



**Title of Project:**

Exploring Construct and Practicality of an Interactional  
Test for L2 Oral Pragmatic Performance

**Researcher:**

Naoki Ikeda  
University of Melbourne  
[na02ikeda@hotmail.com](mailto:na02ikeda@hotmail.com)



Naoki Ikeda

**Research Supervisors:**

Dr. Carsten Roever  
University of Melbourne

Dr. Ute Knoch  
University of Melbourne

---

**Project Summary**

**Motivation for the Research**

Pragmatic competence is one of the components of communicative competence models (Bachman & Palmer, 2010; Canale 1983; Canale & Swain, 1980; Purpura, 2004), and accordingly is a crucial part of human communication in real life (Roever, Fraser, & Elder, 2014). It should therefore be a key domain in language assessment (Kasper & Ross, 2013). Assessment of second language (L2) pragmatics has been developed by conceptualizing the test construct, as well as how it is conceptualized in task formats. Traditionally, the focus of pragmatics has been on offline knowledge of speech acts isolated from interaction. Despite considerable groundwork in pragmatics research, the findings in the literature are constrained by the narrowly defined construct and task design that restrict participants' performances to pre-planned scenarios, limiting conclusions about how test takers pragmatically perform in authentic discourse contexts.

Recent arguments (Kasper & Ross, 2013) and empirical studies (Grabowski, 2009; Youn, 2013) in the field, however, have highlighted performance-based assessments eliciting L2 speakers' oral extended discourse, which would allow for a broader conclusion about test takers' L2 pragmatic performances and for a stronger extrapolation from task performance to reality. Although informative about test takers' pragmatic abilities, performance-based assessments would negatively impact practicality (McNamara & Roever, 2006), for which suggestions should be made for actual practitioners of the assessments. Another recent activity worthy of note in the field is validation, which examines the extent to which a proposed score interpretation and use is justified and how so. Recent studies on validation in the field of L2 pragmatics assessment (e.g., Roever, et al., 2014), although still very limited, showed a stronger awareness of test validation, which informs test users of what the test scores mean and how useful the test scores are (Chapelle, 2008; Kane, 2006). With the insights provided by the previous studies, the researcher designed and evaluated an oral performance-based assessment instrument of L2 pragmatics, following an existing argument-based framework (Chapelle, 2008; Kane, 2006; Knoch & Elder, 2013). Unlike most validation studies, this study also addressed how the designed instrument can be implemented practically while avoiding construct under-representation (Messick, 1996).



### **Research Questions**

1. Investigation of measurable construct of oral pragmatics: What abilities of L2 oral pragmatics utilized in extended discourse contexts are measurable in inter-personal settings at university?
2. Test validation: To what extent is the procedure for assigning test scores appropriate? To what extent does the assessment yield test results consistent across assessment contexts (raters, test tasks, and test sets)? To what extent can test takers' test results be attributed to the construct of pragmatic abilities utilized for language activities at university?
3. Practicality: How and to what extent can monologue tasks serve as an alternative to dialogue tasks?

### **Research Methodology**

This study involved multiple phases of research activities and a range of analytical methods to address the research questions. First, the test tasks were developed through five stages of drafting and pilot studies, including discussions with domain experts. The tasks simulated university situations where a student addresses professors, administrators, and classmates to obtain assistance for the student's academic work. Two parallel test sets (with each set comprising three dialogue tasks and three monologue tasks) were created. The dialogue tasks simulated conversational situations where a student (test taker) physically interacts with an addressee. The monologue tasks simulated situations requiring a student to leave a voice-message to the intended addressee. Once the test tasks were finalized, the test administration in the main study was subsequently administered to 67 L2 students in Australia comprising university students (N=44) and pre-entry students (N=23), each of whom were assigned either Test Set 1 or Test Set 2. The pre-entry students were studying at a language school to satisfy the English language requirement for university admission. They did not have experiences at an English-medium university and had little exposure to English-speaking environments. The collected data also included the test takers' perceived similarity between their task and authentic performances and their self-assessed task performances (both of which were elicited by Likert scale questionnaires) in addition to their audio-recorded task performance samples.

