Chapter 11

The effects of task type and L2 proficiency on discourse appropriacy in oral task performance

Monika Ekiert, LaGuardia CC, City University of New York

Sofia Lampropoulou, Department of English, University of Liverpool

Andrea Révész, Institute of Education, University College London

Eivind Torgersen, Norwegian University of Science and Technology

Abstract

Conceived within the TBLT framework, the present study examined pedagogic tasks as vehicles for demonstrating L2 learners' discourse appropriacy in oral production. Eighty ESL learners' discourse appropriacy was measured using three pragmatically-oriented task types (complaint, refusal, and advice) across four different proficiency levels. The findings showed that, for all task types, as general proficiency increased, ratings of discourse appropriacy also increased. We found that there was a pronounced difference in discourse appropriacy between the intermediate and advanced proficiency levels, and that for learners at higher levels of proficiency, discourse appropriacy did not vary from task to task. In contrast, task type made a difference for less proficient learners in that the refusal task was particularly challenging compared with other tasks.

Introduction

In a *task-based language* teaching (TBLT) framework, the goal is to allow classroom learners to develop the ability to function successfully in real-life communicative settings. To achieve this goal, second language (L2) learners need to acquire a range of linguistic resources, as well as the ability to evaluate layers of contextual information, select the most appropriate language tools, and use them efficiently. Hence, in order to capture a fuller array of learning outcomes associated with learning to accomplish real-world language tasks, L2 performance needs to be measured not only with traditional linguistic indices of complexity, accuracy, and fluency (CAF), but also with measures of communicative adequacy (Pallotti, 2009; Révész, Ekiert, & Torgersen, 2016). Although different definitions exist, communicative adequacy or discursive appropriacy generally refers to learners' ability to adequately recognize and respond to the expectations of what to say and how to say it, contingent on contextual specifics (Young, 2011). This definition closely reflects the core construct underlying pragmatic competence.

Learning the social rules of speaking, or the *pragmatics* of conversation (Beebe, 1995), has been acknowledged as a fundamental aspect of language learning. Yet, traditional approaches to L2 pedagogy focus on vocabulary- or grammar-oriented activities and often overlook the social aspects of language use. Over the past few decades, calls from both the field of SLA and L2 pragmatics have been made for including pragmatics within L2 teaching (Roever, 2009; Taguchi, 2011a). As noted by Ortega (2003), L2 learning, in addition to strictly linguistic development, entails the development of discursive and sociopragmatic repertoires that learners can use appropriately in relation to particular communicative demands. Similarly, Beebe (1995) contends that learning social rules of speaking is about what to do with words, depending on the sociocultural context. Hence, pragmatics, just like grammar and lexis, need to be incorporated

into L2 pedagogy to provide a complete picture of target language use. TBLT, which calls for meaning-focused and goal-oriented language use via real life-like tasks, can offer a useful pedagogical framework to enhance L2 pragmatic development and its assessment.

Pragmatic knowledge involves two complementary dimensions: pragmalinguistics and sociopragmatics (Leech, 1983; Thomas, 1983). Pragmalinguistics refers to the linguistic resources available for performing language functions, while sociopragmatics refers to a language user's assessment of the context in which those linguistic resources are implemented and one's ability to respond to that context and communicative event or task. These two dimensions in tandem contribute to learners' pragmatic knowledge that can be demonstrated in real-life communication. Because pedagogic tasks in the TBLT framework focus on goal-oriented communicative language use, many of the tasks used in L2 classrooms and assessment settings place the task context within a foreseen or emerging sociorhetorical situation (Swales, 1990). These tasks can therefore be explored as a vehicle to both elicit and assess L2 learners' pragmatic ability (Ross & Kasper, 2013). The present study aims to explore whether such communicative tasks can indeed provide a useful platform for assessing L2 pragmatic ability. We investigate L2 users' discourse appropriacy using three pragmatically-oriented task types across various proficiency levels.

