BRILL

Contrastive Pragmatics and Second Language (L2) Pragmatics: Approaches to Assessing L2 Speech Act Production

Naoko Taguchi
Professor, English Department, Northern Arizona University,
705 S. Beaver Street, PO Box 6032, Flagstaff, AZ 86001, U.S.A.
naoko.taguchi@nau.edu

Shuai Li
Associate Professor, Department of World Languages and Cultures,
Georgia State University, 25 Park Places, Suite 1933, Atlanta, GA 30303
sli12@gsu.edu

Abstract

This state-of-the-art paper discusses common approaches to the assessment of pragmatic competence. Two approaches have dominated the assessment practice in the field of second language (L2) pragmatics. One approach, rooted in the tradition of contrastive pragmatics, involves comparing and contrasting L2 learners' pragmalinguistic forms with those of native speakers to determine whether L2 forms approximate native speaker forms. The other approach, rooted in the tradition of performance-based assessment, involves using a rating scale to evaluate L2 pragmatic performance based on multiple criteria (e.g., clarity of intention, formality level of expressions, and interaction abilities). Focusing on the area of speech act assessment, this paper presents an overview of these two approaches, highlighting their advantages and disadvantages. By doing so, the paper intends to illustrate the interface between contrastive pragmatics and L2 pragmatics. The paper concludes with critical insights in terms of what is missing in these approaches under the current trend of globalization and intercultural communication.

Keywords

1   Introduction

Second language (L2) pragmatics is a subfield of second language acquisition
(SLA) that investigates L2 learners' ability to perform communicative func-
tions in a social context, how such ability develops over time, and what factors
affect the process of development (Taguchi and Roever, 2017; Taguchi, 2019).
The primary practice of the field has been to collect data on L2 learners' prag-
matic performance and to evaluate their performance so we can understand
their current stage of development.

When evaluating pragmatic performance, or more narrowly speech act per-
formance, two approaches have dominated the field's practice. One approach
(contrastive linguistics approach), rooted in the tradition of contrastive prag-
matics, involves identifying learners' linguistic strategies in speech acts and
comparing them with those of native speakers to see how their linguistic forms
approximate native speakers' forms. Within this approach, similarities to
native speaker forms are considered as a sign of learners' development, while
differences are considered to indicate their underdevelopment.

Another approach, rating scale approach, comes from the tradition of
performance-based language assessment. This approach involves evaluating
learners' speech acts by using a rating scale (holistic or analytic) that includes
a series of predetermined score bands. The most typical implementation of
this approach in the field has been to recruit native speaker raters to assign
scores on learners' speech acts based on a set of preconstructed rating cri-
terion, and interrater agreement is sought to confirm the reliability of their
scoring. Criteria in rating scales used in this approach often focus on pragmat-
ics concerns, such as degree of politeness, directness, and formality of speech
acts, as well as other dimensions (e.g., grammatical accuracy and aspects
of interaction). While a holistic rating scale is used to assign one score to a
pragmatic utterance based on an overall evaluation of all dimensions under
consideration, an analytic rating scale is used to assign multiple scores to a
pragmatic utterance based on multiple dimensions under investigation. These
two common approaches, both prominent in the field, have both advantages
and disadvantages.

In order to illustrate how the work in contrastive pragmatics has featured in
L2 pragmatics research, this state-of-the-art paper represents a comprehensive
review of the above-mentioned approaches when assessing pragmatic com-
petence. In the following, we first present the current understanding of the
construct of pragmatic competence by articulating what elements are entailed
in that construct. Then, we review existing speech act studies using the two
approaches to assessment – contrastive linguistics analysis and rating scale
method – and discuss which elements of pragmatic competence are assessed

using these approaches. We also summarize generalizations coming from empirical findings and discuss strengths and limitations of these approaches. Finally, we critically discuss the two approaches together by highlighting what they are lacking under the current trend of globalization and intercultural communication.

2   Background: Theoretical Construct of L2 Pragmatic Competence

What makes someone pragmatically competent? This question is fundamental when we define and operationalize pragmatic competence for assessment purposes. Researchers in SLA and Applied Linguistics are invested in clarifying the construct of pragmatic competence – what knowledge and skills are involved in the construct and how they interact with each other. The early definition of pragmatic competence draws on Thomas's (1983) two-dimensional conceptualization of the construct. Thomas defined pragmatic knowledge as consisting of two distinct yet complementary elements – *pragmalinguistics* and *sociopragmatics*. The former refers to the knowledge of linguistic resources for performing a communicative act. For example, when greeting someone, we need to know a variety of linguistic forms to perform this communication act (e.g., "Hi", "What's up?") (for a recent work on rituals in contrastive pragmatics, see Kadar & House, 2020). However, knowing these forms alone does not make us pragmatically competent. We need to have sociopragmatic knowledge of which forms to use to greet whom in what context. Hence, knowledge of social conventions and interactive norms is a critical part of pragmatic knowledge. Knowledge of pragmalinguistics and sociopragmatics together gives us capacity to perform a communicative act appropriately and effectively in a social situation.

Thomas's two-dimensional operationalization of pragmatic knowledge is also reflected in the theoretical models of communicative competence that appeared in the 1980s and 1990s. In Canale and Swain (1980) and Bachman and Palmer's (1996) models, Thomas's two dimensions are understood as two types of knowledge – functional and sociolinguistic knowledge. The former involves the knowledge of form-function mappings (i.e., knowledge of linguistic forms for a communicative function), while the latter extends the form-function mappings to contexts of use (i.e., selecting appropriate forms to use in a specific context). Hence, these models consider knowledge of form-function-context mappings as a core of pragmatic knowledge.

With the surge of interactional competence in the last two decades (Kasper, 2006; Young, 2011), the early conceptualization of pragmatic competence has changed drastically (for a summary, see Taguchi, 2018a, 2019). Pragmatic

competence is no longer considered as knowledge of form-function-context mappings alone; instead, pragmatic competence is understood to involve the ability to use the knowledge in a flexible and adaptive manner in interaction. Under this view, form-function-context mappings are not fixed or stable in discourse. Rather, they are contingent on an unfolding course of interaction and are collaboratively constructed among people participating in interaction. With the emphasis on interactional competence, pragmatic knowledge is currently viewed as *interactional resources*. As Young (2011) claims, participants draw on a number of linguistic and interactional resources to co-construct meaning with their interlocutors. Those resources include knowledge of form-function-context associations like register-specific linguistic forms and speech acts, but they also extend to interactional skills such as topic management, turn-taking skills, and repair. Critically, these resources are not the property of individual participants. They are resources shared among participants as they co-construct meaning in interaction. Hence, a fundamental aspect of pragmatic competence is the ability to adapt one's resources to the dynamic course of interaction and achieve a communicative act collaboratively with others.

