K., & Hubley, N. (2007*). A practical guide to assessing English language learners*. Michigan: The University of Michigan Press.

Gorsuch, G., & Griffee, D. (2018). *Second language testing for student evaluation and classroom research*. Charlotte, NC: IAP.

Heaton, J. B. (1988). *Writing English language tests.* New York: Longman Inc.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.

Miller, M. D., Linn, R. L., & Gronlund, N. (2013). *Measurement and assessment in teaching* (11th ed.). New Jersey, NJ: Pearson Education.

About the Reviewer

Ferit Kılıçkaya is currently working as an associate professor at the department of English language teaching at Burdur Mehmet Akif Ersoy University, Turkey. His main area of interest includes computer-assisted language learning and testing, language teacher education, and language teaching methodology. He has published several book chapters, articles, and reviews in journals such as British Journal of Educational Technology, Computer Assisted Language Learning, Teaching English with Technology, and Educational Studies. He can be contacted at ferit.kilickaya@gmail.com .