Literature Review

Pragmatics in L2 Pedagogy

Pragmatic competence, or the ability to convey and interpret meaning appropriately in a social situation, has been studied extensively in the fields of SLA and L2 assessment. Previous research in L2 pragmatics has shown that general proficiency and pragmatic ability may follow separate trajectories toward their full development (Bardovi-Harlig, 2000, 2012). While a

threshold level of grammatical ability is needed for L2 learners to perform pragmatic functions, grammatical competence is not sufficient for successful pragmatic performance. To assist L2 learners' acquisition of pragmatics researchers have argued for the importance of explicit teaching in L2 pragmatics (Bardovi-Harlig, 2013; Beebe & Waring, 2005; Taguchi, 2015; Takahashi, 2010).

Difficulty in learning pragmatics comes from the culture-specific nature of pragmatics. Some pragmatic functions may be universal, but linguistic and non-linguistic means to engage in those functions, as well as norms and conventions behind the practice, exhibit considerable variation across cultures (Taguchi, 2012). These conventions are also partly activity-specific and partly context-specific, and have to be worked out by L2 learners during the process of meaning making (Kasper & Rose, 2002). In addition, linguistic behaviors and social conventions of speaking are not easily observable. Taguchi (2012) remarks that learners often experience difficulty in detecting how target language speakers project appropriate levels of politeness or how they communicate meaning indirectly. When learners transfer their L1 sociocultural and sociorhetorical norms to L2 practice, they may end up with what Thomas (1983) calls *pragmatic failure* or a failure to convey the intended meaning, which occurs when two languages operate under different conventions. Pragmatic failure can also occur from not understanding contextual features of communication, that is, the relative power of the speaker over the hearer, social distance between them, their rights and obligations, as well as the degree of imposition involved in a communicative act (Thomas, 1995).

Existing studies in L2 pragmatics have revealed slow acquisition of pragmalinguistic forms (e.g., Barron, 2006; Schauer, 2004; Iwasaki, 2010). L2 learners tend to acquire coping strategies in target-language pragmatic acts relatively easily, but the precise syntax and lexis

needed to encode pragmatic intentions in those pragmatic acts do not develop as quickly (Taguchi, 2012). As noted by Taguchi, slow progress in the acquisition of pragmalinguistic forms indicates the unbalanced development between grammar and pragmatics among adult L2 learners. One promising way to promote pragmalinguistic development is by creating an instructional context that provides opportunities for acquiring pragmalinguistic features of the target language. TBLT can offer such a context by offsetting the lack of authentic communicative contexts in traditional language classrooms. By emphasizing language use in an authentic social context, task-based instruction allows exposure, raises awareness, and helps L2 learners practice language use through relevant instances of social interaction (Olshtain & Celce-Murcia, 2001).

Pragmatics and TBLT

TBLT is a strong form of communicative language teaching, where learners are encouraged to discern the language system through communication, and more specifically, through tasks. It is an analytic approach to language teaching, requiring the learner to discover the structures and meanings of language while engaged in communicative activities. Ellis (2003) defines a task "[as]…a workplan that requires learners to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate propositional content has been conveyed" (p. 16). Thus, TBLT purports that proficient use of a language involves a mastery of functional usage of the language within a social context, which is also a concern in pragmatics.

With its primacy of meaning, orientation towards real-world language use, and focus on both linguistic and non-linguistic outcomes (Ellis, 2003; Skehan, 1998), the task in TBLT is, by definition, a communicative and social act. When performing a task, participants need to

accomplish goals as social actors who do not just get things done but also attend to their interpersonal relationships with other participants. Ellis (2009) also emphasizes the authenticity of tasks, both *situational* (when a task corresponds to a real-world activity) and *interactional* (when a task instigates the same kind of interactional processes that arise in naturally occurring language use). These situational and interactional dimensions of a task also correspond to pragmatics that involves language use in social interaction.

So far TBLT researchers who have investigated task performance and learning outcomes in relation to task design have overwhelmingly focused on investigating the capacity of tasks for promoting the acquisition of grammar and vocabulary (e.g., Baralt, 2013; Kim, Payant, & Pearson, 2015; Révész, 2009; Révész, Sachs, & Hama, 2014; see, however, Kim & Taguchi, 2015), and tended to neglect the potential of tasks for promoting pragmatics-related language performance and learning. To fill this gap, the present study aims to utilize the TBLT framework to examine pragmatic dimensions of task performance in an L2 assessment context. In our investigation, we focus on L2 learners' abilities to deliver discourse-appropriate task performance in an L2. This entails conveying meaning at discourse-level as opposed to grammar and sentence level while attending to contextual factors such as the relationship with the interlocutor and goals of communication.