Parallel to the recognition of interactional competence, recently the concept of learner agency has also influenced our understanding of pragmatic competence. LoCastro (2003) defines agency as a self-reliant capacity that affects one's behavior. It is now understood that L2 learners do not blindly conform to conventionalized pragmatic norms (i.e., normative form-function-context mappings). Rather, they are active agents who make their own linguistic choices and create social positions for themselves. When conventionalized local norms contradict their desired social identity, learners sometimes resist adopting those norms. Learners' enactment of agency has been supported by several studies (Ishihara and Tarone, 2009; Kim and Brown, 2014). These studies showed that American learners of Japanese and Korean resisted the local norm of using honorifics when talking to seniors because they valued the egalitarian social structure and wanted to express solidarity by opting out honorifics. Clearly, agency can shape learners' pragmatic performance. Knowing the normative form-function-context mappings and applying the knowledge to interaction is part of pragmatic competence, but deciding whether or not to actually use the knowledge is the learners' own choice. Based on their beliefs and values, learners make their own linguistic choices even when their choices do not conform to the normative form-function-context mappings in the local community.

This three-layered conceptualization of pragmatic competence is summarized in Taguchi (2019: 4) as follows:

> [p]ragmatic competence in the current era is best understood as a multi-dimensional and multi-layered construct that involves several knowledge and skill areas: (1) linguistic and sociocultural knowledge of what forms to use in what context; (2) interactional abilities to use the knowledge in a flexible, adaptive manner corresponding to changing context; and (3) agency to make an informed decision on whether or not to implement the knowledge in the community.

Considering these historical developments, assessment of pragmatic competence should address all three dimensions – knowledge of form-function-context mappings, interaction skills, and agency. In the following section, we review existing assessment literature using two approaches: contrastive linguistic analysis and rating scale approach. Our purpose is to illustrate which dimension of pragmatic construct is assessed under these approaches and what they reveal in terms of pragmatic development.

To identify appropriate empirical studies for our purpose, we turned to Nguyen's (2019) review of research methods in L2 pragmatics literature. Nguyen conducted database searches to identify 246 empirical studies in L2 pragmatics published since the 1980s. Among the 246 studies that Nguyen compiled, we focused on those studies that examined speech acts in spoken modality for two reasons. First, speech acts have been the most representative units of analysis in L2 pragmatics research. Second, the two focal approaches (i.e., contrastive linguistics and rating) have been mainly used to assess speech acts. By presenting a comprehensive critique of the two major approaches to the assessment of speech acts, this state-of-the-art paper intends to highlight the contribution of the field of contrastive pragmatics to the study of L2 pragmatics and discusses how the two fields can be complementary in advancing the current practice of L2 pragmatics assessment.

3   Approaches to L2 Speech Acts Assessment

3.1   *Contrastive Linguistics Approach to Speech Act Assessment*
Contrastive linguistics' approach to speech act assessment is most clearly observed in cross-sectional studies that compare speech act strategies across different L2 groups. A 'group' can be created based on a variety of factors such as L1 language and cultural background (including native and nonnative speakers), L2 proficiency, age, length of study, and course level. Researchers often compare types of strategies and linguistic forms used to perform speech acts

in context across L2 groups. Hence, under the contrastive linguistics approach, pragmatic competence is primarily viewed as learners' knowledge of form-function-context mappings.

A project that served as foundation of cross-sectional, contrastive linguistics research is Blum-Kulka, House and Kasper's (1989) Cross-Cultural Speech Act Realization Project (CCSARP), which documented speech act strategies across seven languages (i.e., German, Hebrew, French, Danish, and three varieties of English). Using a discourse completion test (DCT),[1] the researchers elicited requests and apologies from participants, and categorized the strategies they used to complete these requests using a uniform coding framework. Comparison of speech act strategies across language groups revealed how many speech act strategies exist in a language, which strategies are direct or indirect, and how those strategies vary depending on contextual parameters (e.g., speaker relationship, degree of imposition). For example, Hebrew speakers were found to prefer using direct request three times more often than Australian English speakers, who mainly relied on conventional indirect strategies. The CCSARP has been replicated in over a hundred of studies, which together provided empirical descriptions of linguistic patterns across language groups and speech act types (e.g., refusals, compliments, thanking, and complaints). The contrastive pragmatics approach is still prominent today. Chen (2010), for example, surveyed compliment and compliment response patterns across 13 languages, revealing culture-specific patterns of compliment strategies. Ogiermann and Bella (2020) also revealed cross-cultural differences in request strategies among English, German, Greek, Polish, and Russian speakers.

While the primary contribution of CCSARP is in the field of contrastive pragmatics, its impact (in terms of replicability) extends to the field of L2 pragmatics. The DCT and coding framework have been adopted by a number of researchers who wish to evaluate L2 learners' speech act strategies and make a judgement about their pragmatic development (for a recent innovation of DCT, see Hashimoto and Nelson, 2020). The judgement about speech act development is often made by comparing learners' strategies with those of native speakers' strategies (baseline data). When learners use strategies similar to native speakers, they are judged to be in the more advanced stage of development; when their strategies differ from baseline strategies, they are considered to fall behind in development. For example, when making a request, learners of English were found to use longer justifications than native English speakers,

---

1 A typical DCT involves a situational scenario followed by blank space for participants to fill in their speech acts according to the situation. Participants are asked to imagine situation and produce the response as if they were performing the role indicated in the situational description.

sounding more verbose (Blum-Kulka et al., 1989). Learners also used a greater number of direct strategies and fewer syntactic/lexical mitigations (e.g., past tense and conditional clause; downgraders such as 'if you can' and 'possibly'). Similarities and differences between native speakers and learners' strategies have also informed potential areas of pragmatic failure coming from L1–L2 differences. For example, Maeshiba et al (1996) showed that Japanese-specific strategies such as apologizing when making a request often appeared in L1 Japanese learners' requests in English, indicating negative pragmatic transfer from L1 to L2.

While native vs. non-native comparison of speech act strategies has been the dominant practice in speech act assessment, cross-sectional research in the 1990s and beyond has expanded the scope to include studies comparing different L2 groups. Those studies analysed speech act data collected from two or more L2 groups of different proficiency levels, lengths of formal study, or duration of residence in the target language country. Any between-group differences found in the data were considered to signal certain levels of pragmatic competence. Findings about the effect of proficiency and length of study in L2 speech act performance are largely mixed and inconclusive. Some studies revealed a positive influence of proficiency and length of study on increased pragmatic competence, while others did not, suggesting that pragmatic competence is a complex construct influenced by a number of factors simultaneously (for a review, see Taguchi and Roever, 2017).

