

Understanding Language Testing. Dan Douglas. London, England: Hodder. 2010. Pp. x +156.

Understanding Language Testing is a valuable contribution to language teacher education; it addresses the essential facets of language testing needed by language test users, including practitioners, program administrators, and future language test developers, to effectively evaluate and assess the language ability of language learners. The book covers the key considerations in the development and use of language tests today in a succinct yet comprehensive manner. It begins with a quick introduction of what language testing entails, and a discussion of what counts as language ability, the object of measurement in language testing. Douglas also examines the contextual elements of language use necessary to create authentic contextualized tasks that provide for valid and reliable inferences about a test taker's language ability. After that, he outlines the steps in test development, starting with a needs analysis through which the purpose of the assessment is defined, the targeted population is identified, and the target language use task and language characteristics are clearly described. Douglas then focuses on the development of test tasks. In Chapter 4, he considers alternatives in assessment, including conference and portfolio assessments, self- and peer-assessment, task-based and performance assessments, and finally, dynamic assessment. Chapter 5 presents a mini-course in statistics. The book ends with Chapter 6, in which issues in technology and language testing, technology and task types, and automated scoring, are examined.

Chapter 1 presents a coherent rationale for the value of testing language, addressing skeptical reactions to the growing institutional emphasis on standardized tests. Douglas acknowledges that language tests are like "rubber rulers" with certain features that may mean they are not such good measuring devices. He explains, for example, that there is disagreement about the nature of the units of measurement in a language test and what each of those units mean, or that the units are not uniform along the scale. Despite these handicaps, Douglas discusses the steps that test developers and users can take to ensure the usefulness of the results, such as test development procedures to ensure an acceptable level of accuracy and analytical procedures to help interpret the results appropriately and fairly. Douglas makes a case for ensuring the ethical use of language tests by practices that ensure reliability and by designing tests that provide for valid interpretations of test performance.

Chapter 2 examines the process of measuring language ability in order to make decisions about the learners. While many types of decisions, (e.g., admission to university or certification of non-native speaking air traffic controllers) are made in whole or in part on the basis of language test results, it is important to keep in mind that the object of measurement of any language test is language ability, which is demonstrated in various contexts through the skills of reading, writing, speaking and/or listening. To explore the issue of validity, Douglas builds on his own framework of the components of communicative language ability, which incorporates the Bachman and Palmer (1996) model. Given the notion that language is communicative and situated in a context, test tasks should provide relevant contextual information. Thus, to make valid inferences about language ability, language performance must be solicited through the creation of test tasks in authentic contexts that simulate the target language use domain. Douglas describes in detail the contextual features relevant to language use represented in Hymes' acronym *SPEAKING* (Setting, Participants, Ends, Act Sequence, Key, Instrumentation, Norms of interaction, and Genre) (Hymes, 1972). These facets of target language use play a part in

communicative language performance, and therefore, must be considered in measuring language ability.

Furthermore, in this chapter, Douglas establishes some principles for creating validity arguments and suggests various types of evidence to support the validity of the interpretations of test results. This chapter first addresses consistency of measurement by proposing that we must control both test taker factors and assessment factors in order to make interpretations of performance as reasonable as possible. Tests should also reflect the content learned as well as the learning approach. Also, they should be neither too difficult nor too easy, and they should be of appropriate length. Finally, the test setting should be as comfortable as possible and optimize a test-taker's performance. Assessment factors also have to be controlled (e.g., clear articulation of what is being measured and why, the number of tasks, the time allotted for each task). The task might also include *task input* or a prompt, and an expected response. Attending to both the test setting and to the assessment features of a test can provide for results that can be interpreted with greater confidence as accurate measures of language ability.

Chapter 3 addresses the steps involved in test development. Through a needs analysis, which can be more or less formal and rigorous depending on the purpose and scope of the assessment, we can determine what we need to test. Before developing the actual language test tasks, the test developer needs to first define the purpose of the test. From there, primary data about the target language use domain can be collected – a sort of environmental scan including a description of the context, target language used in that context, participants, and the nature of communication in that context. Douglas describes a subsequent stage of secondary data collection, for example, consulting experts in a discipline regarding texts or technical vocabulary. At that point, an analysis of task characteristics (the circumstances under which the texts are produced and used) and language characteristics (a systematic analysis of the elements of language in the texts) according to the Bachman and Palmer (1996) and Douglas (2000) frameworks for test task characteristics are in order. This chapter includes a brief overview and discussion of the different types of test tasks, and attention is dedicated to the need for consistent and optimal procedures in test administration to reduce any unnecessary threats to reliability. Douglas considers computer software currently available for the development of test tasks and weighs the advantages and limitations for its application, particularly for classroom assessment. Douglas, therefore, takes us through a step-by-step process in developing tests, from the identification of task characteristics and a set of criteria enabling the scoring of the test taker's performance, to the development of a blueprint for the test instrument.

