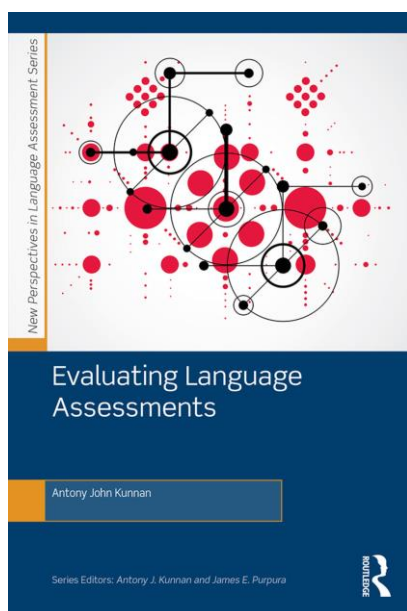




Book Review



Evaluating Language Assessments, by Antony John Kunnan, Routledge, 2018, XIX + 273 pp., 34.99 GBP (Paperback), ISBN 978-0-415-89777-8

The volume makes a unique contribution to the field of language assessment focusing on the issue of establishing fairness and justice among language assessments. In so doing, it provides the reader with a new perspective towards assessing learners meaningfully with justice and fairness based on the robust theory of ethics. Given this unique feature, the book deserves a new subsidiary title of “a philosophical investigation into the principle of language assessment” or something to this effect. Indeed, the author successfully accounts for the deep philosophical issues to be dealt with in down-to-earth words, in a way in which he disposes readers to practice them in their contexts.

The author starts by tracing the history of educational assessments in general, thereby helping the reader to fathom the importance of the issue from a wider perspective. Chapter 1 is so rich in content that it may be read as an independent work particularly for those who need to know about the history of educational testing in general and that of language testing in particular. The reader is advised to read this chapter by examining each example case that is showcased in history as an exercise critical review of the quality of assessments in terms of fairness and justice.

Chapter 2 starts with the observation that evaluation of assessments does not have a long history as it deserves. Readers will agree with the author when he claims that the “main reason for this has been the erroneous belief that assessments are infallible because they seemed to be beneficial to society” (p. 34). Despite this assumption, however, we are well aware albeit tacitly that assessments are not always fair, just nor flawless. The author then raises an important question: “as individual test takers, can we do anything about any of these situations given the obvious power imbalance between test takers and assessment institutions?” (p. 34). The answer to this question is in the affirmative and the entire book is devoted to addressing this issue.

Chapter 3 calls our attention to the fact that a philosophical basis has been lacking in the previous approach to validating language assessments. This observation makes the entire volume unique by distinguishing itself from any other work that has been published to date in the field of educational assessments. This chapter provides the overview of major philosophical positions, ranging from religion-based ethics, utilitarianism, deontology, which are useful preambles of the ethical issue of language assessments, the topics that are covered in the subsequent chapters. The explanatory account for these theories is followed by the discussion on Rawls and Sen, whose theories provide a basis for the subsequent argument. Based on their theories, the author proposes two general principles; that an assessment ought to be fair to all test takers with the presumption of treating every test taker with equal



respect, and that an assessment institution ought to be just, bringing about benefits in society, promoting positive values, and advancing justice through public reasoning.

Chapter 4 comprises the methodological component of this book, where Toulmin's (1958) approach is employed to link the principles of fairness and justice. Though this model has already been widely used for validating language assessments (e.g., Chapelle, Enright, & Jamieson, 2008), it may be daunting to those who are not familiar. Indeed, it is such an audience who will benefit from this book, as it gives an overview of the method of the argument-based approach to validating assessments. It is to be noted that as the author repeatedly emphasizes the method can be applied to teacher-made assessments or low-stakes assessments as well as large-scale assessments. This chapter will serve as a manual for those who wish to employ it to validate assessment in their own contexts.

Based on the theoretical underpinnings that are provided in the previous chapters, Chapters 5, 6, and 7 illustrate the process of building an argument for fairness by examining typical cases concerning the principle of fairness, whereas Chapter 8 examines the claim of beneficial consequence relating to the principle of justice.

The notion of opportunity-to-learn (OTL) (Chapter 5) stipulates that the test-takers be given adequate educational opportunities and learning resources before they take an assessment. The OTL is to be guaranteed by qualified teachers, clear and safe facilities, up-to-date books and quality learning materials, and school conditions that promote these. One of the valuable suggestions offered involves that the learner has to be given the chance to learn from the assessment. This means that test takers have to be provided with the diagnostic information, which is "an overlooked aspect of assessment systems in schools" (p. 135).