Prior to the rating, the researcher developed the rating criteria based on the researcher's review of related literature and qualitative discourse analyses of the test takers' collected speaking samples, which identified features frequently seen in the test takers' performances and discriminating them. The findings were used to inform the rating rubrics, which comprised the six criteria (a) to (f) (described in the next section) with four band levels for each. The features summarized in the six criteria addressed research question 1. Generally, the lower scores (i.e., 2 and 1) were awarded to performances where negative features outweighed positive features, whereas the higher scores (i.e., 4 and 3) were awarded to those showing the reversed case. The rating was conducted by three raters, including the researcher himself. The test scores allowed for descriptive statistics, reliability estimates, multi-faceted Rasch analysis, and correlational analysis. The test takers' perceptions were analyzed quantitatively. The results of the quantitative analyses were evaluated to address research Questions 2 and 3.

### **Summary of Findings**

The qualitative discourse analyses of the test takers' task performances led to setting six criteria (a) to (f) summarized below, each of which discriminated the test takers in the current study. Performances of pragmatically competent test takers were constructed by (a) adequate social actions tailored for the context from opening through to closing; (b) smoothly and clearly delivered contents with sound variation and controlled repair; (c) linguistic resources varied and employed naturally to deliver intended meanings, minimizing the addressee's effort to understand what the speaker was trying



to say; (d) linguistic resources varied and employed naturally to mitigate imposition; (e) understanding of the discourse context and use of varied patterns of evaluative feedback well-tailored for the context without noticeable scaffolding from the interlocutor, and (f) naturally taking and releasing conversation turns so that the interlocutor understands when to take turns. Criteria (e) and (f) were used exclusively for dialogue performances. Generally, pragmatically competent test takers showed stronger evidence for their competencies. However, performances of less competent test takers were characterized by inconsistent evidence of the positive features described above and/or negative features undermining their performances.

The Rasch analysis showed the raters' internal consistency in rating. Although they showed a difference in severity in rating the test takers' performances, the severity difference was much smaller than the Rasch-estimated separation of the 67 test-takers, implying that the impact of raters' severity difference on test-taker scores was small. It was also confirmed that the raters operated the assessment band levels appropriately. For the tasks, inter-item correlations were high. The two paralleled test sets showed comparability in characteristics of test takers, raters, and rating criteria. These results, combined with the results of the raters' performances, suggest that the assessment can yield test results consistent across assessment contexts. Another important result was found regarding the test takers' Rasch-estimated pragmatic ability. Some university students, who were deemed pragmatically more competent than pre-entry students because of their experiences at English-medium universities and because they had higher proficiency levels, were outperformed by pre-entry students. However, overall the test takers were separated according to their proficiency levels and exposure to the language activities in the target domain as identified in the literature. The test takers perceived their language use in the tasks to be similar to that in reality. Highly proficient university students in particular perceived a stronger similarity between their task and authentic performances. Their self-assessment of task performance was moderately correlated with their Rasch-estimated ability, as expected for a non-test criterion. The test takers' Rasch-estimated abilities under the dialogue and monologue conditions were highly correlated, suggesting that test takers who performed positively under the monologue condition would perform just as positively under the dialogue condition.

### **Implications**

Although the methodology and the rationale are different among the existing studies on assessing L2 pragmatics, they have a common role of addressing a crucial component of human communication, which should ultimately contribute to enriching L2 learners' language activities. The current study has addressed the same goal as the literature, from the perspective of language assessment, by targeting L2 students' pragmatic abilities for language activities at an English-medium university. First and foremost, the constructs of pragmatics targeted in the study can be assessed, as they discriminate the test takers. Overall, the test scores generated from the designed assessment were indicative of L2 students' oral pragmatic abilities for university activities.

Secondly, the test takers' task performances, which have been qualitatively and quantitatively described in the study, have highlighted the meaningfulness of pragmatics assessments at both university pre-entry and post-entry stages. In the study, 19 university students were regarded as pragmatically competent as their average scores were 3 or above (out of four band levels). For the remaining 48 test takers (including 25 university students and 23 pre-entry students), their average scores were below 3, which indicated that their pragmatic performances were more negative than positive, according to the rating criteria. The results also imply that being a university student in an English-medium context does not necessarily mean that their pragmatic abilities are sufficient for language activities in that domain. Based on the qualitative and quantitative analyses, the study has



found that performances of all of the pre-entry students, as well as the substantial number of university students, reserve much room for improvement. Students could be aware of the need for improvement through a pragmatics assessment.

