



## Introduction from the Guest Editor

*Zohreh R. Eslami*<sup>1</sup>

The centrality of pragmatics in communication has precipitated research in several aspects of interlanguage pragmatics. However, most of the research has focused on instructional aspects and investigated the effects of various instructional approaches on learners' pragmatic competence development. Only recently have issues related to pragmatic assessment attracted serious attention from second language acquisition (SLA) researchers'. Despite the recognized necessity of teaching pragmatics, assessment of second language (L2) pragmatics is still a very young field of inquiry, awaiting further research and development (McNamara & Roever, 2006).

This thematic issue on pragmatic assessment covers theoretical and empirical investigations that explore important aspects related to the assessment of pragmatic competence. In the opening paper, **Cohen** presents an overview of classroom based pragmatics assessment. This paper is a revised version of chapter 14 from Ishihara and Cohen (2010). It examines tasks designed to assess both comprehension and production of speech acts and considers various ways to collect student responses – oral role play, written discourse as if spoken, multiple-choice, and short-answer responses. It focuses on practical issues including how to make the speech act situations realistic and how to rate for key aspects. The assumption being made is that classroom teachers tend to avoid extensively assessing pragmatics for various reasons. Nonnative teachers may especially feel that they themselves are incapable of judging what constitutes correct behavior. In the second contribution, **Bardovi-Harlig and Shin** focus on the whys and hows of assessing L2 pragmatic knowledge. Bardovi-Harlig and Shin explain that pragmatic assessment measures proposed by different studies were mainly researcher tools and not intended to be used to make decisions about test takers. However, in typical language assessment contexts, tests are normally used to make decisions about test-takers (Bachman & Palmer, 2010). Their study addresses how particular uses of L2 pragmatic assessment guide the design and development of pragmatic measures and motivate scoring.

Two contributions in this issue explore rater effects in pragmatic assessment. The study by **Liu and Xie** focuses on minimizing the impact of rater inaccuracy or bias on ratings of speech acts elicited by written discourse completion tests (WDCT). The study uses many-facet Rasch measurement. Based on the Myford & Wolfe (2003; 2004) model, five

---

*1Texas A&M University, USA. E-mail:zeslami@tamu.edu*



types of rater effects are investigated and discussed: leniency/severity, central tendency, randomness effect, halo effect, and differential leniency/severity. Results revealed significant differences in terms of rating severity, with a general tendency towards leniency. Though the raters could effectively and consistently employ the rating scales in their ratings, some showed a certain degree of halo effect. Most raters were also found to exhibit bias across both traits and testees. **Tajeddin and Alemi** focus on how rater training may affect rater accuracy and bias in pragmatic assessment. The study aimed to investigate whether a training program focused on pragmatic rating would diminish bias of non-native English speaker (NNES) ratings and have a beneficial effect on the accuracy of NNES ratings of refusal production as measured against native English speaker (NES) ratings. To this end, 50 NNES teachers rated EFL learners' responses to a written discourse completion task (WDCT) before and after attending a rating workshop. The same WDCT was rated by 50 NES teachers who functioned as a benchmark. Before training, non-native ratings as measured against the native benchmark in terms of mean, SD, mean difference and native-non-native correlation revealed that non-native raters tended to be highly divergent in their ratings. Rater training led to more accurate and consistent ratings. To measure rater bias, a FACETS analysis was run. FACETS results showed that both before and after training, many of the raters were outliers. The findings reveal that rater training can positively influence non-native teachers' ratings of pragmatic performance and increases the level of consistency in their scores.