In Chapter 6, the term *meaningfulness* is taken as an alternative to *validity*. Indeed, the ultimate goal of validation is to establish that the assessment provides stakeholders with information that is meaningful and relevant to the given purpose. The examples cited in the chapter are all familiar to us as they are related to classroom language assessment. Amongst many such examples, the most relevant to language teachers would be the claim that a classroom language assessment is fair to all test takers (general claim), which means that the assessment is meaningful in that language, content, and the topic are based on curriculum objective (subclaim). This example convinces us of the usefulness of building up an argumentation however naïve it might be in terms of meaningfulness concerning the issue of fairness, so they may be checked against warrants or evidence. By so doing, we will be able to enrich our assessment practices.

The topic of Chapter 7, *bias* in language assessment, involves the type of knowledge that might appear to be rather technical to those readers who have little knowledge about statistics. However, statistics are important for understanding issues concerning a *biased test*, the very definition of which is the kind of test that "yields scores that have a different meaning for members of one group from their meaning for members of another" (Jensen, 1980, p. 516, cited on p. 167). To make the topic familiar to our own context, it would be serviceable to consider the differential meaning of *bias* to the type of assessments for different purposes. By way of illustration, consider the case in which one develops a proficiency assessment to general test takers. The topic of a reading passage ought to be unbiased and fair to all the students irrespective of their majors. However, if the purpose of the assessment is to examine learners' language ability along with the subject matter knowledge they are supposed to have acquired in other subjects at school, an attempt to assess that knowledge by the test would not invite a bias but rather comprise an essential element of the assessment, as is the case in the framework of Content Language Integrated Learning (CLIL) (Watanabe, 2021).

Chapter 8, covering the topic of washback and consequences, cites examples from high-stakes examinations, such as *the U.S. Naturalization Test*, which may not be familiar to all readers. However, the basic tenet of this chapter applies to the type of assessment we conduct in the classroom. Particularly important is the notion of *intended* and *unintended* consequences. We conduct an assessment with a specific purpose, and we normally check to see if the purpose has been achieved by the student's performance on the test. However, it is not so common to pause to consider that the assessment might

have brought a negative unintended consequence. We may administer a challenging test expecting to motivate learners, but it may turn out to demotivate them. This is the aspect of consequence that is easy to escape our attention, and hence, important to examine carefully. In this sense, a *claim* can be taken as synonymous with a *hypothesis* to be tested.

In Chapter 9, the author proposes ways of promoting justice and fairness in language assessments by the pre-service and in-service training programs. It is noteworthy that the list has test-takers in addition to stakeholders such as teachers, assessment professionals, policy makers, etc., as one of those who “need to be made aware of ethical thinking that can bring about responsible behavior related to assessments” (p. 226). The range of issues is then illustrated by scenarios simulating the situation where every one of us may encounter. One among many scenarios involves such a familiar case where though compensation to the test-takers who were misclassified as failed, the results were available only after the university admission has been completed. By carefully inspecting the scenario, readers will be able to deepen their understanding of the problem that they may encounter in their context.

Chapter 10 recapitulates and expands the issues that have been covered in the previous chapters by elaborating Toulmin’s model of argumentation in an attempt to help readers apply them to their context. In so doing, the author analyses more than twenty assessments from different countries in terms of various features, and summarises them in tables. Meanwhile, he urges us to investigate justice and fairness in teacher-made, low-stakes assessments for classroom use as well as high-stakes assessments. The summary tables will be an extremely useful resource to that end.

Written in a reader-friendly manner, the author challenges the reader by posing many thought-provoking issues about the topic of fairness and justice of language assessments. In doing so, the book reminds us of the importance of taking language assessments so seriously that we may gain the greatest benefit from language assessment practices. The book should be read by all language teachers, researchers and whoever wish to examine the ultimate goals of testing and assessing language learners.

References

- Chappell, C.A., Enright, M. E., & Jamieson, J. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.
- Jensen, A. (1980). *Bias in mental testing*. The Free Press.
- Toulmin, S. E. (2003). *The uses of argument, updated edition*. Cambridge University Press.
- Watanabe, Y. (2021). Assessing Soft-CLIL. In Ikeda, M., Izumi, S., Watanabe, Y., Pinner, R., and Davis, M. (pp. 121 – 147), *Soft-CLIL and language teaching*. Routledge.

Yoshinori Watanabe
The Graduate School of Languages and Linguistics, Sophia University, Japan
yjwatana@gmail.com

(Received October 24, 2021; Revised November 20, 2021; Accepted December 18, 2021)