For the practice of a performance-based assessment of oral pragmatics, the study suggested a possible use of a monologue assessment. As confirmed by the high correlation between test takers' abilities estimated from the dialogue and the monologue data, both types of assessment instruments functioned similarly, in separating and ranking L2 students according to the pragmatic abilities. Dialogue assessments would allow for providing more comprehensive diagnostic information, including interactional criteria, but at the expense of increased resource intensiveness in its test administration. Because available resources and what to prioritize differ depending on individual assessment contexts, it would not be possible to provide an absolute solution for the appropriate balance between instrument practicality and construct coverage of pragmatics. However, if diagnosis of test takers' features that are exclusively seen in dialogue performances is not a central concern, the monologue assessment could be an alternative to the dialog assessment.



## References

- AERA, APA, & NCME (1985). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.
- Ahn, R. C. (2005). *Five measures of interlanguage pragmatics in KFL (Korean as a foreign language) learners* (Unpublished doctoral thesis). Honolulu, HI: University of Hawai'i at Manoa.
- Al-Gahtani, S. (2010). Request made by L2 learners of Arabic: *Pragmatic development, methodological comparison, and politeness* (Unpublished doctoral thesis). Melbourne, Australia: University of Melbourne
- Al-Gahtani, S., & Roever, C. (2012). Proficiency and sequential organization of L2 requests. *Applied Linguistics*, 33(1), 42-65.
- Al-Gahtani, S., & Roever, C. (2013). 'Hi doctor, give me handouts': Low-proficiency learners and requests. *ETL Journal*, 67(4), 413-424.
- Atkinson, J., & Heritage, J. (Eds.). (1984). *Structure and social action*. Cambridge, UK: Cambridge University Press.
- Austin, J. L. (1962). *How to do things with words*. Oxford, UK: Oxford University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1-34.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford, UK: Oxford University Press.
- Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning*, 63(1), 68-86.
- Bardovi-Harlig, K., & Hartford, B. (Eds.). (2005). *Interlanguage pragmatics: Exploring institutional talk*. Mahwah, NJ: Erlbaum.
- Blum-Kulka, S., House, J., & Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in human science* (3rd ed.). New York, NY: Routledge.
- Bouton, L. F. (1999). Developing non-native speaker skills in interpreting conversational implicatures in English: Explicit teaching can ease the process. In E. Hinkel (Ed.), *Culture in second language teaching and learning* (pp. 47-70). Cambridge, UK: Cambridge University Press.
- Brennan, R. L. (2013). Commentary on "Validating the interpretations and use of test scores". *Journal of Educational Measurement*, 50(1), 74-83.
- Brown, A. (2005). *Interviewer variability in oral proficiency interviews*. Frankfurt, Germany: Peter Lang.



- Brown, J. D. (2001). Six types of pragmatics tests in two different contexts. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 301-325). New York, NY: Cambridge University Press.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. New York, NY: Cambridge University Press.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 2-27). London, UK: Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254-272.
- Chapelle, C. A. (2008). The TOEFL validity argument. In C. A. Chapelle, M. E. Enright, & J. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 319-350). New York, NY: Routledge.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). Test score interpretation and use. In C. A. Chapelle, M. E. Enright, & J. Jamieson (Eds.), *Building a validity argument for the Test of English as a Foreign Language* (pp. 1-25). New York, NY: Routledge.
- Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Chomsky, N. (1965). *Aspect of the theory of syntax*. Cambridge, MA: MIT Press.
- Coulmas, F. (1979). On the sociolinguistics relevance of routine formulae. *Journal of Pragmatics*, 3, 239-266.
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.
- Cohen, A. D., & Olshtain, E. (1993). The production of speech acts by EFL learners. *TESOL Quarterly*, 27(1), 33-56.
- Crystal, D. (1997). *The Cambridge encyclopedia of language* (2th ed.). New York, NY: Cambridge University Press.
- Demeter, G. (2007). Role-plays as a data collection method for research on apology speech acts. *Simulation & Gaming*, 38(1), 83-90.
- Ducasse, A. M. (2010). *Interaction in paired oral proficiency assessment in Spanish*. Frankfurt, Germany: Peter Lang.
- Educational Testing Service. (2014). *TOEFL iBT scoring guides (rubrics) for speaking responses*. Retrieved April 2, 2017 from [https://www.ets.org/s/toefl/pdf/toefl\\_speaking\\_rubrics.pdf](https://www.ets.org/s/toefl/pdf/toefl_speaking_rubrics.pdf)
- Filipi, A. (2015). Authentic interaction and examiner accommodation in the IELTS speaking test: A discussion. *Papers in Language Testing and Assessment*, 4(2), 1-17.
- Galaczi, E. (2013). Interactional competence across proficiency Levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics*, 35(5), 1-23.
- Gumperz, J. (1982). *Discourse strategies*. Cambridge, UK: Cambridge University Press.