**Ishihara and Chiba's** study is a qualitative study exploring the benefits and limitations of teacher-based and interactional assessment of young learners' pragmatic development. A variety of teacher-based assessment instruments are used in this study and the outcome of the assessment instruments are compared with the analysis of audio- and video-recorded classroom interactions. It is shown that the findings from the teacher-based assessment were not necessarily compatible with those from the interactional assessment. Some potential limitations of the teacher-based assessment instruments provided incomplete, inconsistent, or inaccurate data at times. On the other hand, the interactional data served as a valuable source of information, revealing the dynamic learning process in collaborative class interactions. The discourse data enabled the researchers to scrutinize the interaction recurrently over time and to analyze evidence of the learners' pragmatic competence. However, the laboriousness of the transcription process involved in analyzing interactional data, prevents interactional assessment from becoming a realistic instrument in most classroom contexts. The findings reveal that teacher-based and interactional assessment inform each other regarding the learning outcome and the process of knowledge co-construction and if used in combination, they may be able to capture both the learning outcome and the process of knowledge co-construction more effectively.

The final two papers in the volume delve into the use of different types of DCT elicitation techniques in pragmatic assessment. **Usó-Juan and Martínez-Flor** research focuses on the use of interactive discourse completion tasks to assess learners' performance when complaining and apologizing. The potential for alternative techniques to obtain learners' information as regards the planning and execution of speech act production is also



highlighted by the creation of a retrospective verbal report that may be used in combination with the elicitation method being elaborated. In this paper, data collection techniques employed in interlanguage pragmatics are reviewed. Then, the authors explain the particular instruments they used to assess learners' knowledge of complaints and apologies. Finally, concluding remarks and pedagogical implications are suggested concerning the use of the proposed techniques as both test instruments for researchers as well as pedagogical teaching materials for instructors in the English as a second/foreign language.

In the last paper in this volume, **Eslami and Mirzaei** draw on elicited speech act data collected from Persian language users to extend the research on data collection measures in pragmatic instruction, research, and assessment to non-Western languages. The researchers acknowledge that although most speech act research has relied on elicited data, it is not clear whether data elicitation techniques such as the discourse completion test are cross-culturally valid means of data collection and assessment (Rose, 1994). The authors assert that while several studies have dealt with the validity of DCTs in Western contexts (e.g., Rintell & Mitchell, 1989; Bardovi-Harlig & Hartford, 1993), very few have addressed the issue of cross-cultural validity of these pragmatic assessment instruments (except Rose, 1994). In order to fill the gap in the literature, this paper reports the results of a study that used two discourse completion task types (Oral DCT vs. Written DCT) in a non-Western language (i.e. Persian) compared to English. The findings show that response length, range and content of the expressions, level of formality and directness, and the formulas used were more noticeably different in Persian data compared to English data. The authors attribute this finding to the fact that Persian language has markedly noticeable differences between its oral and written variety. Furthermore, the correlation between the appropriateness scores of the two methods for Persian data was not high enough to support the claim that they measure exactly the same trait and thus construct validity of these elicitation tasks cannot be established. The findings indicate that more work is needed to both extend the scope of speech act studies to non-Western languages and also refine the methodologies used in pragmatic research and pragmatic assessment.

Bachman, L. F., & Palmer, A.S. (2010). *Language Assessment in Practice*. Oxford: Oxford University Press.

Bardovi-Harlig, K. & Hartford, B. S. (1993). Refining the DCT: Comparing open questionnaires and dialogue completion tasks. *Pragmatics and Language Learning*, 4, 143–165.

Ishihara, N. & Cohen, A. D. (2010). *Teaching and learning pragmatics: Where language and culture meet*. Harlow, England: Longman Applied Linguistics/Pearson Education.

McNamara, T. & Roever, C. (2006). The Social Dimension of Proficiency: How Testable Is It?. *Language Learning*, 56, 43–79.

Myford, Carol M., & Wolfe, Edward W. (2003). Detecting and measuring rater effects



- using Many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-422.
- Myford, Carol M., & Wolfe, Edward W. (2004). Understanding Rasch measurement: detecting and measuring rater effects using Many-facet Rasch measurement: Part II. *Journal of Applied Measurement*, 5(2), 189-227.
- Rintell, E. M., & Mitchell, C. J. (1989). Studying requests and apologies: An inquiry into method. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies* (248-272). Norwood, NJ: Ablex
- Rose, K. R. (1994). On the validity of DCTs in non-western contexts, *Applied Linguistics*, 15, 1-14.