

Language Assessment in Practice

Lyle Bachman and Adrian Palmer. Oxford, UK: Oxford University Press. 2010. Pp. v + 510.

Language Assessment in Practice by Lyle Bachman and Adrian Palmer is a defining text within the field of language assessment. With an expected readership of students, language teachers with varying experience, applied linguists, professional language testers, materials developers and textbook writers, *Language Assessment in Practice* is ambitious at its core. Nonetheless, the authors' logical approach and straightforward writing style make this a highly accessible text for all members of its intended audience. In an inviting first chapter titled "Objectives and Explanations, or why we need another book about language testing," Bachman and Palmer explain their reasons for taking on this project, especially in light of their earlier, formative work *Language Testing in Practice* (Bachman & Palmer, 1996). Using Bachman's articulation of the Assessment Use Argument (AUA) as its organizing theoretical framework, *Language Assessment in Practice* represents a distinct departure from the earlier text. At nearly twice the number of pages of text (excluding the projects in the final chapters of both books), and incorporating extensive resources available online, this volume represents a significant evolution in the authors' thinking about assessment.

Language Assessment in Practice is organized into three major parts which are divided categorically into 21 related chapters, most of which end with a set of exercises, suggested readings, and notes. Part I, titled *Conceptual Foundations* (Chapters 2-5), presents the theory behind language testing and establishes a framework to guide assessment development and use. Part II, *Constructing an Assessment Use Argument* (Chapters 6-12), is arguably the most important contribution of this volume, specifically, its elaboration of the Assessment Use Argument (AUA). By guiding the reader through a step-by-step articulation of the practical reasoning behind crafting a test and interpreting its results, this section provides an innovative framework through which language assessment can be conceptualized. Part III, *Developing and Using Language Assessments in the Real World* (Chapters 13-21), gives guidelines to assessment developers and users who are working with tests in the imperfect academic world where theory and practice may collide. Finally, Part IV, *Projects*, gives three examples of how this assessment framework can be applied to design real assessments, and provides useful models for those who intend to implement this system for their own assessment needs.

Looking to the text chapter by chapter, Chapter 1 functions as an introduction in which Bachman and Palmer define their aim: to enable the reader to become competent in the development and use of language assessments. *Competency* is thus defined as an ability to show stakeholders that intended assessment uses are justified, thereby introducing the text's adherence to a structure of the justification of assessment use. The authors assert that reaching a level of competence in developing and using language assessments is necessary; though unfortunately, as stated in the final chapter, such a level is not often achieved by those stakeholders making important assessment decisions on behalf of schools or institutions. In Chapter 1, the authors illustrate the three most common misconceptions about language assessment and provide alternatives to complicate these widely held, erroneous beliefs. The first misconception lies in the belief that there is a single ideal test for any given testing situation. Such a misconception can easily lead test-developers to favor certain tests or test-types which are ill-suited or inappropriate for different groups of test-takers. The authors state explicitly that the perfect test does not exist,

and that test developers should focus instead on developing assessments whose uses can be justified to stakeholders. The second misconception relates to justification. Some test developers may believe that bringing in an expert test designer may help to deflect some of the difficulties of justifying test use to stakeholders and help to create a better test; however, this process creates a largely fragmented test development process. The authors believe that stakeholders and practitioners who are closest to the instructional and assessment practices are better suited for the assessment development process and should participate cooperatively. The third misconception relates to the notion that tests are “good” or “bad” in accordance to an inherent quality of the test itself. Bachman and Palmer believe that this emphasis is misplaced; rather, a test should be evaluated according to the degree to which its intended uses can be justified through an AUA, thus contextualizing the process of test selection and use. These alternatives to the three misconceptions provide the rationale for the language testing approach presented in this volume.