In utilizing the term *discourse appropriacy*, we draw on the approaches to task which place the task context within a sociorhetorical situation (Swales, 1990). In other words, following Swales, we posit that, typically, the sociorhetorical situation, foreseen or emerging in the task, will be represented by a language-specific discourse community. Discoursal conventions, including their pragmatic dimensions, are used by a particular discourse community with a view to accomplish communicative goals, just as tasks have communicative outcomes. To be

effective, L2 performance thus needs to reflect the discoursal conventions of a particular discourse community.

In the present study, we intend to explore how L2 proficiency and task type affect L2 learners' discourse appropriacy in spoken production. The relationship between proficiency and pragmatics in L2 use has received some attention in recent years (cf. Taguchi, 2011b, 2011c), but so far almost no studies have investigated how L2 proficiency interacts with task type in predicting discourse appropriacy in L2 learners' oral performance. The effect of task type on discourse appropriacy is also underexplored in the existing research. While a number of studies have examined how task type affects interactional patterns (e.g., Gurzynski-Weiss & Révész, 2012) or complexity, accuracy, and fluency of L2 performance (e.g., Skehan & Foster, 1997), the link between task type and discourse appropriacy has not been considered. The present study is an effort to advance research in this direction.

Research Questions

Conceived within the TBLT framework, the present study examined to the extent to which pragmatically-oriented pedagogic tasks can reveal learners' discourse appropriacy in oral production. Specifically, the study explored whether and how discourse appropriacy of L2 users varies across different L2 proficiency levels and across different pragmatic task types (complaint, refusal, and advice). Discourse appropriacy reflects a set of decisions made by learners during task performance in order to meet the expected pragmatic conventions of a given task in spoken discourse situations.

We addressed the following research questions:

1. To what extent does learner L2 proficiency predict discourse appropriacy during task performance?

2. To what extent does task type predict discourse appropriacy during task performance?

3. To what extent do learner L2 proficiency and task type interact in predicting discourse appropriacy during task performance?

Methodology

Data

The data for the present study included 300 task-based performances. Eighty ESL learners and twenty native speakers (NSs) of English completed three oral tasks. The ESL data were collected as part of a placement test, which was used to place students into appropriate levels in a community language program at a North American university (Kim, 2006). The test was theme-based and consisted of five sections (listening, speaking, grammar, reading, and writing). When selecting ESL participants, we considered both their overall scores and speaking section scores. Correlation between the overall scores and speaking scores was high ($r = .93$). For this study, 20 ESL participants were selected from each of the four proficiency levels (low-intermediate, intermediate, low-advanced, and advanced; Purpura, 2004) from among 600 test-takers. In order to control for their native language backgrounds, 10 Japanese and 10 Spanish test-takers were randomly selected from each proficiency level, because most test-takers had either Spanish or Japanese as their first language. The median age of the ESL participants was 29.5, and the mean age was 31.80 ($SD = 7.02$). Seventy-five percent were female. The participants had resided in an English-speaking country between 11 months and 5 years ($M = 2.25, SD = 1.48$). One-way ANOVAs with age and length of residence as dependent variables yielded no significant differences among the ESL participants $F (3, 76) = .333, p = .80$; and $F (3, 76) = .222, p = .88,$ across the four proficiency levels. The median age was also similar among the proficiency levels (median age range: 29.5-32). The NSs recruited for this study were all

studying at the same university. Their average age was 34.55 ($SD = 8.23$), and seventy percent were female.

Speaking tasks

The three speaking tasks used in this study asked participants to perform three different speech acts (Searle, 1979): making a complaint about a catering company, refusing a teacher's suggestion, and giving advice based on a radio commentary. They were all integrated tasks, drawing on various input types (see a sample task script in Appendix A). Although originally developed for assessing speaking performance, all three tasks fulfilled frequently-cited task criteria (e.g., Ellis, 2003): they were likely to generate a primary focus on meaning; they mirrored real-life activities; the speakers had to resort to their own linguistic resources while completing a task; and the tasks had both linguistic and non-linguistic task outcomes.