The contrastive linguistics approach to speech act assessment still permeates the field today. In more recent studies, researchers have used this approach to reveal characteristics of advanced speech act production by documenting which pragmalinguistic forms appear in more advanced-level learners' speech acts, which are missing in beginning-level learners' data (e.g., Al Masaeed et al., 2020; Felix-Brasdefer, 2007; Sabatéi Dalmau and Gotor, 2007; Rose, 2009; Taguchi, 2011a; Chang, 2010, 2016; Flores Salgado, 2011; Bella, 2012, 2014; Göy and Otcu, 2012; Liu and Ren, 2015; Savic, 2015; Lee, 2016). One generalization found in the data is that, in a high-imposition speech act addressed to someone in higher social status and larger social distance, advanced-level learners tend to use more indirect strategies with a greater number of external modification devices than lower-level learners, although their use of internal modifications (e.g., syntactic and lexical mitigations) is still minimal (for a review, see Taguchi, 2018). For example, Lee (2016) used a spoken DCT to elicit refusals from L2 learners of English in three grade levels in secondary schools. He found that, as the grade level increased, learners' use of direct strategies decreased (e.g., saying "I don't want to."). Instead, upper-level learners often used the indirect strategy of giving a reason for refusal, combining it with a statement of regret. The shift from direct to indirect strategies was also found

in Bella's (2012) study that examined requests by L2 learners of Greek at three proficiency levels. More advanced-level learners used indirect strategies at similar frequency of native speakers. They also used twice the number of modifications in a wider range, including imposition minimizers, considerators (e.g., if you can), and downtoners (e.g., perhaps). The diversification of strategies at the advanced-level was also found in Liu and Ren's (2015) study, which examined apologies among 40 first and third-year students of English in a Chinese university. Results showed that the high-proficiency group (third-year students) used upgraders (e.g., please) more often than the lower-proficiency group (first-year students) to intensify the intention of apology. The higher-proficiency group also acknowledged the likelihood of causing offense more often than their lower-proficiency counterparts.

In addition to the use of indirect strategies and wide-ranging pragmalinguistic repertoire, another characteristic of advanced speech act competence is found in the emergence of complex syntactic structures. For example, Rose (2009) found that upper-level learners of English used a gerund complement structure (e.g., Would you mind + a gerund). In Savic's (2015) study on Norwegian learners of L2 English, the 6th graders used a complex, bi-clausal structure (e.g., Do you have … that I can borrow?), while these forms were absent from the 2nd and 4th graders' productions. These findings add to the generalization that more advanced learners' speech acts are characterized by pragmalinguistic sophistication, as seen in the use of a variety of complex and compound structures, external/internal modifications, and syntactic/lexical mitigations.

3.2 *Pros and Cons of the Contrastive Linguistics Approach to Speech Act Assessment*

As described above, a common practice in the contrastive linguistics approach has been to elicit L2 speech acts using a structured instrument (e.g., DCT) and to categorize speech act strategies using a coding framework. Strategies and linguistic forms appearing in L2 data have been compared with those appearing in native speaker data to allow researchers to pinpoint similarities and differences between the two groups. The cross-sectional comparison has been extended to the comparison among L2 groups of different proficiency levels or lengths of formal study. General findings indicate that, as learners' proficiency increases, they tend to use more indirect strategies, along with longer, more complex syntactic structures (e.g., bi-clausal forms and complement structures) and a wider range of external/internal modifications. From these findings we can conclude that pragmatic competence in speech act production

is, in part, reflected in learners' ability to produce speech acts that are linguistically elaborate and complex, and furnished with a range of supportive moves and mitigation devices.

A primary advantage of the contrastive linguistics approach is that it has provided a means to systematically analyze pragmalinguistic strategies in L2 speech acts. Comparison of the strategies with native speakers and across L2 groups can clearly reveal which strategies are available in learners' current linguistic repertoire, which strategies are missing, and which strategies are overused or underused. These linguistic-level comparisons help us understand the level of learners' pragmatic competence in terms of the size of their repository of pragmalinguistic strategies and ability to select appropriate linguistic forms from the repository according to situations. Hence, among the dimensions of pragmatic competence described in the previous section, the contrastive linguistics approach can directly assess learners' knowledge of form-function-context mappings – which linguistic forms to use when performing a speech act in what social situation.

However, this exclusive focus on pragmalinguistic forms is the fundamental weakness of the contrastive linguistics approach. Using this approach, researchers can only assess learners' knowledge of speech act strategies; other performance-level features involved in speech act production – fluency, clarity of intention, and comprehensibility of utterances – are largely neglected. Yet these features are critical to determine the perlocutionary effect of a speech act because the listener's reaction depends on his/her understanding of the speech act. If the speech act is incomprehensible or its intention is too obscure, it does not lead to expected outcomes. More critically, under the current trend of interactional competence (Kasper, 2006; Young, 2011), a speech act is understood not as a unidirectional act *from* the speaker *to* the listener, but as a collaborative construction *between* the speaker and listener. Linguistic resources such as direct/indirect strategies and syntactic/lexical mitigations are not fixed, stable, or pre-determined in discourse; the decision regarding which strategies or mitigations to use is contingent upon the unfolding course of discourse. For example, indirect strategies and downgraders may appear initially in request-making, but the speaker might shift to more direct strategies with upgraders when the listener does not comply to his/her request. Similarly, the speaker can accomplish a request without saying anything; the listener may anticipate the speaker's move and offer to do the expected act before anything is said. These interactional moves in a speech act are never assessed using the coding framework and counting frequency of pragmalinguistic forms. Therefore, utterance-level analyses of speech act strategies

cannot account for learners' interactional abilities to co-construct a speech act with their interlocutors (see also Culpeper, Mackey, and Taguchi, 2018; House and Kadar, 2021).

Another limitation of the contrastive linguistics approach is that the comparison of L2 data with native speaker baseline does not align with the current discourse of intercultural communication and lingua franca framework, because it promotes a deficit-model of L2 pragmatic competence (e.g., Seidlhofer, 2011; Jenkins, 2015; Cogo and House, 2017). The current lingua franca literature argues that the use of native speaker norms should be abandoned when assessing L2 competence because learners do not always use target language to mimic native speakers; rather, they use the language in intercultural communication to achieve mutual understanding, develop personal relationships, and express identity with other nonnative speakers. Because speakers in intercultural communication attend to mutual intelligibility and rapport-building rather than attempting to imitate native speakers, native speaker norms do not serve as a reference point for assessment of L2 performance (Seidlhofer, 2011).

Related to this point, the idea of single, uniform native speaker norms has been criticized in the current discourse of multilingualism and transculturalism. The common practice of using native speaker data as baseline for comparison to L2 data is problematic because it overlooks variations among native speakers who come from different regional, educational, or generational backgrounds (Mori, 2009; Barron, 2019). Given these various backgrounds, when judging the appropriateness of L2 pragmatic behaviors, native speaker performance should not be regarded as having single, identical standards. However, despite the varied norms existing among native speakers, studies under the contrastive linguistics approach continue to use native speaker data as a single benchmark for assessing L2 speech acts. This problem is even more serious when considering that the majority of these studies collect data from only a small group of native speaker participants (typically no more than 30 participants) (e.g., Felix-Brasdefer, 2007; Taguchi, 2011a; Bella, 2012, 2014). These studies only consider age, gender, and educational level of native speaker participants to be comparable to L2 participants, without addressing who those native speakers are. These shortcomings of the contrastive linguistics approach will be revisited in the final part of this paper when we present future directions.