Chapters 2-5 (Part I) relate the conceptual and theoretical foundations behind language assessment. Many of the structural components from Bachman and Palmer (1996) have been retained here and integrated into the larger text. Chapter 2 defines terms of language assessment and use, beginning with sections differentiating the nature and uses of language assessment. In the nature of language use section, the authors acknowledge the lack of agreement in term use within the field of language assessment, but largely discount the value of making fine distinctions between terms when developing or using language assessments. The use of language assessment is defined in terms of “collecting information for making decisions” (p. 22). Included in this chapter are notions of how decisions are made and the consequences which result, as well as an examination of assessment as a formative tool informing instruction and learning. This chapter incorporates many visual representations of complex concepts, but the figures occasionally lack sufficient explanation to be beneficial for the uninitiated reader. Chapter 3 introduces the conceptual framework for language use and details the differences between reciprocal and non-reciprocal language use. Later, the construct of language ability is discussed as it underlies assessment performance. Apart from an expanded “knowledge of genres” section, the components outlined in the “language knowledge” section (pp. 44-48), are largely reproduced from the former volume and maintain their relevance here. Strategic competence and language skills sections preserve the essence of the authors’ earlier perspective in their treatment of metacognitive strategies and the limitations of a four skills ability model. In Chapter 4, the concept of Target Language Use (TLU) domain is described as “a specific setting outside of the test itself that requires the test taker to perform language use tasks” (p. 60) and has been similarly retained from the earlier volume. Significant explanation is dedicated to this concept which lays fundamental groundwork for subsequent chapters. A framework of language task characteristics, originally proposed by Bachman (1990) maintains its pertinence here as it describes five aspects of tests (setting, assessment rubric, input, expected response, relationship between input and expected response). The use of tests, accountability, and justification for test use are discussed in Chapter 5. Figure 5.1, which establishes the multi-step relationship from expected test consequences to assessment performance, is reprinted here in similar form to figures from Chapter 2, (2.2 and 2.3), and clarifies the concepts of interest. Accountability is discussed here, and refers to a process through which test developers and decision-makers convince stakeholders that the use of a test is justified, thereby fulfilling a necessary requirement of equitable and beneficial test use. Justification is critical notion within this text, and much of the subsequent development of argumentation is based on this idea. The authors state that justification serves

two purposes: (1) it provides a guide through the development of the language assessment and serves as a basis for quality control throughout the process, and (2) it provides the basis for test designers and decision makers to be held accountable to test users and others affected by testing decisions. Clearly, these initial chapters function as a precursor to the focal point of the text – the articulation of the AUA.

Originally introduced in Bachman (2005), the AUA is explored in this text as a conceptual framework by which to justify the intended uses of assessment. Borrowing a model of argumentation from British logician Stephen Toulmin (Toulmin, 2003), Bachman and Palmer utilize a framework of Practical Reasoning whereby argumentation can be understood and applied in an assessment context. Through an elaborate chain of logical arguments which builds on Michael Kane's work on validity and validation (Kane 2002, 2006; Kane, Crooks & Cohen, 1999), the authors provide a new way to think about test development through incorporating the elements of the data, claims, warrants, backing, rebuttal, and rebuttal backing into an argument. In the simplest terms, test users and developers collect basic situational data and make a subsequent claim about that data. Warrants support the claim that is made, and backing provides evidence supporting the truth of the warrant. However, a rebuttal suggests conditions under which a warrant may not support the claim, thereby weakening the overall argument being made. Rebuttals provide an alternate explanation to the assertions, and rebuttal backing furnishes evidence for the rebuttal statement, which in turn weakens the original claim. Much of the rest of the chapter is dedicated to the meaningful application of this framework to a larger language assessment context. Particular attention is paid to four claims of interpretation and use in an AUA which are examined in depth: Claim (1) Consequences and their related decisions are beneficial; Claim (2) Decisions take community values and equitability into account and are made based on test interpretations; Claim (3) Interpretations about language ability incorporate meaningfulness, impartiality, generalizability, relevance, and sufficiency; and Claim (4) Assessment records maintain consistency. Understanding the AUA and all its configurations requires close textual analysis, but early comprehension of the model's structural framework advantages the reader considerably in the chapters that follow.

