ume an invaluable resource for SLA researchers, CALL and CALT developers, and instructors alike.

DOROTHY M. CHUN
University of California, Santa Barbara

TESTING

JOHNSON, MARYSIA. *The Art of Nonconversation: A Reexamination of the Validity of the Oral Proficiency Interview.* New Haven, CT: Yale University Press, 2001. Pp. xi, 230. $30.00, paper. ISBN 0–300–09002–1.

In the overview of the book, the author presents the two questions that drive her argument: (a) Is the Oral Proficiency Interview (OPI) a valid instrument for assessing language speaking proficiency? and (b) What is speaking? Curiously, the author considers the first question practical and the second theoretical.

In the first chapter, "The Genesis and Evolution of the OPI System," Johnson traces its origins from the 1950s to the present. Curiously, there is no mention of the SOPI, the new-generation tape-mediated Simulated OPI that is currently available. In chapter 2, "A Critical Appraisal of the OPI," the author evaluates the OPI, accompanied by a historical definition of validity and criticisms of the OPI made by researchers such as Bachman, Lantolf and Frawley, Van Lier, Ross and Berwick, Young and Milanovic, and Johnson and Tyler. The third chapter is a discussion of the second question (What is speaking? or What is a speech event?) that starts with the theoretical bases provided by Hymes, Levinson, Grice, and the conversational analysts (e.g., Sacks & Schegloff) and the discourse analysts (e.g., Schiffrin, Van Lier, Allwright). This discussion leads to a summary of the prototypical features of a conversation, an interview, and a classroom interaction in terms of the turn-taking mechanism, the repair mechanism, the topic and the question type.

These chapters set up the insightful chapter 4 that follows, entitled "A Discourse Analysis Study of the OPI." The data were 35 audiotaped OPIs, 10 at each level and 5 at level 1 of both male and female test takers. But there is a problem here: The OPIs are telephonic OPIs, not the standard face-to-face OPI used in U.S. government institutions. The key question, therefore, is whether the telephonic OPI and the face-to-face OPI are equivalent or similar in terms of turn taking, topic nomination, and repairs (among other features), given that visual interactive clues are absent in the telephonic OPI and that linguistic features of phone conversations need to be deployed. Therefore, there is a difficulty in accepting the author's argument, given that it is directed at the face-to-face OPI, but is formulated based on evidence from the telephonic OPI. Because the author does not discuss this critical difference, it could constitute a serious flaw of the study.

This problem notwithstanding, the first strength of the book lies in its discourse analysis (DA) of the telephonic OPI through carefully transcribed and analyzed data. The author concludes this chapter by stating that "the DA study results indicate that the OPI speech event represents an interview of two types: a survey research interview and a sociolinguistic interview. It is not representative of real-life conversation as the ETS claims" (p. 120).

The author then uses native speakers' perceptions of a speech event by having four OPI English testers from the Defense Language Institute and four naive native speakers (i.e., not language testers) of English investigate the telephonic OPI from an outsider perspective. These participants listened to 16 OPIs (4 for each base level) and filled out Semantic Differential (SD) forms. The OPI was divided into six categories, with four pairs of words for each of the categories. Some categories are the following: for format, the pairs of words were formal/informal, conversation-like/interview-like, natural/contrived, spontaneous/controlled; for tester, the pairs were spontaneous/controlled, uncooperative/cooperative, active-involved/passive-uninvolved, formal/informal; for questions, the pairs were varied/repetitive, unnatural/natural, conversation-like/interview/like, formal/informal, and so on. However, no quantitative analysis of the SD study is presented. Instead, readers are referred to Johnson's (1997) dissertation, and we are left with the author's conclusion of this analysis: "The findings of the SD study support the DA study findings reported in chapter 5. They confirm that the OPI does not test speaking ability in a real-life context—conversation" (p. 139).