3.3   *Rating Scale Approach to Speech Act Assessment*

The rating scale approach is a widely adopted method in speech act assessment studies with a quantitative orientation. Unlike the contrastive linguistics approach, the rating scale approach is observed in research with both

cross-sectional and longitudinal designs. It is also represented in several sub-domains of L2 pragmatics research, including pragmatic development in instructed, uninstructed, and study abroad environments, as well as pragmatics testing, to name just a few.

Given the inherent quantitative nature of the rating scale approach, it is not surprising that this method was first proposed in research focusing on L2 pragmatics assessment. A foundational project in this area is Hudson, Detmer and Brown's (1992, 1995) project on developing a prototypical test battery for assessing L2 speech acts (i.e., requests, apologies, and refusals). This project created six assessment measures including a multiple-choice test, a written DCT, an oral DCT, a role play task, a DCT self-assessment task, and a role play self-assessment task. To evaluate written and oral production of speech acts, Hudson et al. developed 5-point analytic rating scales for assessing six aspects of speech act production: (1) use of correct speech acts, (2) use of formulaic expressions, (3) amount of speech/information, (4) levels of formality, (5) levels of directness, and (6) levels of politeness. The design of their rating criteria clearly reflects considerations of both pragmalinguistics and sociopragmatics. For example, the dimension of formulaic expressions mainly concerns the use of typical pragmalinguistic expressions in a given situation; on the other hand, the dimensions of directness, formality, and politeness address not only pragmalinguistic strategies, but also the sociopragmatic appropriateness of using those strategies in context. Moreover, Hudson et al.'s rating criteria were also informed by relevant research findings. For example, the dimension concerning amount of speech was included because excessive amount of speech was found as an indication of either circumlocution (due to low proficiency) or verbosity (due to high proficiency).

Hudson et al.'s project was influential in that a number of subsequent studies directly adopted their rating criteria or made minor revisions to it. While many studies in this group focused on pragmatics assessment (e.g., Yamashita, 1996; Brown, 2001; Hudson, 2001; Liu, 2006; Youn, 2007; Taguchi, 2011b), a few studies examined pragmatic development in L2 speech act production (e.g., Taguchi, 2011c). One minor revision in Hudson et al.'s rating criteria involved combining the dimensions of directness, formality, and politeness to create a holistic scale because these three dimensions are often difficult to separate (e.g., Liu, 2006; Taguchi, 2011b). Another minor revision made was to remove certain dimensions (e.g., amount of speech, formulaic expressions) from the original criteria in order to cater to the goals of individual studies (e.g., Taguchi, 2011b, 2011c).

As researchers continue to adopt and adapt Hudson et al.'s rating criteria, a major development has been to add a dimension of grammatical accuracy into speech act assessment (e.g. Sasaki, 1998; Taguchi, 2007, 2012; Grabowski,

2013; Li, 2014; Chen and Liu, 2016; Li, Taguchi, and Xiao, 2019; Xiao, Taguchi, and Li, 2019). While Hudson et al. (1995) made it clear that the grammaticality of speech act production was not part of their consideration (p.165), researchers have argued that grammar and pragmatics are inevitably interconnected (Kasper and Rose, 2002; Bardovi-Harlig, 2003). Such understanding is clearly reflected in the 5-point analytic rating criteria developed in Taguchi's (2012) study, where she assessed appropriateness (directness, politeness, and formality combined) and grammaticality (grammatical and lexical accuracy combined) using two separate scales. Another example comes from Grabowski's (2013) study in which she developed several 5-point analytic rating criteria for assessing learners' speech acts in role plays. The rating criteria were based on Purpura's (2004) theoretical model of language knowledge consisting of grammatical and pragmatic knowledge. Whereas grammatical knowledge subsumes grammatical form and grammatical meaning, pragmatic knowledge entails sociolinguistic (social norms and preferences), sociocultural (cultural norms and preferences), and psychological dimensions (affective stance and tone). It is also noteworthy that, unlike Taguchi (2012) and Grabowski (2013), who adopted separate rating criteria for grammatical accuracy and pragmatic appropriateness, several studies developed holistic rating criteria combining both dimensions. For example, Taguchi, Xiao and Li (2016) adopted a 6-point rating scale that simultaneously addressed appropriateness, grammaticality, and clarity of speech acts. Finally, the emphasis on pragmalinguistic forms and grammatical accuracy in evaluating speech act production is clearly featured among instructional studies that aimed at teaching specific pragmalinguistic strategies (e.g., Fukuya an Zhang, 2002; Li, 2012; Li and Taguchi, 2014). For example, Fukuya and Zhang (2002) designed a study to teach request-making forms in English. Pre- and post-written DCT data were collected. Learners' requests (collected via written DCT) were evaluated on the main criterion of appropriateness (i.e., use of target request-making forms), as well as the secondary criterion of grammaticality.

Another under-represented yet important development in the rating scale approach is a shift of assessment focus from *illocutionary force* to *perlocutionary effects* (Roever, 2005; Cohen and Shively, 2007). The rating criteria discussed in the previous section largely focus on assessing speech acts for their illocutionary force from the perspective of the speaker, that is, the extent to which the speaker's communicative intention is achieved; they do not consider the effects of the speaker's utterance on the part of the hearer, that is, how the hearer would respond to the speaker's utterance. To address this issue, Cohen and Shively (2007) evaluated requests and apologies (elicited via written DCT) on a 4-point rating scale focusing on the overall success of speech acts, that

is, the level of compliance (for requests) and forgiving (for apologies) on the part of the hearer. In another study, Roever (2005) evaluated speech acts (elicited through a computerized written DCT with rejoinders) based on a holistic scoring rubric concerned primarily with the likelihood of a learner utterance leading to a pre-determined rejoinder.