Chapters 6-12 (Part II) progress from the purely theoretical framework of constructing an AUA to the step-by-step practical application of this framework to language assessment examples. Chapter 6 concerns itself with the process of assessment development and use and the conceptual steps that must be taken in that process. The five identified stages in assessment development and use (initial planning, design, operationalization, trialing, and assessment use) mark an expansion of Bachman and Palmer's previous system of stages in which there were only three categories (design, operationalization, and test administration). These five stages are thoroughly discussed here as well as in later chapters. Chapter 7 is dedicated to the process of initial planning and includes a discussion of the expenditure of resources in test development and administration as well as questions which must be asked in order for test developers and decision makers to plan effectively. Two useful examples, one of a Kindergarten ELL speaking and writing assessment (adapted from Frey & Fisher, 2003), and another of a University ESL reading test, illustrate what planning looks like in a real assessment situation. It should be noted that both of these examples, as well as a third relating to a Chinese mini-speech test (materials developed by Michelle Fu and Hongyin Tao) are found in their entirety in the Projects section (Part IV) of the text. The step-by-step design of an AUA, including its articulation through claims, warrants,

and rebuttals, is found in Chapter 8, though the reader is so frequently referred back to Chapter 5 for clarification that joining these two chapters may have been justified. The elements of an AUA are outlined here in a general and highly complex table, and include sections relating again to the four claims within the AUA. In the pages that follow, this table is re-formulated into a much more coherent and user-friendly question/answer format. Toward the end of the chapter, the reader is guided through an example of a Chinese mini-speech test to better understand the elements of an AUA in a real context. Chapters 9 through 12 focus in great detail on the four claims mentioned above (consequences, decisions, interpretations, and assessment records) and are framed according to the question/answer format stemming from the earlier “AUA articulation and assessment development guide.” Chapter 9 guides the reader through the actual articulation of claims, including a test’s intended consequences, and the different ways in which its use could affect stakeholders. An articulation of various warrants and their specific backing is found in this chapter, and the reader is led step-by-step through the initial logical phases in developing an AUA. Claims and warrants are presented in a generic form first, and then adapted according to contextualized examples relating to a kindergarten and a university assessment. By applying this complex theoretical framework to actual assessment scenarios, this model becomes much more transparent. Chapter 10 discusses the ways in which decisions are made when developing assessment. Again, referring to the university and kindergarten examples, the authors classify the many different kinds of decisions that are made based on tests. Again, working through the “AUA articulation and assessment development guide questions,” this chapter answers questions relating to which decisions need to be made, how societal and educational values are reflected in those decisions, and the assurance that these decisions will be equitable. In interesting discussion of cut scores is found here. Chapter 11 follows suit with a similarly detailed discussion of interpretations in relation to test-takers’ language ability. A critical feature of this component of the AUA is to define the construct of language ability. This chapter provides an enlightening discussion of the many different configurations of this construct, depending upon the specific characteristics of the assessment situation. The authors draw specific attention to the fact that they have discarded the concept of a four skills construct of language learning and assessment (based on speaking, listening, reading, and writing), since the natural integration of these skills make it extremely difficult to evaluate them individually. A brief Chapter 12 examines the use of assessment records as a means by which to maintain consistency. The aforementioned four chapters result in a nicely integrated and contextualized section which reinforces and clarifies key elements of the AUA.

Part III (Chapters 13-21) takes what has been established conceptually in Parts I and II and applies it to the real world of teaching and testing. Bachman and Palmer’s treatment of the inherent limitations of testing demonstrates great insight into the actual practice of designing and administering language tests. The authors’ sense of humor is interspersed in this section, exemplified in a subheading entitled “Welcome to the Real World.” An entertaining discussion is conducted relating to the uncertainty of language assessments when one strays from the theoretical realm into the practical realities of the classroom. A nice treatment of practicality as an impediment to assessment use is found in Chapter 13, though interestingly, it is not included here as a component of ‘usefulness’ as it was in Bachman and Palmer’s former volume. In fact, there are no references to the qualities of test usefulness in this text, which reflects a significant departure from their previous work. In chapter 14, the authors present the process of creating a Design Statement, a document which serves three main purposes: (1) leading the test developer

