Book review


Reviewed by: Franz Holzknecht, University of Innsbruck, Austria

Although listening is often considered “the most fundamental skill of all” (Buck, 2018, p. xi), it has traditionally been underrepresented in language testing research compared to reading, writing, and speaking. Possible reasons for this comparative dearth of studies might be the complexity of the listening construct, the difficulties associated with collecting data on listening test takers’ response processes, and the many practical challenges of designing authentic listening test tasks. However, recent years have seen an influx of studies on testing listening, particularly in English as a Second Language listening. Despite the surge of studies, not many books have been published in this area. Ockey and Wagner’s edited volume Assessing L2 Listening: Moving Towards Authenticity addresses this need and is an important and timely contribution, as it discusses four of the key challenges for assessing L2 listening in the twenty-first century: the use of authentic listening texts, the inclusion of a variety of accents, the role of audio-visual texts, and the rise of interactive and integrated listening tasks. The book will be of interest to language testers involved in designing and researching L2 listening tests, as well as to graduate students and language teachers who would like to deepen their understanding of listening assessment.

The volume is organized into four main sections, each following one of the themes outlined above. Within each section, Ockey and Wagner provide a useful overview of the theme discussed, before presenting a number of empirical studies on various aspects related to the theme. Section I deals with the oft-debated notion of using authentic, real-world spoken texts for testing L2 listening. Authentic spoken texts are generally regarded as being more construct-relevant than scripted texts for most assessment contexts, but scripted texts have the advantage that they can be tailored to test specifications and are therefore deemed more practical by test developers. It is perhaps for this reason that the vast majority of current high-stakes L2 listening tests feature scripted texts instead of unscripted authentic materials, as highlighted by Ockey and Wagner in their introductory overview of this section (Chapter 2). Using scripted spoken texts in listening assessment, however, can potentially lead to negative washback of using scripted texts in classrooms. Chapter 3, thus, presents research that suggests a compromise between authentic and
scripted materials. In this study, Wagner provides initial evidence that it is possible to “authenticate” scripted texts to some degree by changing the speech rate and inserting filled pauses, redundancies, false starts, and backchannels. In Chapter 4, Brown and Trace also demonstrate the importance of including authentic language in listening tests, based on their findings in a study comparing two different types of connected-speech dictations with two traditional listening tests. The authors of both chapters caution against the use of inauthentic materials and provide strong arguments for including real-world connected speech in listening test tasks.

Section II then turns to consider the use of different accents in listening texts for L2 listening assessment. Globalization and the rise of English as a Lingua Franca (ELF) have called into question to some extent the traditional practice of assessing only the standard speech variety of a certain context. As outlined by Ockey and Wagner in the introduction to this section (Chapter 5), many contemporary (English) contexts include an amalgamation of different accents, which language testers need to consider when designing listening tests. One possible approach is to include a number of accents in the test that are close to the standard accent of the assessment context, in order to avoid disadvantages owing to unfamiliarity. However, to achieve this, the strength of accents first needs to be judged on a rating scale. In Chapter 6, Ockey provides evidence that a “strength of accent” scale can be used reliably by a mix of L1 and L2 raters to differentiate between the comprehensibility of different speech varieties. Encouraging results also emerge in Chapter 7, where Harding reports that test takers exposed to highly unfamiliar accents make use of a number of compensatory strategies to aid comprehension. Harding’s exploratory findings are particularly interesting for ELF contexts where the exposure to unfamiliar accents is a crucial part of the construct (e.g., aviation, healthcare, academia). Finally, in Chapter 8, Kang and Moran show that national (US) and international undergraduates, graduates, and teachers perceive standard English varieties to be significantly less accented and more comprehensible than non-standard English varieties. However, the authors did not study whether these perceptions had any effect on test outcomes.