Participants first read, listened to, or viewed the task stimulus and then were prompted to respond to the stimulus. The tasks were delivered via computer without a live interlocutor. The planning time varied between 20 to 30 seconds, whereas the available speaking times were set at either 45 or 60 seconds. As a practice task, participants were asked to introduce themselves. The test tasks were administered to the participants in a fixed order. The three tasks are summarized in Table 1.

Table 1.

Task Summary (Révész et al., 2016)

Task	Task Description	Contextual Features
Complaint	The participant placed an order for his/her boss's birthday party but received a late and	Customer—Catering Service Input: Aural 20 seconds planning time

	incomplete delivery. The participant calls the caterer to complain about it.	45 seconds response time
Refusal	The participant who is unhappy with a professor's suggestion to take a lower-level class visits the professor and politely refuses the suggestion by providing reasons.	<u>Student</u>—Professor Input: Aural 30 seconds planning time 45 seconds response time
Advice	Upon listening to the radio commentary, the participant offers his/her opinion and advice on electric cars to a friend who is considering buying one.	<u>Friend</u>—Friend Input: Aural 30 seconds planning time 60 seconds response time

Note. The underlined part is the role that the participants were asked to play.

Tasks and analysis of discourse appropriacy

We assessed discourse appropriacy of each task performance. Discourse appropriacy was defined as the ability to use language to perform a speech act appropriately according to the sociocultural and sociorhetorical conventions of the task context (see Révész et al., 2016 for the measures of accuracy, linguistic complexity, and communicative adequacy[1] based on the same dataset). Because the three tasks involved different norms and conventions, first, the data from each task were analyzed qualitatively to identify discourse features elicited from each task. We focused on the use of linguistic and discourse-level features that reflected participants' understandings of the contextual parameters of each task. Based on the initial analyses of task relevant linguistic and discourse features, we rated each sample by using the task-independent discourse appropriacy scale consisting of five levels (see Appendix B). The rating scale was accompanied by the list of task-relevant linguistic and discourse features drawn from our initial

analyses of task performance. The following section presents the task-relevant features addressed in this study.

Task 1 required participants to make a direct complaint. This is a speech act that involves the expression of displeasure on the part of the speaker (Searle, 1979) as a result of an act that has affected him/her negatively. Complaints involve various communicative strategies that convey negative emotions. In this task, learners were expected to perform a direct complaint to a socially distant addressee over the telephone. Therefore, aspects of negative politeness (Brown & Levinson, 1987) as well as norms and conventions of the genre[2] of a telephone interaction were taken into account. In particular, we assessed aspects of self-presentation (i.e., whether and how test takers introduced themselves) in the opening of the telephone conversation and the ways in which the actual complaint was justified. Justification involved the reasons expressed to defend the speaker's position (DeCapua, 1998), as well as the provision of background information before proceeding with the complaint. We also analyzed participants' use of negative politeness strategies, including the use of address terms, conditionals, and sentence structures that help establish social distance between the speaker and listener. We also considered the complexity of the ways in which participants expressed criticism and negative emotions through the use of appropriate adverbs and adverbials, epistemic verbs, and repetition. Finally, the presence or absence of requests for repair was taken into account.

Task 2 involved the speech act of refusal to a senior person (from a student to a professor). Because refusal is a face-threatening act, it is often realized through indirect strategies and linguistic mitigations that can reduce the face threat. In this particular task, the potential face threat is large because of the social distance and relative status difference between the speaker and listener (Brown & Levinson, 1987). In light of this, we considered the use of mitigation

devices such as *could, would* and *if*-conditionals, as well as terms of address that help maintain

the social distance and power between speaker and listener. Additionally, the presence or

absence of semantic strategies (i.e., use of regrets and apologies, offers of reasons/explanations,

offers of alternative proposals, postponements and wishes) was considered (Beebe, Takahashi, &

Uliss-Weltz, 1990). Finally, the presence or absence of an adjunct to refusal, such as expression

of willingness, gratitude or initial agreement, was taken into account.

