

**Salaberry, M.R. & Burch, A.R. (Eds.) *Assessing Speaking in Context: Expanding the Construct and its Applications. Multilingual Matters. 2021. Pp. 317.***

In recognition of spoken interaction as a dynamic, jointly constructed, and context-embedded activity (He & Young, 1998; Kramsch, 1986; McNamara, 1997), there is unanimous agreement that the conceptualization of the speaking construct should move beyond a psycholinguistic approach to a social perspective (Chalhoub-Deville, 2003; Fan & Yan, 2020). A handful of theoretical papers (e.g., Galaczi & Taylor, 2018) and empirical studies (e.g., Lam, 2018; May et al., 2020; Zhang & Jin, 2021) have, therefore, been devoted to the expanded speaking construct and its assessment. There are, however, growing arguments, debates and disagreements among language testing researchers and practitioners on how to viably incorporate and operationalize the co-constructed nature of social interaction in speaking tests. To further advance the scholarly development in this area, *Assessing Speaking in Context: Expanding the Construct and its Applications*, edited by M. Rafael Salaberry and Alfred Rue Burch, brings together research presented at the Rice University Center for Languages and Intercultural Communication's conference in 2018, with the prime objective of developing and implementing procedures targeting a sociolinguistic definition of speaking ability. In so doing, this volume is a timely and valuable contribution to the field: it pushes the construct definition of speaking tests into a broader conceptual terrain and provides a foundation for discussing the merits and pitfalls of different methodologies for assessing such an expanded construct.

Incorporating contributions from both leading and emerging researchers in the field, the volume consists of 13 chapters organized into four parts. Part 1, Conceptual and Theoretical Issues, includes three chapters. Chapter 1 by Salaberry and Burch presented a panoramic view of the driving forces leading to the expansion of the speaking construct, the problems and challenges encountered in operationalizing the expanded speaking construct, and possible solutions to address these challenges through examples of research findings from conversation analysis (CA). This chapter also laid out the main contents of the subsequent chapters, providing an overview of the book for prospective readers.

Roever and Dai in Chapter 2, with the aim of conceptualizing the construct of interactional competence (IC) for language testing, deliberated on the differences between speaking and talking-in-interaction, and addressed in great depth practical concerns in operationalizing a 'talking' construct in assessment contexts. The authors argued for a methodic needs analysis to narrow down the target domain and a purposeful

IC marker selection by means of the combination of “etic, researcher-based judgment with emic, data-based evidence” (p. 36). Finally, this chapter zoomed in on how the IC construct can be broadened to incorporate social role enactment (an under-researched sub-trait) as a rating criterion, employing an innovative approach, membership categorization analysis (MCA).

Chapter 3 by Plough centered on another less-explored area in second language speaking assessment, that is, nonverbal behavior (NVB). Acknowledging the fundamental role of NVB in face-to-face communication, Plough held that NVB should be incorporated into the speaking construct and that it should not be viewed as static, but “part of dynamic interactional processes” (p. 62). To assess NVB, Plough advocated adopting the emic-based approach of ethnomethodological conversation analysis (EMCA) and constructing a continuum of (im)permissible behaviors rather than a dichotomous conceptualization of acceptable and meaningful gesture versus non-gesture which is unacceptable and meaningless.

Part 2, *Collecting and Rating Speaking Data*, assembles three chapters. In Chapter 4, Burch and Kasper carried out a fine-grained analysis of how roleplay instructions transition the interaction from the interview-structured portion to the roleplay performance in the oral proficiency interview (OPI) through employing multimodal CA. Findings indicated that the examiner’s task instruction delivery, both verbal and nonverbal, was remarkably similar across the four video-recorded OPI cases and that variation occurs when intersubjectivity is at risk. This chapter also revealed that participants tended to be cooperative to achieve understanding and further stressed the pressing need for video-recorded data to observe how participants manage the interaction through a rich repertoire of semiotic resources.

Using think-aloud protocols, Chapter 5 by Youn and Chen investigated raters’ scoring processes and strategies before and after training on recognizing IC-related features elicited by paired roleplays. Before training, raters intuitively relied on factors frequently used in large-scale language tests, namely fluency, accuracy, and task completion. After training on the interaction-sensitive analytical rating criteria, raters closely interacted with the specific features included in the rubrics although variations occur in their decision-making processes, influenced by differential pragmatic norms and scoring styles.

