

Ducasse, A. M. *Interaction in paired oral proficiency assessment in Spanish. Language Testing and Evaluation Series.*

Frankfurt am Main: Peter Lang. 2010. Pp 189.

Assessing speaking proficiency through paired/group oral tests has attracted increasing interest over recent years. Testing candidates in pairs or groups can elicit a wider range of functions and more symmetrical contributions among the speakers than in traditional examiner-candidate oral interviews, and therefore, more international language tests have come to use these formats in both high- and low-stakes contexts (Nakatsuhara, 2013). Among a number of studies in the area of paired oral tests, Ducasse's book makes a distinct contribution to the field in that it focusses on the beginner-level achievement test setting in an undergraduate Spanish language programme, employing extensive qualitative data analysis to gain insight into the construct of paired interaction.

This book is based on Ducasse's PhD study which sought to bring together classroom language teaching and in-house tests at her university. When it was decided that the university's Spanish programme would introduce paired speaking activities in class, the need arose for developing appropriate assessment tools accordingly. The advantages of using the paired oral test format included the authenticity of the interaction (more 'conversation-like' performance than in the traditional oral interview format), representativeness of in-class activities, the transparency and standardisation of rating across the first year groups of beginners, reduction of time and cost of testing, and reduced workload for the raters by removing the interlocutor role in the test. With these benefits in mind, Ducasse's goal was firstly to define and operationalise 'interaction,' and secondly to develop an empirically based rating scale for peer interaction. The book, therefore, consists of two main studies: Study 1 on the raters' and candidates' perceptions of peer interaction in a paired test format, and Study 2 on the development of a rating scale based on the findings from Study 1.

After an introductory chapter which gives the research background and an overview of the book, Chapter 2 discusses a wide range of literature in the field of oral interviews. The first part reviews the history, advantages and issues in testing speaking in paired and group formats, as well as the interlocutor effects on the scores, such as proficiency, familiarity and personality. The latter part focuses on the different approaches to rating scale development and validation.

Chapter 3 then introduces three research questions (RQs) and methodology:

- RQ1. *What features of peer interaction do raters attend to in paired task test performance?* (Study 1)
- RQ2. *How do candidates view interaction in a paired oral?* (Study 1)
- RQ3. *Can candidate peer performance samples from a paired test form the basis for developing a rating procedure for interaction?* (Study 2)

Study 1 (RQs 1 & 2) involved 17 video recordings of the candidates performing the paired oral tasks as the stimuli for raters and candidates to report on the features they noticed while watching and/or reflecting on the interaction. The tasks required each pair of candidates to spend 10 minutes talking about either set of three topics taught in class: 1) *family, holiday and friends*, or 2) *weekends, hobbies and summer holidays*. For each pair of candidates, one rater was present, observing and keeping time, who gave separate marks for the two candidates afterwards. The marks were based on the initial 5-level rating scales put together intuitively on *communication, comprehension, grammar and vocabulary*.

RQ1 was aimed at exploring the peer construct and what the word ‘interaction’ meant to raters (i.e. Spanish teachers). Twelve Spanish expert teachers worked in pairs, observing and commenting on videoed paired performances. The resultant verbal report data from the raters was then transcribed for content analysis in order to identify what features they noticed and paid attention to.

Similarly, RQ2 was intended for exploring candidates’ perceptions of the construct, which followed the suggestion by Galazci (2004) to include candidates’ own interpretations of the interaction. Twenty-five individual L2 Spanish learner candidates performed retrospective stimulated verbal recall on the video recording of their own paired performance. The data was then transcribed for content analysis to see if the candidates and the raters observed the same features of interaction.

Using the same data as RQs 1 and 2, key features of peer interaction were identified for RQ3. Eight video clips and eight key feature sets were then selected and given to three teams of L2 Spanish specialists, who participated in the procedure for developing an Empirically-based, Binary-choice, Boundary-definition (EBB) scale (Upshur & Turner, 1995). Following a modified version of the procedure proposed by Upshur and Turner (1995), each of the three teams of specialists came up with a series of yes/no questions which, when applied to a candidate’s performance, served as criteria to navigate raters to arrive at a final rating. Later, the three sets of questions were integrated to create a final EBB scale.