Chapter 4 explores alternatives in assessment, but begins briefly with a consideration of various notions in assessment that have shaped test development, such as norm-referenced vs. criterion-referenced tests, discrete-point vs. integrative tests, and formative vs. summative assessment. Douglas stresses that communicative language testing most closely reflects the construct of language proficiency today, namely communicative competence or the ability for language use. The latter half of this chapter discusses alternative approaches to assessment. First, the author observes that qualities proposed in favor of alternative assessments have been embraced as desirable alternatives in assessment practices (e.g., assessments as extensions of normal classroom teaching activities familiar to the learners, as well as the production of language in natural, real-world contexts that are also culturally sensitive). Alternative

assessments might be viewed as more formative than a standardized test, as they include a component of student involvement. Conference assessments, for example, involve one-on-one meetings between the learner and the instructor. Portfolio assessments usually comprise a collection of a student's work, including a component of self-reflection, all selected by the student. Peer- and self-assessments also involve the participation of the learner in the assessment process, resulting in the enhancement of self-motivation and learner autonomy. A section is dedicated to task-based and performance assessments, which lend themselves to authenticity and the complexity of language use in target language use domains. However, these types of assessments also present challenges in the design of these complex tasks, the evaluation of performance of these tasks, and the valid interpretations of the results. Finally, dynamic assessment is an alternative based on a Vygotskian concept that we need to assess what the learner will be capable of doing in the future as a result of mediation by a teacher/assessor.

Perhaps one of the most valuable chapters in this volume for a language educator and language test user is Chapter 5, which presents a mini-course in statistics necessary for language test users to make meaningful, valid interpretations of test results. Douglas provides several illuminating examples to demonstrate how statistics, such as the standard deviation, can be a useful tool to help us interpret performances and provide a basis for determining whether tests are fair and accurate measures. He explains in terms that the layperson can understand how key assumptions like a normal distribution allows us to interpret with confidence the degree to which the results of a particular assessment is representative of a norm. He delineates clearly and simply how the average or mean allows interpretation of an individual student's performance on a test in comparison to that of the other test-takers. His explanation of how standard deviations are calculated and can be used to correlate performances on different measures of language ability makes this an invaluable introductory resource for the language educator and test developer. He walks us through correlation with clear, edifying examples that illustrate well the concept of statistical significance, and subsequently, the t-test of the difference between two averages. As an introductory text in a language evaluation and assessment course at the graduate level, this overview of statistics would be an extraordinary resource.

Chapter 6 paints a fantastic picture of a computer technology-assisted assessment featuring a richly contextualized performance-based task with instantaneous scores and video and audio feedback to the test taker. Douglas reminds us that this is not the current state-of-the-art, but that it is not out of the realm of possibility in the near future. While some may protest the use of computers in language testing, Douglas explores several more productive questions (e.g., the effects of technology on test taker attitudes like test anxiety or motivation and the difference in language performance relative to different technologies). Douglas raises an interesting question about the effect of technology use on the construct of language ability that is measured. He maintains that if a particular technology, such as email, is a relevant aspect of a language use situation, then the language user who is not competent in that technology cannot be considered competent in communication in that context. Some would contend that competence in the use of email is not a part of the construct of language ability. Part of this dialogue is, of course, the pros and cons of automated scoring, which Douglas only briefly discusses. This chapter ends with a detailed and convincing review of a few more sophisticated computer programs, such as *DIALANG* for testing listening, that leave us again with anticipation for what future technology-supported assessments will bring.

With *Understanding Language Testing*, a volume in the *Understanding Language Series*, Douglas has presented language educators and language test developers and users with an invaluable tool, a guidebook to navigate the sometimes overwhelming thicket of “Language Testing.” Topics range from concepts essential in an introductory or basic course in language evaluation and assessment to contemporary trends, such as the use of technology in assessment. The explanations of testing concepts are clear, concise, and understandable, and the examples are edifying. This is a text that is appropriate for a wide audience. It provides a succinct overview for the more serious language educator, who will eventually be inspired to choose a specialization in assessment, as well as an accessible introduction to the educator who simply wants a working knowledge of language testing in order to create classroom assessments that can be used with confidence to measure the language ability of their students.

CYNTHIA WISEMAN, EdD

Borough of Manhattan Community College, CUNY

REFERENCES

- Bachman, L.F., and Palmer, A.S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge, UK: Cambridge University Press.
- Hymes, D. (1972). Models of the interaction of language and social life. In J.Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 35-71). New York: Holt, Rinehart, & Winston.