through the last stages of assessment development and use, (2) providing backing for several warrants in the AUA, and (3) providing interpretive information about the test to stakeholders. Through a ten-step process, the reader is guided through the university scenario to facilitate understanding of the complexity of creating such a document. The task characteristics template from chapter 4 reappears here and provides the reader a logical context to follow. A delicate discussion of “no specific TLU domain” is conducted to clarify cases in which a TLU domain is very difficult to define. A useful section delineating the considerations in estimating, planning, and allocating resources is also found in this chapter. Chapter 15 extends to the critical notion of the design of assessment task types and how test-designers choose between multiple versions of task types. The authors describe three situations for developing assessment tasks: (1) modifying a TLU task template for it to be usable as assessment task specifications, (2) the TLU task template can be used as an assessment task without changes, and (3) when there is no specific TLU domain to begin with. In these cases, tasks development will follow according to adherence to the warrants of the AUA, always keeping the characteristics of potential test-takers in mind to assure appropriateness of tasks. Assessment task specifications are a detailed description of this process, and include steps of (a) defining the construct, (b) describing the characteristics of the setting, (c) the characteristics of the input, response, and relationship between the two, (d) methods for recording performance, and (e) instructions for response. The robust section in Chapter 16 on recording assessment performance addresses feedback, ability leveling, rating scales, and score assignment, making it highly practical for the practitioner and designer alike. Relevant sample rubrics are provided as models for assessment use. Chapter 17, Blueprints, discusses the documents that guide test development. Providing a skeletal framework for writing the actual test, this chapter breaks down test-design into manageable component parts. Dedicated to the art of effectively creating test instructions, Chapter 18 suggests techniques for making instructions understandable and accessible to the test-taker. Chapter 19 presents the process and methods of collecting feedback during the final two stages of assessment development: (a) trialing and (b) assessment use. The authors highlighting the importance of collecting feedback at multiple stages of test development, as well as once a test is being used in order to achieve maximum benefit from the consequences of use and the resulting decisions. This chapter also details ways that feedback can be used to solve unexpected problems encountered in testing situations, and therefore have a positive impact on the subsequent testing processes. Major uses and sources of feedback, amounts of resources, and methods for obtaining feedback are usefully discussed at length in this chapter. Chapter 20 presents the use of resources in each stage of assessment development, and conducts a purposeful discussion of the human resources necessary for larger-scale language assessment design. Chapter 21 speaks to the allocation of responsibility amongst test-designers, test users, and decision-makers within the scope of language testing, and identifies areas which are as yet under-developed, including questions of competence, test literacy, and being a responsible test user.

Part IV presents three assessment projects in their entirety – a kindergarten English Language Learner (ELL) speaking and writing assessment, a university ESL reading test, and a university elementary Modern-Chinese speaking assessment, which are consistently reproduced throughout the text as partial and fragmented examples. Positioning these examples as appendices, and referring to them frequently throughout the text may decrease confusion and redundancy for the reader, and will likely support the authors’ eventual goal of teaching the reader how to build a cohesive assessment document one step at a time. Sixteen more examples

of language assessment projects are available free of charge by registering online through Oxford University Press.

Language Assessment in Practice is a highly recommended, essential text for its intended audience. By guiding the readership through a methodical and well articulated process of test development and use, Bachman and Palmer achieve their stated goal of increasing competence in the development and use of language assessments. Though it presents new theoretical perspectives, *Language Assessment in Practice* should not be considered a complete departure from *Language Testing in Practice*, but rather, an improvement upon it, as it retains and expands key elements and discards those that have been outmoded. Despite the authors' claims otherwise, *Language Assessment in Practice* does not easily lend itself to being read in parts due to the cohesive nature of the text and its reliance on the logical progression from the theoretical to the practical. Simply stated, without being familiar with the AUA framework, using this text as a quick reference would be difficult. This evaluation should not be misunderstood as a criticism — far from it. The authors have succeeded in their attempt to make this text accessible, humorous, and comprehensive, and a prominent place should be cleared for it on the bookshelves of any professional connected to the field of language testing.

BETH CLARK-GARECA

New York University

REFERENCES

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1-34.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Frey, N., & Fisher, D. (2003). Linking assessments with instruction in a multilingual elementary school. In C. A. Coombe & N. J. Hubley (eds.): *Assessment practices* (pp. 63-74). Alexandria, VA: Teaching English to Speakers of Other Languages.
- Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*. 21(1), 31-41.
- Kane, M. (2006). Validation. In R. L. Brennan (ed.): *Educational measurement* (4th ed.). New York: American Council on Education and Praeger Publishers.
- Kane, M. Crooks, M., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*. 18(2), 5-17.
- Toulmin, S. E. (2003). *The uses of arguments* (updated ed.). Cambridge: Cambridge University Press.