Each research question has a whole chapter that presents the results. Chapter 4 presents the findings for RQ1, where thematic analysis was used on the raters’

verbal report data. It was found that ‘success’ in interaction was judged by raters as a function of *non-verbal communication* (visual signals and body language), *interactive listening* (signalling comprehension and engagement; e.g. back-channelling) and *conversation management* (including *topic management* and *turn taking*).

In Chapter 5, the results for RQ2 are presented with three components that emerged from thematic analysis on the candidates’ verbal report data: *interpersonal non-verbal communication* (gesture, laughter, facial expression etc.), *interactive listening* (comprehension, listening (or not listening) to predict an interactive move) and *interactional management* (listening, taking turns and introducing a topic). A great deal of overlap was found between the raters’ and candidates’ comments, which provides additional validation evidence to raters’ judgements on peer interaction.

Chapter 6 corresponds with RQ3 and shows how the experts arrived at the consensual EBB questions. All three teams of experts agreed that candidates were distinguished in terms of the levels of *supportiveness* and *interactional management*. The first question of the final EBB scale was on the uses of supportive body language, followed by questions on the relevance of questions to the topic, and then on interactional management (i.e. cohesion between and within topics, turn length and response time).

Chapter 7 (Discussion) presents more of a recap of the findings from the previous chapters, rather than a discussion of the results in light of literature. However, this is understandable considering that the literature in Chapter 2 was reviewed in order to provide justifications for the instruments and format, as well as to inform the methodology used in this largely-qualitative study. Given the broad scope of this ambitious study and its wide range of data and breadth of analyses, readers will benefit from a summary of findings at this point.

Since this study aimed at addressing the needs in a particular context (i.e. devising a beginner-level achievement test in Spanish language which reflect the classroom activities), the limitations naturally lie in the generalisability of the results. As Ducasse states in Chapter 8 (Conclusion), employing different task types and casting stricter control over the candidates’ proficiency levels will be useful for future research. Moreover, a discussion on preventative measures against when a candidate is paired with an uncooperative partner – which is more likely to lead to ‘unsuccessful interaction’ – would have been desirable. Nevertheless, such limitations do not undermine the importance of this book.

It offers useful implications for developing and validating paired oral tests and rating scales, both theoretically and methodologically. Also, the author should be commended for the final EBB scale due to its ease of use for language teachers as not all strong research findings and resultant products can be useful in classroom practice. The greatest contribution of this study lies in the unveiling of the role of *interactive listening* in paired interaction. It is eye-opening that, as well as showing engagement and giving encouragement and support for the partner to continue, listening to the partner in order to *ask* relevant questions is more important than listening in order to *answer* questions in managing co-constructed interaction. Although Ducasse's study was situated in a local, classroom-based context, its implications have been accepted by a wider community of language testing. Recognising the importance of *interactive listening* in one's interactional competence, even a large-scale international examination board, Trinity College London, has introduced *Interactive Listening* as one of the rating criteria in the Speaking & Listening module of its recently revised Integrated Skills in English (ISE) tests. A number of solid, relevant studies on interaction have been published since, such as May (2011), Nakatsuhara (2011) and Isaacs (2013) to name but a few, and this area of research is likely to continue to develop.

Review by Chihiro Inoue

Centre for Research in English Language Learning and Assessment (CRELLA),
University of Bedfordshire, UK.

References

- Galaczi, E. D. (2004). *Peer-peer interaction in a paired speaking test: The case of the First Certificate in English*. Unpublished PhD dissertation, Teachers College, Columbia University, New York.
- Isaacs, T. (2013). International engineering graduate students' interactional patterns on a paired speaking test: Interlocutors' perspectives. In: K. McDonough & A. Mackey. (eds.). *Second language interaction in diverse educational settings* (pp.227-245). Amsterdam: John Benjamins.
- May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly* 8 (2), 127–145.
- Nakatsuhara, F. (2011). The relationship between test-takers' listening proficiency and their performance on the IELTS Speaking Test. *IELTS Research Reports*, vol.12. Available at:
https://www.ielts.org/pdf/vol12_report_4.pdf

- Nakatsuhara, F. (2013). *The co-construction of conversation in group oral tests*. Language Testing and Evaluation Series. Frankfurt am Main: Peter Lang.
- Upshur, J. A. & Turner, C. E. (1995). Constructing rating scales for second language tests. *ELT Journal*, 49(1): 3-12.