Task 3 asked participants to give advice about electric cars based on a radio commentary

that they heard. This speech act is part of directives (Searle, 1979). It is a non-impositive speech

act (Haverkate, 1984) because the objective is to benefit the hearer (Trosborg, 1995). Although

advice is given in the interest of the hearer, it is considered as a face-threatening act because it

intrudes on the hearer's world (Brown & Levinson, 1987). In fact, many L2 learners in this study

talked about the downsides of an electric car, which challenged the positive face of the addressee

who was considering buying one. When evaluating learners' task performance, we considered

the complexity of the justification provided for the suggestion (e.g., use of data and facts, such as

I heard there are problems with electric cars. The first problem is…). We also considered the

use of explicit suggestion expressions (*I advise/I suggest that you…*), conventionalized forms

(*have you thought about, you should, you need to* etc.) and indirect forms (*I've heard it's not the*

best idea) that accompanied the justification.

As we explained previously, these linguistic and discourse features were addressed when

assessing speech samples using the discourse appropriacy rating scale. The second author rated

the entire corpus of speech samples based on the five-point rating scale. The fourth author

analyzed a portion of the data (20%), which was selected through stratified random sampling by

taking proficiency level and task type into account. Interrater-reliability was high ($r=.92$ based on 60 samples, $p<.001$; Rater 1: $M=3.17$, $SD=1.30$; Rater 2: $M=3.04$; $SD=1.29$).

Data analysis procedures

First, descriptive statistics of discourse appropriacy ratings were analyzed across proficiency levels and task types. Then, to examine the effects of proficiency (RQ1), task type (RQ2), and their interaction (RQ3) on discourse ratings, a series of ANOVAs was conducted. As post-hoc analyses, a series of dependent samples t-tests was carried out. We adopted a conservative alpha level of .01 to control for Type 1 error due to the use of multiple comparisons. Eta-squared and partial eta-squared values were calculated to provide estimates of effect sizes for the ANOVAs (Norouzian & Plonsky, 2018), and Cohen's d values were computed to assess the effect size of the t-tests. Eta-squared values of .06, .16, .36 and Cohen's d values of .60, 1.00, 1.40 were considered small, medium and large respectively (Plonsky & Oswald, 2014).

<div align="center">

Results

</div>

Table 2 displays the descriptive statistics for the discourse appropriacy ratings by proficiency level and task type. The mean ratings increased as proficiency level increased for all task types. Overall, the difference in the mean ratings between the intermediate and advanced groups was more pronounced than the difference between the two intermediate and the two advanced groups, respectively.

A mixed-model ANOVA was conducted with proficiency level as a between-subjects variable and task type as a within-subjects variable. Results yielded a significant effect for task type, $F(2, 190) = 6.65$, $p < .01$, $\eta_p^2 = .07$, $\eta=.01$, for proficiency level, $F(4, 95) = 105.70$, $p < .01$, $\eta_p^2 = .82$, $\eta=.74$, and for the interaction between task type and proficiency, $F(8, 190) =$

2.70, $p < .01$, $\eta_p{}^2 = .10$, $\eta=.01$. Proficiency level explained 74% of the variation in the discourse appropriacy ratings, whereas task type and the interaction between proficiency and task type both accounted for only 1% of the variance. This means that level of proficiency had a large, positive impact on discourse ratings, while task type led to small differences in discourse ratings across the five levels of proficiency.

Table 2.

Descriptive Statistics for Discourse Ratings

Proficiency	N	Task type	Mean	SD	95% *CI Dif*	
					Lower	Upper
		Complaint	2.13	.79	1.75	2.48
Low-Intermediate	20	Refusal	1.70	.66	1.45	2.00
		Advice	1.78	.73	1.45	2.10
		Complaint	2.30	.57	2.05	2.55
Intermediate	20	Refusal	1.75	.64	1.50	2.00
		Advice	1.85	.75	1.55	2.15
		Complaint	3.15	.75	2.85	3.45
Low Advanced	20	Refusal	3.00	.73	2.70	3.30
		Advice	3.33	.65	3.05	3.60
		Complaint	3.35	.75	3.05	3.65
Advanced	20	Refusal	3.30	.66	3.05	3.60
		Advice	3.38	.48	3.18	3.60
		Complaint	4.80	.41	4.60	4.95
Native	20	Refusal	4.85	.37	4.70	5.00

| | Advice | 4.85 | .37 | 4.70 | 5.00 |

Note. The maximum score was 5 points as assessed on the 5-point discourse appropriacy rating scale.