The attention to perlocutionary effects in the development of the rating scale approach echoes more recent theorizations of pragmatic competence informed by interactional competence (Kasper, 2006; Young, 2011; also see Section 2). Informed by the updated conceptualizations of pragmatic competence, there has been a trend to utilize interactive tasks such as role plays for assessing pragmatic competence. Correspondingly, researchers have developed new rating criteria to evaluate such interactive data attending to both the speaker and the hearer (Timpe, 2013; Youn, 2015, 2018a). The first attempt in this trend is Timpe's (2013) study that created a Skype-mediated role play task. She developed two sets of rating criteria (each containing 6 levels) for assessing pragmatic and discourse competencies. The rating scales on pragmatic competence tapped all aspects proposed by Hudson et al. (1995) except for the ability to use correct speech acts (because each of the role play scenarios involved different speech acts). The rating criteria for assessing discourse competence addressed multiple dimensions, including appropriateness in opening, pauses, turn-taking, cohesion/coherence, and closing. More recently, Youn (2015, 2018a) adopted a conversation analytic approach to develop and validate rating criteria. She evaluated L2 English learners' role play performance involving different speech acts (e.g., requests, refusals, and negotiations). Her rating criteria encompassed five dimensions: (1) contents delivery (i.e., clarity and fluency in turn delivery), (2) language use (i.e., use of appropriate pragmalinguistic and grammatical forms to achieve pragmatic functions), (3) sensitivity to situation (i.e., awareness of sociopragmatic characteristics of specific scenarios and offering appropriate explanations as needed), (4) engaging with interaction (i.e., ability to maintain interaction and establish shared understanding with interlocutors), and (5) turn organization (i.e., ability to follow appropriate turn-taking conventions). Youn (2018a) claimed that the last three criteria were most closely related to interactional organizations. Comparing Timpe and Youn's rating criteria, it is clear that they are not restricted to the traditional understanding of pragmatic competence consisting of pragmalinguistic and sociopragmatic components; instead, their criteria address co-construction of meaning in interaction.

In summary, the rating scale approach to speech act assessment has shown major developments corresponding to evolving conceptualizations of pragmatic competence. The foundational proposal by Hudson et al. (1995) rooted in a traditional view of pragmatic competence has been gradually enriched

by incorporating constructs reflecting (broadly defined) grammatical competence and various features of interaction. Still, in view of the most recent understanding of pragmatic competence as consisting of linguistic and sociocultural knowledge, interactional abilities, and agency, current practices of the rating scale approach fall short of attention to learner agency.

3.4 Pros and Cons of the Rating Scale Approach to Speech Act Assessment

There are several advantages associated with the rating scale approach. First, this approach affords a relatively high level of construct coverage, i.e., the extent to which targeted constructs are appropriately represented in an assessment instrument. Compared to the contrastive linguistics approach, the rating scale approach enables researchers to address a variety of dimensions of speech act production (e.g., aspects of interaction, meaning co-construction, and perlocutionary effects) in addition to form-function-context mappings. The rating scale approach thus allows researchers to evaluate speech act production in a more comprehensive manner to reflect the field's evolving theorization of pragmatic competence.

Second, a related advantage of the rating scale approach is that it allows flexibility in tailoring evaluation criteria according to specific research studies. This flexibility is manifested in several ways. First, researchers can add or remove certain dimensions of assessment based on their goals. The studies that added grammatical accuracy (e.g., Taguchi, 2012; Grabowski, 2013) and features of interaction (e.g., Timpe, 2013; Youn, 2015) into existing rubrics are good examples. Another example is the set of rating criteria proposed by Ishihara (2010) for classroom assessment of speech acts. Her comprehensive rating criteria encompassed seven dimensions, including sociocultural norms, organizations, directness/politeness/formality, grammar strategies, semantic moves, word choice, and tone. Flexibility of the rating scale approach can also be found in studies where researchers adjusted the relative weights of different dimensions of assessment. For example, Fukuya and Zhang (2002) and Li (2012) both assigned more weight to pragmalinguistic forms than to grammatical accuracy when evaluating L2 learners' requests.

Despite the aforementioned advantages, the rating scale approach is limited in several ways. Unlike the comparative linguistics approach, the rating scale approach is unable to provide detailed documentation of pragmalinguistic forms involved in speech acts. While it is possible that researchers can design scoring rubrics in such a way that scores can reflect the use or non-use of certain pragmalinguistic strategies (for example, see Fukuya and Zhang, 2002; Li, 2012), such scores do not show which specific strategies are used (or not used).

Conceivably, in studies examining a wide range of pragmalinguistic strategies, rating scores would at best be a very coarse means for understanding learners' mastery or non-mastery of particular form-function-context mappings. One possible solution to this problem would be to combine the rating scale approach and the contrastive linguistics approach in assessing speech acts. For example, Li (2014) adopted a holistic rating scale that simultaneously assessed realization of communicative intention, appropriateness, and grammaticality of requests produced by L2 Chinese learners over a semester abroad. After confirming the significant gain in speech act ratings from pre- to post-study abroad, he conducted a follow-up analysis using the contrastive linguistics approach to examine changes in learners' pragmalinguistic forms over time.

Another limitation of the rating scale approach is related to the flexibility that it affords in developing and adapting rating criteria. Admittedly, this approach allows us to adjust assessment rubrics based on evolving conceptualizations of pragmatic competence as well as specific research goals. However, such flexibility also makes it difficult to compare findings across studies, due to differences among researchers in operationalizing pragmatic competence (see the studies reviewed in Section 3.3). This issue is further complicated by the two different forms of rating criteria, that is, holistic and analytic. Studies in L2 performance assessment have revealed differences in rating scale functioning and rating processes due to this difference in rating criteria format (e.g., Barkaoui, 2010). Similar issues could exist in pragmatics assessment, and this poses challenges for researchers to compare results across studies using different rating rubrics.

Last but not least, just like the contrastive linguistics approach, the rating scale approach also relies on native speakers in assessment. With a few exceptions (e.g., Walters, 2007; Sydorenko, Mayard, and Guntly, 2014; Tajeddin and Alemi, 2014), the predominant practice in the field has been to recruit native speakers to evaluate learners' pragmatic production. Those raters may include experienced L2 instructors, graduate students, and researchers in the relevant field(s) (e.g., Roever, et al., 2014), as well as native speakers with little relevant teaching and/or research experience (e.g., Taguchi, 2011b). Usually, native speaker raters are instructed to rely on their intuition when evaluating learner production (e.g., Hudson et al., 1995; Liu, 2006); in other cases, native-likeness is part of the rating criteria (e.g., Roever et al., 2014). Either way, the assumption is that native speakers constitute a homogeneous group when it comes to interpreting rating criteria and assigning scores. This assumption, however, has been challenged by empirical evidence. For example, Taguchi (2011b) showed that native speaker raters coming from different cultural backgrounds (i.e., African American, Asian American, and Australian) brought

with them quite different perceptions regarding appropriateness, politeness, and formality of speech acts. Li et al. (2019) further reported that even native speakers with highly comparable cultural, educational, and professional backgrounds showed considerable variations in interpreting the same set of rating criteria. These findings echo recent critical reflections on the assumption of a uniform native speaker norm in the context of multilingualism and transculturalism, a point made earlier in discussing the disadvantages associated with the contrastive linguistics approach. The findings of Taguchi (2011b) and Li et al. (2019) also point to the importance of implementing appropriate rater training programs as part of the assessment process. While the field of L2 pragmatics assessment has yet to investigate the effects of rater training on rating processes and rating outcomes, one can be informed by relevant research on performance-based language assessment at large. In particular, rater training has been found effective in reducing the impact of rater-induced variances (e.g., different linguistic, cultural, and professional backgrounds) on rating outcomes (e.g., Kang, Rubin, and Kermad, 2019; Xi and Mollaun, 2011). If these findings were also replicated in L2 pragmatics assessment, it would provide empirical evidence supporting the expansion the pool of raters to include both native and non-native raters.