Sandlund and Sundqvist in Chapter 6 investigated rater identities, defined as raters’ orientations to severity/leniency in their reflective practices, by combining a CA approach and observational tools from MCA. Results indicated that raters tended to view leniency

as more problematic than severity, as lenient raters were perceived to be inexperienced and overgenerous, while severity was treated as a safer option which embodied professionalism. The authors held that examining rater identities via membership categories can offer a new approach for research on rater variability and cognition.

Part 3 (Chapters 7-10), *Designing Speaking Assessment Tests*, centers on the implementation of novel procedures that assess the expanded speaking construct. Chapter 7 by Kley, Kunitz and Yeh focused on two L2 Chinese learners' use of repair practices in conversation with a native speaker from a CA perspective. Findings showed that repair practices were teachable, even for beginning L2 learners, but that pedagogically intended objectives and actual repair practices only partially align. To be specific, learners could initiate repair when non-understanding occurred. However, half of the learning objectives were not deployed and students used different varieties of repair practices. Furthermore, evidence indicated that the rating scale used in this study could not accurately differentiate between higher- and lower-level learners; refinements are thus required.

Drawing on data from an intermediate US university-level learner of French who completed one scenario out of a series of six, van Compernelle in Chapter 8 examined the potential of using dynamic strategic interaction scenarios (DSIS) for observing and assessing IC. Different from other IC assessment procedures, DSIS can track both learners' actual and scaffolded capabilities through mediation to assess what they can do alone and with support. Moreover, DSIS can elicit learners' authentic outcomes by focusing on those instances where troubles arise in interaction. Thus, learners were forced to negotiate and manage the interactional mechanisms rather than simply use prompted practices. Van Compernelle also called for longitudinal studies to track the growth of learners' repertoire of interactive practices and their ability to deploy resources in a context-sensitive way over time.

Chapter 9 by Dunkle showcased the potentiality of another novel procedure, social deduction board games (SDGs), to assess L2 learners' interactional competence in formative assessment. Sixteen international students, whose English proficiency was rated as Intermediate Mid to Advanced High according to the ACTFL scale, participated in this study, engaging with either SDGs or traditional roleplays. Their performance on three target skills, i.e., negotiation, bluffing, and turn-taking, was compared. Findings showed that SDGs can elicit chatty utterances that more closely resemble native speaker speech, while roleplays tended to generate turns aligning more with monologues than dialogues.

Barth-Weingarten and Freitag-Hild in Chapter 10 proposed a systematic description of IC in terms of the generic organizational contingencies in interaction as basic assessment categories under the conditions of scarce resources. Using turn-taking (TT) as an exemplar, the authors developed a CA-informed rubric of relevant sequence-organizational key points to assess German EFL learners' TT skills. Results indicated that the rubric designed in this chapter can make finer distinctions in learners' TT skill profile, providing robust evidence for the validity of the assessment instrument. The authors also observed that learners' L1-specific IC practices may contribute to the smooth interaction with their peers and called for further investigation to explore the effects of the L1-specific IC practices on interaction with L2 native speakers.

The last part of the volume, *Using New Technologies to Assess Speaking*, consists of three chapters, displaying the potential applications of new technologies in supporting the design and administration of assessment instruments that operationalize an expanded speaking construct. Chapter 11 by Song and Hsu investigated the potential of using Virtual Reality (VR) for a classroom-based L2 oral assessment. Collecting 24 Korean learners' performances on two roleplay-like DCT tasks, Song and Hsu compared the rating scores on, the language functions elicited by, and learners' perceptions of VR and F2F modes. Results showed that the two modes elicited comparable scores and language functions, although, under the VR condition, the mean scores were slightly lower and two language functions (asking for information and negotiation of meaning) were significantly less used. Regarding learners' perceptions, they generally showed positive attitudes towards VR assessment in terms of providing a relaxing and enjoyable atmosphere.

Chapter 12 by Iwashita, May and Moore comprehensively reported on the affordances and constraints of computer-mediated assessment of L2 speaking and its recent developments with the use of technological advances. The authors proposed that, despite the limitations of technological environments to incorporate the assessment of IC, emerging technologies such as videoconferencing, web-based virtual environments and automatic speech recognition and interactive spoken dialogue systems will pave the way for the design of innovative speaking tasks which can enable multiple candidates to interact with each other.

The final chapter, by Salaberry and Burch, concludes this part as well as the whole volume with a critical reflection on the challenges of designing and implementing appropriate procedures to assess speaking ability instruments in the post-COVID-19 era. The authors argued that it is viable to integrate the theoretical development of the expanded speaking construct with logistical affordances or constraints and advocated a

bi-dimensional construct of IC, that is, generalizable competence (stable) and interactional repertoires (variable).