The following section (Section III) explores the use of audio-visual texts on L2 listening tests, a topic that has also been much discussed within the field over the years. Some scholars maintain that decoding visual information is not part of the listening construct, and listening tasks should thus not be video-mediated. In addition, the majority of current high-stakes listening tests do not make use of audio-visual texts. However, an increasing number of researchers – including Ockey and Wagner in their overview of the topic in Chapter 9 – argue that the ubiquity of video in the twenty-first century warrants its inclusion in listening assessment. According to the authors, not including video in domains in which the listener can see the speaker would result in construct-underrepresentation. To shed more light on this issue, Chapter 10 presents a study by Suvorov that investigated in detail how listeners perceive and use visual information in a video-based multiple-choice listening test. Suvorov’s results show that test takers find lecture-related aspects of the video (e.g., textual information) more helpful than speaker-related aspects (e.g., the speaker’s gestures). He argues that not including construct-relevant visual support in listening assessment can jeopardize the validity of decisions based on test scores. In line with these results are findings by Batty (Chapter 11), who compared
test takers’ performance on audio-only and video-mediated versions of the same multiple-choice listening items. Batty reports that items were easier in the video version compared to the audio version, with implicit items more so than explicit items. This is interesting, as implicit items have traditionally been associated with higher difficulty than explicit items in audio-only listening tests.

The fourth and final section of the book deals with interactive listening as part of the construct of interactive and integrated oral test tasks. Recent years have seen a move away from the traditional four-skills approach in language testing towards an increased focus on integrated test tasks. For many assessment contexts where the targeted construct includes the processing of multiple modalities, integrated listen-to-speak tasks are deemed more authentic and have therefore become increasingly popular. In Chapter 12, Ockey and Wagner review the rapidly growing literature on this topic and conclude with two useful recommendations on how to assess listening as part of integrated oral test tasks. One of the challenges associated with integrated listen-to-speak tasks is the potentially differential item functioning depending on candidates’ listening and speaking abilities. To investigate this further, Ockey studied how the difficulty of speaking tasks that do not require listening compares to the difficulty of listen-to-speak tasks (Chapter 13). He reports that integrated listen-to-speak tasks are more difficult and that this effect is larger for lower proficiency candidates. Chapter 14 then focuses on oral tasks mediated by an examiner. In this study, Nakatsuha shows that some examiners seem to adjust their language according to perceived candidate proficiency, which results in differential listening demands for candidates. Finally, in Chapter 15, Choi and So propose a complex measurement model for reporting listening ability in integrated listen-to-speak tasks. In this technical chapter, the authors use data from independent listening and speaking tasks as well as integrated listen-to-speak tasks to estimate accurately the listening ability needed for successful completion of listen-to-speak tasks.

In sum, Ockey and Wagner’s volume comprehensively discusses four of the key challenges for designing more authentic L2 listening tests: using authentic listening texts, targeting a variety of accents, incorporating audio-visual texts, and including interactive and integrated listening tasks. The individual contributions will be of great interest to language test developers and researchers alike. Many of the chapters are also relevant for students and language teachers, although some of the studies may be relatively technical for these particular audiences. However, students and teachers would find the editors’ excellent chapter overviews and their introduction and conclusion to the volume highly illuminating.

The overarching message of the volume is that many current L2 listening exams still do not represent enough of the construct. All too often, for various complex (and practical) reasons, listening tests use inauthentic scripted materials, do not include enough speech varieties, or do not contain visual support when this would be appropriate and necessary. In addition, integrated tasks remain underrepresented, despite a clear rationale for including them in assessment. With a view to improve L2 listening tests in these four areas, Assessing L2 Listening: Moving Towards Authenticity offers deep theoretical insights and highly useful practical guidance in clear and accessible language. One theme that could perhaps have been discussed further is the impact of technology on the listening construct, particularly the role of repeated exposure to listening input and
self-paced listening, which have become increasingly common in many real-world contexts. Overall, however, Ockey and Wagner’s volume covers a substantive number of highly relevant topics related to L2 listening assessment. The book is an important and timely contribution and a great resource for readers interested in how to address some of the key challenges for assessing L2 listening.

**ORCID iD**
Franz Holzknecht https://orcid.org/0000-0002-1218-2062

**Reference**