In summary, the strengths and weaknesses associated with the rating scale approach can be understood from two perspectives, namely, design and use of rating criteria. From the design perspective, the rating analysis approach allows a relatively high level of pragmatics content coverage as well as great flexibility in accommodating updated conceptualizations of pragmatic competence. However, the downside of this approach lies in the difficulty in documenting specific pragmalinguistic strategies as well as in comparing findings across studies. From the use perspective, while assigning scores based on a set of predetermined criteria may appear to be a fairly straightforward and efficient process, the assumptions underlying our practice of relying on native speakers' judgment have been challenged by empirical findings. Because excessive rater variations in interpreting the meaning of the same rating criteria may pose a serious threat to the quality of the data gleaned through the rating process, researchers may need to be cautious in interpreting their findings based on the rating scale approach.

4　　Conclusion and Future Directions

In light of the field's evolving understanding of pragmatic competence, this state-of-the art paper has reviewed and compared the contrastive linguistics approach and the rating scale approach in terms of their advantages and

disadvantages associated with speech act assessment. By doing so, this paper has illustrated the contributions of the field of contrastive pragmatics to second language pragmatics and vice versa, and has discussed how these two fields can complement each other and overcome limitations of the fields.

Considering pragmatic competence as consisting of pragmalinguistic and sociopragmatic components, the contrastive linguistics approach enables a systematic and detailed documentation of pragmalinguistic strategies that can be compared across participant groups and scenarios. Yet, this approach largely excludes attention to other important aspects of pragmatic performance such as clarity of intention and comprehensibility of speech. In comparison, the rating scale approach allows for a more comprehensive evaluation of speech acts by addressing multiple dimensions such as sociopragmatic knowledge and interactional competence. Hence, the rating scale approach is more advantageous in terms of the coverage of pragmatic construct, even though it falls short of its capacity in documenting pragmalinguistic strategies. Still, different rating criteria across studies make it difficult to compare findings across studies.

On the other hand, neither approach, as currently implemented in the field, has been able to accommodate the assessment of speech acts from the perspective of learner agency involved in pragmatic performance (LoCastro, 2003; Taguchi, 2019). Nevertheless, both approaches afford potentials when combined with other data collection methods. For example, researchers can elicit leaners' retrospective comments on the rationale underlying their speech act production (e.g., Taguchi, 2012). The verbal protocols can be coded and analyzed qualitatively and quantitatively (through specifically developed rating criteria) with a focus on leaners' agency in their choice of specific pragmalinguistic strategies.

The issue of learner agency directs us to consider the role of native speakers in assessing speech acts, a challenge that both approaches face. As previously discussed, adopting the assumption of uniform native speaker norms when assessing L2 learners' performance is problematic, because native speakers vary in their perceptions of politeness and appropriateness (Taguchi, 2011b; Li et al., 2019). Hence, we need to explore alternative baseline models for comparative analysis, and to investigate the role of non-native raters. Recruiting and training non-native speakers to evaluate speech acts would allow the field of L2 pragmatics assessment to better connect to the broader field of performance-based language assessment, which has demonstrated satisfactory rating behaviors among non-native raters that are comparable to native raters (e.g., Kim, 2009), particularly after going through rigorous rater training procedures (Xi and Mollaun, 2011; Kang et al., 2019). Nevertheless, it is importantly to articulate the purpose(s) of pragmatics assessment for specific contexts

and develop assessment strategies accordingly (e.g., which baseline model to adopt and whether to involve non-native raters). For example, in a lingua franca communication context where mutual understanding is a shared goal among speakers, and native speakers constitute only one of the stakeholder groups, it would be appropriate to adopt a set of locally negotiated norms rather than native speaker norms for assessment. To adopt the contrastive linguistics approach in this assessment context, it is necessary to identify locally recognized successful communicators (both native and non-native speakers) and use their performance as the baseline for gauging pragmatic competence. To adopt the rating scale approach in this context, we need to develop a set of shared assessment criteria among all relevant stakeholders. We can also invite qualified native and non-native raters who understand local norms to develop appropriate rating criteria. To this end, task-based pragmatics assessment with well-designed needs analysis in specific assessment contexts (Timpe-Laughlin, 2018; Youn, 2018b) offers a viable option.

Biographical Note

Naoko Taguchi is Professor in the Applied Linguistics Program at Northern Arizona University where she teaches courses in linguistics and second language acquisition. Her research interests include L2 pragmatics, technology-enhanced L2 teaching, intercultural competence, and English-medium education. She is the co-editor of *Journal of Applied Pragmatics*.

Shuai Li is Associate Professor, Associate Department Chair, and Chinese Program Coordinator in the Department of World Languages and Cultures at Georgia State University. His research interests include second language acquisition with a special focus on Chinese as a second language teaching and learning, interlanguage pragmatics, and L2 pragmatics assessment.

References

Al Masaeed, Khaled, Naoko Taguchi, and Mohammed Tamimi. 2020. Proficiency effects on L2 Arabic refusals: Appropriateness, linguistic strategies, and multidialectal practices. *Applied Pragmatics 2*(1): 26–53.

Bachman, Lyle F., and Andria S. Palmer. 1996. *Language Testing in Practice*. Oxford, UK: Oxford University Press.

Bardovi-Harlig, Kathleen. 2003. Understanding the role of grammar in the acquisition of L2 pragmatics. In: Alicia Martínez, Esther Usó Juan, and Ana Fernández Guerra

(eds.), *Pragmatic Competence and Foreign Language Teaching*. Castellon, Spain: Servei de Publicacions de la Univerisitat Jaume I, 21–44.

Barkaoui, Khaled. 2010. Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly 7*: 54–74.

Barron, Anne. 2019. Norms and variation in L2 pragmatics. In: Naoko Taguchi (ed.), *The Routledge Handbook of SLA and Pragmatics*. New York, NY: Routledge, 447–461.

Bella, Spyridoula. 2012. Pragmatic development in a foreign language: A study of Greek FL requests. *Journal of Pragmatics 44*(13): 1917–1947.

Bella, Spyridoula. 2014. Developing the ability to refuse: A cross-sectional study of Greek FL refusals. *Journal of Pragmatics 61*: 35–62.

Blum-Kulka, Shoshana, Juliane House, and Gabriele Kasper. 1989. *Cross-cultural Pragmatics: Requests and Apologies*. Norwood, NJ.: Ablex.

Brown, James Dean. 2001. Pragmatic tests. In: Gabriele Kasper, and Kenneth R. Rose (eds.), *Pragmatics in Language Teaching*. New York, NY.: Cambridge University Press, 301–325.

