BOOK REVIEW

The art of non-conversation:
A re-examination of the validity of the oral proficiency interview

by Marysia Johnson. (2001)
Baltimore: Yale University Press. (Pp. viii + 230). ISBN: 0-300-09002-1
Pbk: JPY ¥4,670 / USD $40 / GBP £21.37

The book is designed into 3 thematic segments. The first offers a very brief synopsis and description of the oral proficiency interview (OPI) in terms of its development, appraisal, and its theoretical foundation. The next section analyzes and interprets the nature and application of the OPI. The final section provides a theoretical rationale for constructing a better oral assessment tool.

Part 1 (Chapters 1 – 3)

After describing the OPI's historical development, Johnson outlines the structure of the interview, which comprises a warm-up stage, an elicitation stage with several role-plays, and then a wind-down. The author's description is no different than the one provided by Yoffe (1997, p. 3 – 9). Johnson describes the OPI as providing a global rating based on four main factors: global functions, context and content, accuracy and text-type. These are described comprehensively in Yoffe (1997) as well.

In this text, Johnson does not discuss spiraling, which is shifting the question types upwards in level difficulty *within* a topical domain, depending on the rater's view of the candidate's language level. It was difficult to find information on this particular topic, and one other reference to this appears in a conference presentation by Scott and Helmuth (2008, slide 25). They describe "spiraling up" only through intermediate level question exemplars, as in Figure 1.
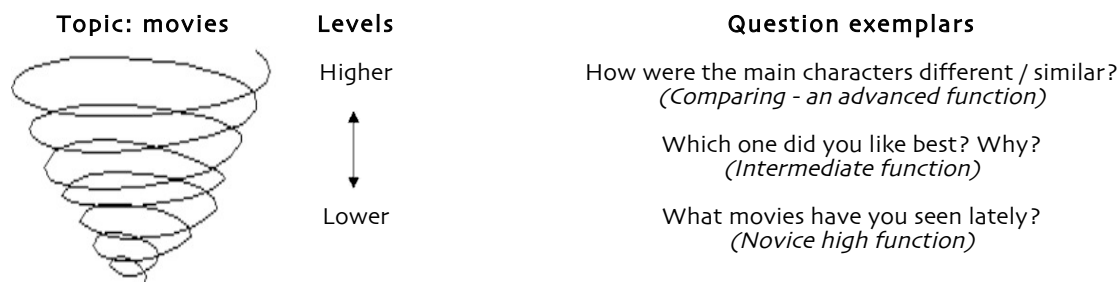

Topic: movies	Levels	Question exemplars
	Higher	How were the main characters different / similar? *(Comparing - an advanced function)*
		Which one did you like best? Why? *(Intermediate function)*
	Lower	What movies have you seen lately? *(Novice high function)*

Figure 1. An example of "spiraling up" questions on the topic of movies

Johnson's text mentions that there are *several* role plays in the course of an OPI interview. However, in my experience there is seldom enough time for this. The role-play is supposed to be an opportunity for the testee to play within a defined context and introduce ideas and content based on their own interests. Johnson does not discuss role-play in any detail, a major deficiency in her treatment of the OPI corpus.

Regarding the wind-down, Johnson does not deal with opportunities for testees to ask questions during this process, i.e. "do you have any questions for me?" I was encouraged in my OPI training to ask such questions as a last-ditch opportunity for testees to provide evidence of their communicative ability. Again, this would represent an opportunity to share the floor and overcome the power asymmetry of OPI and make it more communicative. Scott and Helmuth (2008, slide 22) also describe the wind-down as the stage which "returns interviewee to level at which s/he functions accurately [by giving her a] sense of accomplishment [and] positive feelings – this is important!" How this is accomplished is not mentioned, but presumably the interviewees would be given a chance to ask questions.

Is support for the OPI unfounded in light of recent trends in the testing field? Johnson refers to three main sources indicating that. First, Bachman's views (especially in Bachman, 1990) on testing foreground the idea of situational specificity (context), and in the space of one interview there are simply not enough samples of the testee's performance to predict general oral proficiency with any certainty. Also, she points out that Bachman does not like the idea of a global rating, since speakers may indeed have specific special oral skills that are not evaluated in OPI settings.

Second, Lantolf and Frawley's arguments (as cited in Johnson, 2001, p. 32ff) against making prescriptive guesses regarding task complexity are cited. Are "yes/no" necessarily less complex than proffering detailed information?

Moreover, Liskin-Gasparro (1984) and others claim that the OPI measures real-life performance. However, it invites the raters to compare testees to "ideal well-educated native speaker[s]". Lantolf and Frawley argue that this invalidates the process. Johnson suggests that, since a theory of proficiency is lacking, the OPI should be suspended if and until the psychometrics are sorted out.

Finally, the author refers to work done by Van lier (as cited in Johnson, 2001, p. 35ff) and his question about what the OPI actually measures. Van lier argues that the OPI has the characteristics of an interview rather than a conversation. Despite Liskin-Gasparro's (1984) claim that the OPI is conversation-like, the interactants and raters have very different goals. Raters are trying to make a decision which produces stress among testees, who in turn have little (if any) control over the process.