- Golato, A. (2003). Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics*, 24(1), 90-121.
- Grabowski, K. C. (2009). *Investigating the construct validity of a test designed to measure grammatical and pragmatic knowledge in the context of speaking* (Unpublished doctoral dissertation). New York, NY: Columbia University.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41-58). New York, NY: Academic Press.
- Hall, J. K., & Pekarek Doehler, S. (2011). L2 interactional competence and development. In J. K. Hall, J. Hellermann, & S. Pekarek Doehler (Eds.), *L2 interactional competence and development* (pp.1-15). Bristol, UK: Multilingual Matters.
- He, A. W., & Young, R. F. (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*. Amsterdam, The Netherlands: John Benjamins.
- Heritage, J. (1984). *Garfinkel and ethnomethodology*. Bristol, UK: Policy Press.
- House, J. (1996). Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness. *Studies in Second Language Acquisition*, 18, 225-252.
- Hudson, T. D. (2001). Indicators for pragmatic instruction. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 283-300). Cambridge, UK: Cambridge University.
- Hudson, T., Detmer, E., & Brown, J. D. (1992). *A framework for testing cross-cultural pragmatics* (Technical report #2). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Hudson, T., Detmer, E., & Brown, J. D. (1995). *Developing prototypic measures of cross-cultural pragmatics* (Technical report #7). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Huth, T. (2010). Can talk be inconsequential? Social and interactional aspects of elicited second language interaction. *The Modern Language Journal*, 94(4), 537-553.
- Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Middlesex, UK: Penguin Books.
- International English Language Testing System. (n.d.). *IELTS speaking band descriptors (public version)*. Retrieved April 2, 2017 from [https://takeielts.britishcouncil.org/sites/default/files/IELTS\\_Speaking\\_band\\_descriptors.pdf](https://takeielts.britishcouncil.org/sites/default/files/IELTS_Speaking_band_descriptors.pdf)
- Ishida, M. (2009). Development of interactional competence: Changes in the use of ne in L2 Japanese during study abroad. In H. t. Nguyen & G. Kasper (Eds.), *Talk-in-interaction: Multilingual perspective* (pp. 351-385). Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i.
- Ishida, M. (2011). Engaging in another person's telling as a recipient in L2 Japanese: Development of interactional competence during one-year study abroad. In G. Pallotti & J. Wanger (Eds.), *L2 Learning as social practice: Conversation-analytic perspectives* (pp. 45-85). Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i.