Canale, Michael, and Merrill Swain. 1980. Theoretical aspects of communicative approaches tosecond language teaching and testing. *Applied Linguistics 1*: 1–47.

Chen, Rong. 2010. Compliment and compliment response research: A cross-cultural survey. In Anna Trosborg (ed.), *Handbook of Pragmatics: Pragmatics across Languages and Cultures*. New York/Berlin: Mouton de Gruyter, 79–101.

Chang, Yuh-Fang. 2010. 'I no say you say is boring': the development of pragmatic competence in L2 apology. *Language Sciences 32*(3): 408–424.

Chang, Yuh-Fang. 2016. Apologizing in Mandarin Chinese: A study on developmental patterns. *Concentric: Studies in Linguistics 42*: 73–101.

Chen, Yuanshan, and Jianda Liu. 2016. Constructing a scale to assess L2 written speech act performance: WDCT and e-mail tasks. *Language Assessment Quarterly 13*(3): 231–250.

Cogo, Alessia, & Juliane House. 2017. Intercultural pragmatics. In: Anne Barron, Yueguo Gu, and Gerard Steen (eds.), *Routledge Handbook of Pragmatics*. London/New York: Routledge, 168–183.

Cohen, Andrew D., and Rachel Shively. 2007. Acquisition of requests and apologies in Spanish and French: Impact of study abroad and strategy-building intervention. *The Modern Language Journal 91*(2): 189–212.

Culpeper, Jonathan, Alison Mackey, and Naoko Taguchi. 2018. *Second Language Pragmatics: From Theory to Research*. New York: Routledge.

Félix-Brasdefer, J. César. 2007. Pragmatic development in the Spanish as a FL classroom: A cross-sectional study of learner requests. *Intercultural Pragmatics 4*: 253–286.

Flores Salgado, Elizbeth. 2011. *The Pragmatics of Requests and Apologies: Developmental Patterns of Mexican Students*. Amsterdam: John Benjamins.

Fukuya, Yoshinori, and Yao Zhang. 2002. Effects of recasts on EFL learners' acquisition of pragmalinguistic conventions of request. *Second Language Studies 21*: 1–47.

Göy, Elif, Deniz Zeyrek, and Bahar Otcu. 2012. Developmental patterns in internal modification of requests: a quantitative study on Turkish learners of English. In: Maria Economidou-Kogetsidis, Helen Woodfield, H. (eds.), *Interlanguage Request Modification*. Amsterdam/Philadelphia: John Benjamins, 51–86.

Grabowski, Kirby. 2013. Investigating the construct validity of a role-play test designed to measure grammatical and pragmatic knowledge at multiple proficiency levels. In: Steven Ross, and Gabriele Kasper (eds.), *Assessing Second Language Pragmatics*. New York, NY.: Palgrave MacMillan, 149–171.

Hashimoto, Brett and Kyra Nelson. 2020. Using a corpus in creating and evaluating a DCT. *Applied Pragmatics 1*(1): 80–120.

House, Juliane, and Daniel Z. Kadar. 2021. *Cross-Cultural Pragmatics*. Cambridge: Cambridge University Press.

Hudson, Thom. 2001b. Indicators for pragmatic instruction. In Gabriele Kasper, and Kenneth R. Rose (eds.) *Pragmatics in Language Teaching*. New York, NY.: Cambridge University Press, 283–300.

Hudson, Thom, James Dean Brown, and Emily Detmer. 1995. *Developing Prototypic Measures of Cross-cultural Pragmatics* (Technical Report #7). Honolulu, University of Hawaii, Second Language Teaching & Curriculum Center.

Hudson, Thom, Emily Detmer, and James Dean Brown. 1992. *A Framework for Testing Cross-cultural Pragmatics* (Technical Report #2). Honolulu, University of Hawaii, Second Language Teaching & Curriculum Center.

Ishihara, Noriko. 2010. Assessing learners' pragmatic ability in the classroom. In: Donna H. Tatsuki, and Noel Houck (eds.), *Pragmatics: Teaching speech acts*. Alexandria, VA: Teachers of English to Speakers of Other Languages, 209–227.

Ishihara, Noriko, and Elaine Tarone. 2009. Subjectivity and pragmatic choice in L2 Japanese: Emulating and resisting pragmatic norms. In: Naoko Taguchi (ed.), *Pragmatic Competence in Japanese as A Second Language*. Berlin: Mouton de Gruyter, 101–128.

Jenkins, Jennifer. 2015. *Global Englishes* (3rd ed.). London, UK: Routledge.

Kadar, Daniel Z. and Juliane House. 2020. Ritual frames: A contrastive pragmatics approach. *Pragmatics 20*(1): 142–168.

Kang, Okim, Don Rubin, and Alyssa Kermad. 2019. The effect of training and rater differences on oral proficiency assessment. *Language Testing 36*(4): 481–504.

Kasper, Gabriele. 2006. Speech acts in interaction: Towards discursive pragmatics. In: Kathleen Bardovi-Harlig, J. César Félix-Brasdefer, and Alwiya S. Omar (eds.), *Pragmatics and Language Learning: Vol. 11*. University of Hawai'i at Manoa: National Foreign Language Resource Center, 281–314.

Kasper, Gabriele, and Kenneth R. Rose. 2002. *Pragmatic Development in A Second Language*. Oxford, UK: Blackwell Publishing.

Kim, Eun Young, and Lucien Brown. 2014. Negotiating pragmatic competence in computer mediated communication: The case of Korean address terms. *CALICO Journal 31*(3): 264–284.

Kim, Youn Hee. 2009. An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach. *Language Testing 26*: 187–217.

Li, Shuai. 2014. The effects of different levels of linguistic proficiency on the development of L2 Chinese request production during study abroad. *System 45*:103–116.

Li, Shuai. 2012. The effects of input-based practice on pragmatic development of requests in L2 Chinese. *Language Learning 62*(2): 403–438.

Li, Shuai., and Naoko Taguchi. 2014. The effects of practice modality on the development of pragmatic performance in L2 Chinese. *The Modern Language Journal 98*(3): 794–812.

Li, Shuai, Naoko Taguchi, and Feng Xiao. 2019. Variations in rating scale functioning in assessing pragmatic performance in L2 Chinese. *Language Assessment Quarterly 16*(3): 271–293.

Liu, Jianda. 2006. *Measuring Interlanguage Pragmatic Knowledge of Chinese EFL Learners*. Frankfurt am Main: Peter Lang.

Liu, Jianda, and Wei Ren. 2015. Apologies in emails: Interactions between Chinese EFL learners and their foreign peers. In: Yuan-shan Chen, Der-Hwa Rau, and Gerald Rau (eds.), *Email Discourse among Chinese Using English as a Lingua Franca*. Springer, 205–228.

LoCastro, Virginia. 2003. *An Introduction to Pragmatics: Social Action for Language Teachers.* Ann Arbor, MI: The University of Michigan Press.