The second strength of the book is its rather brief essay on the application of Vygotsky's sociocultural theory to language testing. Here the author draws on ideas from Vygotsky and Bakhtin in an attempt to formulate the features

of a social interactive oral event called the Practical Oral Language Ability.

The book unfortunately suffers from a few slips. For example, the author states Messick's point regarding validity that "validity is a property of test scores, not of tests" (p. 25), but in framing the first question, the author asks whether the OPI is a valid instrument for assessing speaking proficiency. Another is the expression "language speaking proficiency" (p. 1). Perhaps the phrase "second or foreign," as in "second or foreign language speaking proficiency," might disambiguate the expression. Yet another slip is that the author characterizes the Test of Spoken English (TSE) and the Speaking Proficiency English Assessment Kit (SPEAK) as "two popular tests" although the SPEAK is the retired version of the TSE and therefore is identical in terms of test content, questions, features, scoring, reporting, and interpretation.

These slips aside, the author's main argument that the telephonic OPI (and maybe the face-to-face OPI) does not assess speaking ability as it is experienced in real-life conversations is a notable point and deserves wider attention in institutions that use the test.

ANTONY JOHN KUNNAN
California State University, Los Angeles

O'LOUGHLIN, KIERAN. *The Equivalence of Direct and Semi-Direct Speaking Tests.* New York: Cambridge University Press, 2001. Pp. xi, 280. $30.00, paper. ISBN 0–521–66793–3.

The Equivalence of Direct and Semi-Direct Speaking Tests documents test development and research investigations of live and tape-mediated versions of an oral test intended to measure the language proficiency of skilled migrants to Australia. The oral test is part of a battery called *access:* (the Australian Assessment of Communicative English Skills). The Department of Immigration and Ethnic Affairs funded a consortium of Australian universities and the Adult Migrant Education Program to develop *access:*, although the oral component was developed by the Language Testing Research Centre at the University of Melbourne. Given that direct and semi-direct versions of the test were used interchangeably in overseas testing centers, the issue of equivalence of the tests and the corresponding scores is critical.

The present volume investigates the equivalence of test takers' performances on the two *access:* versions, the direct and semi-direct speaking tests. The book includes eight chapters. Chapter 1 provides an overview of the book, laying out the issues that are explored in more depth in later chapters. Chapter 2 provides an overview of the *access:* tests and a detailed description of the live and tape-mediated versions of its oral component. Chapter 3 discusses the case study as the approach employed to investigate the equivalence of the two versions of the test. Chapter 4 describes the Rasch FACETS analyses of trial data collected in 1992. Chapter 5 compares speech samples obtained from test takers who were administered both versions of the oral test. Chapter 6 delves into the test-taking and rating processes tapped by the two versions. Chapter 7 replicates the FACETS analyses reported in chapter 4, using data obtained in 1994. Chapter 8 summarizes the various findings and discusses the types and quality of information obtained from the sundry studies conducted.

The book is a significant contribution to the field at many levels. In terms of the research question, the author argues that concurrent validation based on correlation provides necessary but not sufficient evidence of the equivalence of the language proficiency construct tapped by two different versions of a test. For the most part, the different types of evidence reported in the volume uphold this central argument. While a high correlation ($r = .92$) provides one piece of evidence to establish that the two versions of the test are equivalent, other score information, as well as discourse and process analyses, reveal that test takers draw on diverse skills when taking the two versions of the test. As such, the present volume, with its multifaceted, indepth analyses, lends support to those who advocate abandoning the still-popular practice of providing only a correlation index as evidence of test equivalence.

Given that the volume focuses on comparing two versions of a test that employs different formats, the absence of reference to the considerable literature on test methods in both the measurement and language testing fields is conspicuous. O'Loughlin is obviously aware of the effect test method can produce. For example, he points out that test developers attempted to make the two formats match closely by tightly scripting the live interview so it will match the tape-mediated version. Moreover, he recognizes other test method factors, such as preparation and response time, presence of the interviewer, and the like, as influencing test takers' performances. In short, the volume could have been