To explore the interaction effect between proficiency and task type on discourse appropriacy ratings, we ran a series of repeated-measures ANOVAs separately for each proficiency level with task type as a within-subject factor. As shown in Table 3, task type emerged as a significant, medium-size predictor of discourse appropriacy at the low-intermediate, intermediate, and low-advanced levels. The effect of task type was not significant for advanced-level L2 participants and NS participants. That is, while the low-intermediate, intermediate, and low-advanced participants received significantly different discourse ratings across task types, the discourse ratings of the advanced-level and native speaker participants did not significantly vary as a function of task type.

Table 3.

Summary of Repeated Measures ANOVAs with Task Type Predicting Discourse Appropriacy

Proficiency	*df*	*F*	*p*	η
Low Intermediate	2, 38	4.82	.01	.202
Intermediate	2, 38	6.59	.00	.258
Low Advanced	2, 38	4.21	.02	.181
Advanced	2, 38	.12	.89	.006
Native	2, 38	.16	.85	.008

In order to investigate the interaction effect further, we performed post-hoc dependent samples *t*-tests for the low-intermediate, intermediate, and low-advanced learners. The results of the *t*-tests are presented in Table 4, and the statistically significant patterns are summarized in

Table 5. As shown in Table 5, making a refusal seemed to pose greater difficulty for low-intermediate and intermediate students than making a complaint. Similarly, the low-advanced students performed less successfully on the refusal task in comparison to the advice-giving task. The effect sizes for all these relationships were in the medium range.

Table 4.

Summary of Post-hoc Dependent Samples T-tests for Task Type Comparisons

| Proficiency | Tasks Comparison | *M* | *SD* | 95% *CI Dif* | | *t* | *df* | *p* | *d* |
				Lower	Upper				
LowInt	Comp – Ref	.43	.63	.13	.72	3.00	19	.01	.67
	Comp – Adv	.35	.67	.04	.66	2.33	19	.03	.52
	Ref – Adv	-.08	.65	-.38	.23	-.51	19	.61	-.11
Int	Comp – Ref	.55	.69	.23	.87	3.58	19	.00	.80
	Comp – Adv	.45	.76	.09	.81	2.65	19	.02	.59
	Ref – Adv	-.10	.72	-.44	.24	-.62	19	.54	-.14
LowAdv	Comp – Ref	.15	.49	-.08	.38	1.37	19	.19	.31
	Comp – Adv	-.18	.54	-.43	.08	-1.44	19	.17	-.32
	Ref – Adv	-.33	.47	-.54	-.11	-3.11	19	.01	-.70

Note. Adv=advice; Comp=complaint; Ref=refusal

Table 5.

Significant Differences among Mean Discourse Appropriacy Ratings for Task Types across

Proficiency Levels

Proficiency	Significant Patterns
Low Intermediate	Complaint > Refusal

Intermediate	Complaint > Refusal
Low Advanced	Advice > Refusal

Note. X > Y indicates that participants achieved significantly higher mean discourse appropriacy ratings on Task X than Task Y.

Discussion

The present study investigated ESL learners' discourse appropriacy measured on three pragmatically-oriented task types across different general proficiency levels. Previous TBLT studies that examined task type and task design have offered suggestions for improving learners' linguistic performance in terms of complexity, accuracy, and fluency of L2 production, with the focus of facilitating the acquisition of grammatical and lexical features. The present study intended to complement previous research by applying a performance-based analysis that addressed pragmatic competence at the discourse level.

Overall, our findings support the proposal that pragmatically-oriented tasks can offer unique opportunities for displaying learners' appropriacy in oral discourse (cf. Kim & Taguchi, 2015). Specifically, we found that, regardless of task type, or speech act type involved in the task, as proficiency level increased, ratings of discourse appropriacy also increased. Our findings revealed a pronounced difference in discourse appropriacy ratings between the intermediate and advanced proficiency levels. In other words, L2 proficiency had a strong impact on the perceived appropriacy of L2 learners' spoken discourse in pragmatically-oriented tasks.