This state-of-the-art volume contributes to the field by offering new insights, both theoretically and empirically, for language teachers, oral test researchers, developers and practitioners to better assess speaking in context. Despite the fact that the co-constructed nature of interaction raises thorny issues of validity and generalizability, the present volume seeks to break new ground by introducing innovative approaches to eliciting (DSIS, Chapter 8; SDGs, Chapter 9) and analyzing (MCA, Chapters 2 and 6) co-constructed performances and exploring in great depth raters' practices in recognizing IC-related factors (Chapters 5 and 6). For instance, Chapter 5 promisingly reports that raters can justify their decisions when awarding a separate score to an individual test-taker on paired performances featuring mutual accomplishment, adding robust evidence to the feasibility of assessing the co-constructed nature of paired or group speaking tests.

Another strength of the volume is the due attention paid to the role of nonverbal behavior (embodiment) in interaction from different perspectives (Chapters 3, 4, 10 and 11). For example, Chapter 3 lays a solid foundation for the incorporation of NVB into the speaking construct and provides a pioneering attempt at operationalizing NVB in assessment rubrics, and Chapter 10 presents a comprehensive framework of conventions covering kinetic transcription which makes the fine-grained analysis of NVB possible.

Despite its merits, this volume does have some shortcomings, just like any other publication. First and foremost, this volume brings into focus the concept of IC within the expanded conceptualization of the speaking construct and explores in detail some mechanisms like TT and repair, but does not present an overarching framework specifying the full set of components of the IC construct for speaking assessment, which still makes IC elusive. Certainly, this mission is daunting and requires continuing and joint efforts of the language testing community and language teachers. Second, the order of Part 2 (Collecting and Rating Speaking Data) and Part 3 (Designing Speaking Assessment Tests) should be, arguably, swapped for the convenience of reading, given that test design is expected to precede data collection and rating. Moreover, there are some minor issues with proofreading. For instance, "This volume is divided into three parts" (p. 12) should be four parts.

Overall, this volume contributes to the literature by focusing on both established and new speaking test instruments and presenting cutting-edge research with fine-grained analysis, thus serving as illuminating and inspiring cases for future investigation. Therefore, *Assessing Speaking in Context: Expanding the Construct and its Applications* is

highly recommended for a wide audience such as testing professionals and applied linguists as well as language teachers and graduate researchers.

*Reviewed by Ying Chen and Qi Lu*  
Ocean University of China

### Acknowledgements

This work was supported by the National Planning Office for Philosophy and Social Sciences of the People's Republic of China [20BYY108]. We would like to express our deep and sincere gratitude to the PLTA editors for their constructive and detailed feedback on this book review.

### References

- Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing*, 20(4), 369-383. <https://doi.org/10.1191/0265532203lt264oa>
- Fan, J., & Yan, X. (2020). Assessing speaking proficiency: A narrative review of speaking assessment research within the argument-based validation framework. *Frontiers in Psychology*, 11:330. <https://doi.org/10.3389/fpsyg.2020.00330>
- Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, 15(3), 219-236. <https://doi.org/10.1080/15434303.2018.1453816>
- He, A. W., & Young, R. (1998). Language proficiency interviews: A discourse approach. In R. Young & A. W. He (Eds.), *Talking and Testing: Discourse Approaches to the Assessment of Oral Proficiency* (pp.1-24). Amsterdam: John Benjamins. <https://doi.org/10.1075/sibil.14.02he>
- Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal*, 70(4), 366-372. <https://doi.org/10.1111/j.1540-4781.1986.tb05291.x>
- Lam, D. M. K. (2018). What counts as "responding"? Contingency on previous speaker contribution as a feature of interactional competence. *Language Testing*, 35(3), 377-401. <https://doi.org/10.1177/0265532218758126>
- McNamara, T. (1997). Interaction in second language performance assessment: Whose performance? *Applied Linguistics*, 18(4), 446-466. <https://doi.org/10.1093/applin/18.4.446>

- May, L., Nakatsuhara, F., Lam, D., & Galaczi, E. (2020). Developing tools for learning oriented assessment of interactional competence: Bridging theory and practice. *Language Testing*, 37(2), 165-188. <https://doi.org/10.1177/0265532219879044>
- Zhang, L., & Jin, Y. (2021). Assessing interactional competence in the computer-based CET-SET: An investigation of the use of communication strategies. *Assessment in Education: Principles, Policy & Practice*. <https://doi.org/10.1080/0969594X.2021.1976107>