- Iwashita, N., McNamara, T., Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning*, 51(3), 401-436.
- Johnson, M. (2001). *The art of non-conversation: A re-examination of the validity of the oral proficiency interview*. New Haven, CT: Yale University Press.
- Kane, M. T. (1992). An argument-based approach to validation. *Psychological Bulletin*, 112, 527-535.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38, 319-342.
- Kane, M. T. (2004). Certification testing as an illustration of argument-based validation. *Measurement: Issues and Practice*, 21(1), 31-41.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Westport, CT: Greenwood Publishing.
- Kane, M. T. (2013). Validating the interpretation and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, 18, 5-17.
- Kasper, G. (2000). Data collection in pragmatics research. In H. Spencer-Oatey (Ed.), *Culturally speaking: Managing rapport through talk across cultures* (pp. 316-341). New York, NY: Cassell.
- Kasper, G. (2006). Beyond repair: Conversation analysis as an approach to SLA. *AILA Review*, 19, 83-99.
- Kasper, G., & Dahl, M. (1991). Research methods in interlanguage pragmatics. *Studies in Second Language Acquisition*, 13, 215-247.
- Kasper, G., & Roever, C. (2005). Pragmatics in second language learning. In E. Hinkel (Ed.), *Handbook of research in second language learning and teaching* (pp. 317-334). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kasper, G., & Rose, K. R. (2002). *Pragmatic development in a second language*. Malden, MA: Blackwell.
- Kasper, G., & Ross, S. J. (2013). Assessing second language pragmatics: An overview and introductions. In S. J. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp.1-40). Basingstoke, UK: Palgrave Macmillan.
- Kim, Y. (2009). The Korean discourse markers –nuntey and kuntey in native-nonnative conversation: An acquisitional perspective. In H. T. Nguyen & G. Kasper (Eds.), *Talk-in-Interaction: Multilingual perspective* (pp. 317-350). Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i.
- Knoch, U. (2009). *Diagnostic writing assessment: The development and validation of a rating scale*. Frankfurt, Germany: Peter Lang.
- Knoch, U., & Chapelle, C. (in press, 2018). Validation of rating processes within an argument-based framework. *Language Testing*.
- Knoch, U., & Elder, C. (2013). A framework for validating post entry language assessment. *Papers in Language Testing and Assessment*, 2(2), 48-66.





- Kobayashi, H., & Rinnert, C. (2003). Coping with high imposition request: High vs. low proficiency EFL students in Japan. In A. Martinez-Flor, E. Uso-Juan & A. Fernandez Guerra (Eds.), *Pragmatic competence in foreign language teaching* (pp. 161-184). Gastello: Servei de Publicacions de la Universitat Jaume I.
- Lado, R. (1961). *Language testing: The construction and use of foreign language test*. London, UK: Longman.
- Leech, G. (1983). *Principles of pragmatics*. London, UK: Longman.
- Linacre, M. (2013). *A user's guide to FACETS Rasch-Model computer programs*. Retrieved August 20, 2015, from: Winsteps.com
- Linacre, J. M. (2015). Facets computer program for many facet Rasch measurement (Version 3.71.4). Beaverto, Oregon: Winsteps.com.
- Liu, J. (2006). *Measuring interlanguage pragmatic knowledge of EFL learners*. Frankfurt, Germany: Peter Lang.
- Lunz, M. E., Wright, B. D., & Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3, 331-343.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174
- McNamara, T. F. (1996). *Measuring second language performance*. London, UK: Longman.
- McNamara, T. F. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics*, 18(4), 446-466.
- McNamara, T. F., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: American Council on Education and Macmillan.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23, 13-23.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 29(1), 3-17.
- Mey, J. L. (2001). *Pragmatics: An introduction* (2nd ed.). Oxford, UK: Blackwell.
- Myford, C. M., & Wolfe, E. W. (2000). *Monitoring sources of variability within the test of spoken English Assessment System*. Princeton, NJ: Educational Testing Service.
- Myford, C. M., & Wolfe, E.W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-442.
- Norris, J. M., Brown, J. D., Hudson, T. D., & Yoshioka, J. K. (1998). *Designing second language performance assessment*. Honolulu, HI: University of Hawai'i Press.
- Niezgoda, K., & Roever, C. (2001). Grammatical and pragmatic awareness: A function of the learning environment? In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 63-79). Cambridge, UK: Cambridge University Press.
- Okada, Y. (2010). Role play in oral proficiency interviews: Interactive footing and interactional competencies. *Journal of Pragmatics*, 42(16), 1647-1468.