We also addressed a question of whether task type, or speech act type involved in the task, can predict discourse appropriacy of L2 task performance. Our findings indicate that, at the advanced proficiency levels, task type did not have any impact on L2 participants' discourse appropriacy ratings. That is, at higher levels of proficiency, discourse appropriacy did not vary from task to task. In contrast, task type made a difference at the less advanced L2 proficiency

levels. For learners with low and upper-intermediate and low-advanced proficiency, discourse appropriacy varied depending on task type. L2 speakers whose proficiency had not reached near-native-like thresholds displayed variable discourse appropriacy, struggling with some but not other tasks.

The refusal task turned out to be particularly challenging for the less proficient L2 speakers. Refusals have been described in the L2 pragmatic studies as a major cross-cultural obstacle (cf. Babai Shishavan & Sharifian, 2016). Difficulty involved in refusals has been attributed to this speech act's sociolinguistic complexity and variation in form and content depending on the refusal type (refusal to invitation, request, offer, or suggestion). In addition, refusals are sensitive to contextual variables such as the interlocutors' power difference and social distance, which can affect the propositional content of the speech act itself. Our task involved a refusal to person with greater power, which complicated the already face-threatening speech act. Learners were therefore challenged not only with the linguistic demands of the refusal but also with the need to soften the tone of the refusal to maintain the social distance and power relationship with their interlocutor. Hence, learners had to compensate for the necessity to express some form of disapproval or disrespect (Babai Shishavan & Sharifian, 2016) towards an addressee of higher social power through a series of mitigating politeness strategies. In fact, the relationship between student and professor, expected to be interactionally reflected here, may be conceived significantly differently in learners' native cultures. The relationship, thus, carries different pragmatic connotations associated with diverse linguistic resources. In contrast, other speech acts used in the tasks involved a situation in which participants had either an equal or a higher power to their interlocutor (e.g., complaining to a hired caterer or advising a friend, cf. Taguchi, 2007). As a result, participants did not have to use elaborate linguistic expressions or

discourse strategies to mitigate their force, resulting in their relatively higher scores on these speech acts. It may therefore be concluded that the refusal task served as a particularly stringent test of L2 pragmatics and was the most effective in eliciting the core dimensions of pragmatic competence.

Conclusion

The findings of this study indicate that TBLT as a framework is in a good position to provide L2 practice and to create a platform for assessing knowledge of form-function-context mappings in the target language. While tasks have been used differently in the fields of SLA, L2 teaching, and assessment, TBLT enables curriculum and test developers to prioritize contexts in which learners can use the target language to achieve communicatively appropriate functions. The tasks that we utilized in this study are such examples. The study, naturally, has a number of limitations that need to be addressed in further research. One limitation of the study has to do with its exclusive quantitative orientation. It was beyond the scope of this study to conduct more detailed, bottom-up qualitative analyses of speech samples. Second, because we used only three pragmatically-oriented task types, it would be interesting to explore how discourse appropriacy varies across a larger number of tasks. These limitations notwithstanding, our study provided valuable new insights to the fields of TBLT and pragmatics research and confirmed that exploring further synergies between the two fields is a worthwhile research endeavour.

We conclude this paper with several implications for teaching and future directions. From a task-based perspective, pragmalinguistics and sociopragmatics could be addressed via *input-providing* tasks (Ellis, 2009) or consciousness-raising tasks (Ellis, 2003). In this study, multiple pragmatic targets were embedded in the three integrative skills tasks that could be characterized as mainly *output-providing* (Ellis, 2009). These tasks provided real-world speaking contexts and

offered opportunities to practice target language form-function-context mappings. However, in future studies, it would be interesting to examine the knowledge of discourse appropriacy not only via production-oriented tasks but also via comprehension- and recognition-focused tasks. Following Takimoto's (2012) research on the awareness of pragmatic appropriateness, TBLT researchers can consider using consciousness-raising tasks to promote learners' recognition of social appropriateness of target language use. For example, learners can be encouraged to discover politeness strategies used to mitigate face-threatening speech acts in naturalistic conversations. Learners may also be prompted to explain the pragmatic failure of a speaker in performing a communicative task.

Another promising future direction is to explore the potential of communicative tasks to assess and increase what Taguchi (2012) refers to as *pragmatic fluency*. Taguchi suggests the importance of a conjoined analysis of pragmatic skills and processing fluency in the development of pragmatic ability. The TBLT framework provides a unique platform for investigating and developing pragmatic fluency by requiring learners to produce pragmatic functions in contexts that often mimic real-world language use.