Maeshiba, Naoko, Naoko Yoshinaga, Gabriele Kasper, and Steven Ross. 1996. Transfer and proficiency in interlanguage apologizing. In: Susan M. Gass, and Joyce Neu (eds.), *Speech Acts Across Cultures: Challenges to Communication in A Second Language*. Berlin: Mouton de Gruyter, 155–187.

Mori, Junko. 2009. The social turn in second language acquisition and Japanese pragmatics research: Reflection on ideologies, methodologies and instructional implications. In: Naoko Taguchi (ed.), *Pragmatic Competence in Japanese as A Second Language*. Berlin/New York: Mouton de Gruyter, 335–338.

Nguyen, Thi Thuy Minh. 2019. Data collection methods in L2 pragmatics research: An overview. In: Naoko Taguchi (ed.), *The Routledge Handbook of SLA and Pragmatics*. New York, NY.: Routledge, 195–211.

Ogiermann, Eva, and Bella Spyridoula. 2020. An interlanguage study of request perspective: Evidence from German, Greek, Polish and Russian learners of English. *Contrastive Pragmatics 1*(1): 180–209.

Purpura, James. 2004. *Assessing Grammar*. Cambridge UK: Cambridge University Press.

Roever, Carsten. 2005. *Testing EFL Pragmatics*. Frankfurt: Peter Lang.

Roever, Carsten, Catriona Fraser, Catherine Elder. 2014. *Testing ESL Sociopragmatics: Development and Validation of A Web-based Test Battery*. Frankfurt: Peter Lang.

Rose, Kenneth. 2009. Interlanguage pragmatic development in Hong Kong, phase 2. *Journal of Pragmatics 41*: 2345–2364.

Sabaté i Dalmau, Maria, and Hortènsia Curell i Gotor. 2007. From "Sorry very much" to "I'm ever so sorry": Acquisitional patterns in L2 apologies by Catalan learners of English. *Intercultural Pragmatics 4*(2): 287–315.

Sasaki, Miyuki. 1998. Investigating EFL students' production of speech acts: A comparison of production questionnaires and role play. *Journal of Pragmatics 30*: 457–484.

Savic, Milica. 2015. "Can I very please borrow it": Request development in young Norwegian EFL learners. *Intercultural Pragmatics 12*: 443–480.

Seidlhofer, Barbara. 2011. *Understanding English as A Lingua Franca*. Oxford, UK: Oxford University Press.

Sydorenko, Tetyana, Carson Maynard, and Erin Guntly. 2014. Rater behavior when judging language learners' pragmatic appropriateness in extended discourse. *TESL Canada Journal 32*(1): 19–41.

Tajeddin, Zia, and Minoo Alemi. 2014. Pragmatic rater training: Does it affect non-native L2 teachers' rating accuracy and bias? *Iranian Journal of Language Testing 4*(1): 66–83.

Taguchi, Naoko. 2019. *The Routledge Handbook of SLA and Pragmatics*. New York, NY.: Routledge.

Taguchi, Naoko. 2018a. Data collection in developmental L2 pragmatics research: Discourse completion test, role play, and naturalistic recording. In: Aarnes Gudmestad, and Amanda Edmonds (eds.), *Critical Reflections on Data in Second Language Acquisition*. Philadelphia: John Benjamins, 7–32.

Taguchi, Naoko. 2018b. Advanced pragmatic competence. In: Paul A. Malovrh, and Alessandro G. Benati (eds.), *The Handbook of Advanced Proficiency in Second Language Acquisition*. Wiley-Blackwell, 505–526.

Taguchi, Naoko. 2012. *Context, Individual Differences, and Pragmatic Competence*. New York/Bristol: Multilingual Matters.

Taguchi, Naoko. 2011a. Do proficiency and study-abroad experience affect speech act production? Analysis of appropriateness, accuracy, and fluency. *IRAL – International Review of Applied Linguistics in Language Teaching 49*: 265–293.

Taguchi, Naoko. 2011b. Rater variation in the assessment of speech acts. *Pragmatics 21*(3): 453–471.

Taguchi, Naoko. 2011c. Pragmatic development as a complex, dynamic process: General patterns and case histories. *Modern Language Journal 95*: 605–623.

Taguchi, Naoko. 2007. Task difficulty in oral speech act production. *Applied Linguistics* *28*: 113–135.

Taguchi, Naoko, and Carsten Roever. 2017. *Second Language Pragmatics*. New York, NY.: Oxford University Press.

Taguchi, Naoko, Feng Xiao, and Shuai Li. 2016. Development of pragmatic knowledge in L2 Chinese: Effects of intercultural competence and social contact on speech act production in a study abroad content. *The Modern Language Journal 100*: 775–796.

Thomas, Jenny. 1983. Cross-cultural pragmatic failure. *Applied Linguistics 4*: 91–111.

Timpe, Veronica. 2013. *Assessing Intercultural Communicative Competence*. Frankfurt am Main: Peter Lang.

Timpe-Laughlin, Veronica. 2018. Pragmatics in task-based language assessment. In: Naoko Taguchi, and Youjin Kim (eds.), *Task-based Approaches to Teaching and Assessing Pragmatics*. Amsterdam: John Benjamins, 288–304.

Walters, F. Scott. 2007. A conversation-analytic hermeneutic rating protocol to assess L2 oral pragmatic competence. *TESOL Quarterly 24*(2): 155–183.

Xi, Xiaoming, and Pam Mollaun. 2011. Using raters from India to score a large-scale speaking test. *Language Learning 61*(4): 1222–1255.

Xiao, Feng, Naoko Taguchi, and Shuai Li. 2019. Effects of proficiency sub-skills on pragmatic development in L2 Chinese study abroad. *Studies in Second Language Acquisition 41*(2): 469–483.

Yamashita, Sayoko Okada. 1996. *Six measures of JSL pragmatics* (Tech. Rep. No. 14). Honolulu, HI: University of Hawaii, Second Language Teaching and Curriculum Center.

Youn, Soo Jung. 2018a. Task design and validity evidence for assessment of L2 pragmatics in interaction. In: Naoko Taguchi, and Youjin Kim (eds.), *Task-based Approaches to Teaching and Assessing Pragmatics*. Amsterdam: John Benjamins, 217–246.

Youn, Soo Jung. 2018b. Task-based needs analysis of L2 pragmatics in an EAP context. *Journal of English for Academic Purposes 36*(1): 86–98.

Youn, Soo Jung. 2015. Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing 32*(2): 199–225.

Youn, Soo Jung. 2007. Rater bias in assessing the pragmatics of KFL learners using FACETS analysis. *Second Language Studies 26*(1): 85–163.

Young, Richard. 2011. Interactional competence in language learning, teaching, and testing. In: Eli Hinkel (ed.), *Handbook of Research in Language Learning and Teaching*. New York, NY: Routledge, 426–443.