- Pearson. (2012). *PTE Academic score guide*. Retrived April 2017 from: [http://pearsonpte.com/wp-content/uploads/2014/07/PTEA\\_Score\\_Guide.pdf](http://pearsonpte.com/wp-content/uploads/2014/07/PTEA_Score_Guide.pdf)
- Pekarek Doehler, S., & Pochon-Berger, E. (2011). Developing 'methods' for interaction: A cross-sectional study of disagreement sequences in French L2. In J. K. Hall, J. Hellermann, & S. Pekarek Doehler (Eds.), *L2 Interactional competence and development* (pp. 206-243). Bristol, UK: Multilingual Matters.
- Purpura, J. E. (2004). *Assessing grammar*. Cambridge, UK: Cambridge University Press.
- Read, J. (2015). *Assessing English proficiency for university study*. New York, NY: Palgrave Macmillan.
- Roever, C. (2005). *Testing ESL pragmatics*. Frankfurt, Germany: Peter Lang.
- Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing*, 23(2), 229-256.
- Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463-481.
- Roever, C. (2012, November). *Assessing and teaching pragmatics in the language classroom*. Workshop conducted at the first conference of the Association for Language Testing and Assessment of Australia and New Zealand, Sydney, Australia.
- Roever, C. (2013). Assessment of pragmatics. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-8). Hoboken, NJ: Blackwell Publishing.  
DOI: 10.1002/9781405198431.wbeal0049
- Roever, C., Fraser, C., & Elder, C. (2014). *Testing ESL sociopragmatics: Development and validation of a web-based test battery*. Frankfurt, Germany: Peter Lang.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1-20.
- Sasaki, M. (1998). Investigating EFL students' production of speech acts: A comparison of production questionnaires and role plays. *Journal of Pragmatics*, 30(4), 457-484.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge, UK: Cambridge University Press.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.
- Searle, J. (1975). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 59-82). New York, NY: Academic Press.
- Stokoe, E. (2013). The (in)authenticity of simulated talk: Comparing role-played and actual interaction and the implications for communication training. *Research on Language and Social Interaction*, 46(2), 165-185.
- Tada, M. (2005). *Assessment of EFL pragmatic production and perception using video prompts*, (Unpublished doctoral dissertation). Philadelphia, PA: Temple University.
- Taguchi, N. (2002). An application of relevance theory to the analysis of L2 interpretation process: The comprehension of indirect replies. *IRAL*, 40, 151-176.



- Taguchi, N. (2012). Context, individual differences and pragmatic competence. New York, NY: Multilingual Matters.
- Ten Have, P. (2007). *Doing conversation analysis* (2nd ed.). Los Angeles, CA: SAGE Publications.
- Thomas, J. (1995). *Meaning in interaction*. London, UK: Longman.
- Timpe, V. (2013). *Assessing intercultural language learning*. Frankfurt, Germany: Peter Lang.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- Tran, V. T. T. (2014). *Second language interactional competence: Solicitation effects on suggestions* (Unpublished doctoral dissertation). Melbourne, Australia: University of Melbourne.
- Yamashita, S. O. (1996). *Six measures of JSL pragmatics* (Technical report #14). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Yoshitake, S. S. (1997). *Measuring interlanguage pragmatic competence of Japanese students of English as a foreign language: A multi-test framework evaluation* (Unpublished doctoral dissertation). Novata, CA: Columbia Pacific University.
- Youn, S. J. (2013). *Validating task-based assessment of L2 pragmatics in interaction using mixed methods* (Unpublished doctoral dissertation). Honolulu, HI: University of Hawai'i at Manoa.
- Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, 32(2), 199-225.
- Young, R. F. (2008). *Language and interaction: An advanced resource book*. Abingdon, UK: Routledge.
- van Lier, L. (1989). Reeling, writing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly*, 23(3), 489-508.
- Walters, F. S. (2007). A conversation-analytic hermeneutic rating protocol to assess L2 oral pragmatic competence. *Language Testing*, 24(2), 155-183.
- Weigle, S. C. (2010). Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability. *Language Testing*, 27(3), 335-353.
- Xi, X. (2008). Methods of test validation. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education: Language Testing and Assessment* (2nd ed., Vol. 7, pp. 2316-2335). New York, NY: Springer.
- Xi, X. (2015, September). *Revisiting constructs- Recent theoretical advances and applications to practical test design, development and validation*. Workshop conducted at University of Melbourne, Melbourne, Australia.