Part 2 (Chapters 4 – 7)

In the second section of the book, the author uses discourse analysis to dissect a corpus of OPIs, with samples taken from a variety of language levels. She argues against a conversational structure of the OPI through the analysis of three models: an interview as a speech event, as a research survey interview, and as a sociological interview. With regard to a speech event, Levinson (as cited in Johnson, 2001, p. 46ff) describes conversation as characterized by (1) turn-taking, (2) repair, (3) adjacency pairs (i.e. connecting ideas), (4) topic selection and maintenance principles, and (5) a discourse unit. The survey interview casts the rater as using ideal elicitation techniques, designed to provide the best possible ratable sample. The sociological interview is conversational at the beginning and the end, thus more open-ended. However, the middle is a hybrid of elicitation and interaction. The author goes on to apply the five conversation characteristics of classroom situations, summarized in Table 1, contending that the OPI in many respects closely resembles classroom talk:

Table 1. *A comparison of the pragmatic processes characteristic of classroom talk and the OPI (based on Johnson, 2001, pp. 65 – 72)*

Characteristic	"Typical" Classroom contexts	OPI contexts
Turn-taking	Teacher-controlled	Typical pattern – question is turn point for conversation; timing, direction dominated by rater / teacher
Repair	Teacher dominates feedback and pre-empts students' self-repair	Repair is generally self-initiated; the OPI format expressly prohibits corrective feedback because errors help to discriminate ceiling performance
Topic	What is discussed is irrelevant, the process (how) is critical	Topic changes are contrived, arbitrarily driven by the rater's agenda and choice; topics were treated so formally and had so little relatedness to each other it seemed as though the topic list was chosen ahead of time; tended to extreme insensitivity in some instances (asking Iraqi refugee to comment on Hussein's regime; immigrant to comment on victimology)
Question type	Display and comprehension questions dominate	Predominantly information-seeking / checking, with the candidates occasionally requesting clarification; check / probe – expert questions meant to check baseline performance (floor) and probe for potential, conducted as a formal verbal exchange without any negotiation
Discourse unit	Teacher talk	Overwhelmingly question and answer unit, with some variance in the warm-up stage

The author makes the interesting note that "most OPI testers are classroom teachers," and this may be one reason why the OPI could have the characteristics of classroom speech (Johnson, 2001, 64). The note was anecdotal, and no evidence is given to support this assertion.

Chapter 5, the keystone of this book, analyzes the OPI through the technique of discourse analysis (DA). Drawing from a study of 35 (mainly advanced level) telephone interview samples, the author shows how the OPI is not a conversation based in terms of the criteria listed in Table 1. She continues her analysis of the OPI samples by comparing the perceptions of 4 native speaking testers and 4 non-testers about OPI as a speech event. Both groups were asked to rate the samples across 6 factors (test format, tester, questions, candidate / testee, topic nomination, and turn-taking). Her findings were that, overall, there was no significant correlation on what the OPI represented when all stages of the OPI were collapsed together. When each stage was considered separately, the

agreement improved substantially, with the warm-up / wind-down stages rated as resembling a sociological interview, and the level-check / probe stage as controlled interview. The consensus again was that the OPI was not a conversation as advertised on the whole, and that each stage of the OPI was a distinct genre.

In Chapter 7 the author concludes her description of the OPI's three stages: it resembles a highly controlled research survey interview book-ended by more informal sociological interviews. She suggests providing raters with samples of natural conversation so as to create a more conversation feel to the process, and advises that the lead-in should be done in a way that is less abrupt to reduce its artificiality. While well-intentioned, these suggestions cast the raters as test-producers. However, the raters are not in a position to negotiate the content / makeup of the OPI. Their work is evaluated by ACTFL / ETS, and any significant and consistent straying from the test specifications would result in their decertification. For Johnson's suggestions to have traction, they would need to be grafted into the very composition of the test design.

Part 3 (Chapters 8 – 9)

After offering a historical development of how spoken discourse has been viewed since Chomsky, the author suggests a model for assessing spoken language. Johnson decries the fuzziness between competence and proficiency and contends that context-specific interactional competence, as envisioned by Vygostky (as cited in Johnson, 2001, p. 183ff), is the way forward in speaking assessment.

In Chapter 9 a socio-cultural approach to language testing is outlined. Johnson discusses how language development takes place, and claims that testing should be about evaluating potential rather than actual performance - what *should* be rather than what *is*. Communication, according to Bakhtin (as cited in Johnson, 2001, p. 192), is "half someone else's" and so must be evaluated in terms of the interaction. The conversation is framed, following on from Van lier's ideas, as the most desirable interaction form in the classroom (Johnson, 2001, 194).

Summary

In summary, I applaud the author for finally shedding some light on an issue that Yoffe (1997), and others (Bachman, Lantolf and Frawley, etc.) have drawn attention to: that is, to name the beast that is the OPI. There is an ample body of evidence suggesting that the OPI is not a conversation. However, this does not render the OPI useless, nor does that rationalize the use of conversation as the metric for judging speaking ability.

The art of non-conversation is a book for a highly select group of readers in the language psychometric field. It serves as a readable introduction and a non-technical analysis of an influential foreign language assessment instrument, but unfortunately does not deliver on its promise of providing a user-friendly, practical alternative.

-Reviewed by Gerry Lassche
Miyagi Gakuin Women's University

Works Cited

Bachman, L. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Brindley, G. (1989). *Assessing achievement in the learner-centered curriculum.* Sydney: NCELTR.

Kenyon, D. (2006). Interview-Based Oral Proficiency Assessments. Retrieved July 25, 2008 from http://www.nclrc.org/about_teaching/topics/assessment.html

Liskin-Gasparro, J. (1984). The ACTFL proficiency guidelines: Gateways to testing and curriculum. *Foreign Language Annals, 17* (5), 475 – 489.

Messick, S. (1988). Validity. In R. Linn (Ed.) *Educational measurement.* (pp. 13-103). New York: Macmillan.

Scott, V. & Helmuth, C. (2008). SLA Theories, FL Teaching, and Assessment: (Dis)Connections. Presentation given at Vanderbilt University, January, 2008. Retrieved July 25, 2008 from http://sitemason.vanderbilt.edu/files/gPYliw/DisConnections.ppt

Widdowson, H. (2004). *Text, Context, Pretext: Critical issues in discourse analysis.* Oxford: Blackwell.