Finally, within the framework of TBLT, investigating the relationship between grammar and pragmatics in L2 development should be explored further by looking into the relationship between specific linguistic constructions and discourse appropriacy (cf. Bardovi-Harlig, 2013). In future research, it would be useful to explore this relationship by selecting linguistic features that are relevant to pragmatically appropriate task performance in a given context (e.g., use of embedding clauses in requests). More importantly, promoting learners' attention to the connection among grammatical forms, their social functions, and contexts of their occurrence could significantly enhance the learning potential of pedagogical tasks.

Acknowledgements

Endnotes

1. Communicative adequacy and discourse appropriacy are related but not overlapping constructs. We found a strong, but not perfect correlation between communicative adequacy and discourse appropriacy based on the dataset (Révész et al., 2016).

2. Following Johnson and Johnson (1999), we see genres as types of spoken and written discourse recognized by a discourse community, each having typical features, including linguistic (particular grammatical or lexical choices), paralinguistic (print size, gesture), and contextual and pragmatic (setting, purpose).

References

Babai Shishavan, H., & Sharifian, F. (2016). The refusal speech act in a cross-cultural perspective: A study of Iranian English-language learners and Anglo-Australian speakers. *Language and Communication, 47,* 75-88.

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition, 35*, 689-725.

Bardovi-Harlig, K. (2000). Pragmatics and second language acquisition. In R. Kaplan (Ed.), *Oxford handbook of applied linguistics* (pp. 182-192). Oxford: Oxford University Press.

Bardovi-Harlig, K. (2012). Pragmatics in SLA. In S. M. Gass & A. Mackey (Eds.) *The Routledge handbook of second language acquisition* (pp. 147-162). London: Routledge/Taylor Francis.

Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning*, *63*, 68–86.

Barron, A. (2006). Learning to say 'you' in German: The acquisition of sociolinguistic competence in a study abroad context. In M. DuFon & E. Churchill (Eds.), *Language learners in study abroad contexts* (pp. 59-82). Clevedon: Multilingual Matters.

Beebe, L. (1995). The social rules of speaking: Basics—not frosting on the cake. *The Language Teacher, 19*(3), 4-11.

Beebe, L., Takahashi, T, & Uliss-Weltz, R. (1990). Pragmatic transfer in ESL refusals. In R. Scarcella, E. Andersen, S. D. Krashen (Eds.), *On the development of communicative competence in a second language* (pp. 55-73). New York: Newbury House.

Beebe, L., & Waring, H. (2005). Pragmatic development in responding to rudeness. In J. Frodesen & C. Holten (Eds.), *The power of context in language teaching and learning* (pp. 67-80). Boston, MA: Heinle & Heinle/Thomson Learning.

Brown, P. & Levinson, S.C. (1987). *Politeness: Some universals in language use*. Cambridge: Cambridge University Press.

DeCapua, A. (1998). Pragmatic transfer and cultural stereotyping. *Issues in Applied Linguistics, 9,* 21–36.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*, 221-246.

Gurzynski-Weiss, L., & Révész, A. (2012). Tasks and teacher feedback in the Spanish foreign language classroom. *Language Learning, 62,* 851-879.

Haverkate, H. (1984). *Speech acts, speakers and hearers. Pragmatics and beyond 4*. Amsterdam: John Benjamins.

Iwasaki, N. (2010). Style shifts among Japanese learners before and after study abroad in Japan: Becoming active social agents in Japan. *Applied Linguistics, 31*, 45-71.

Johnson, K., & Johnson, H. (1999). *Encyclopedic Dictionary of Applied Linguistics*. Oxford: Blackwell.

Kasper, G., & Rose, K. (2002). *Pragmatic development in a second language*. Oxford: Blackwell.

Kim, H-J. (2006). Providing validity evidence for a speaking test using FACETS. *Teachers College Columbia University Working Papers in TESOL & Applied Linguistics, 6* (1). From http://journals.tc-library.org/index.php/tesol/article/view/180

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory. *Studies in Second Language Acquisition, 37*, 549-581.

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom

    contexts: The role of task complexity. *Modern Language Journal, 99,* 656-677.

Leech, G. (1983). *Principles of pragmatics.* London: Longman.

Norouzian, R., & Plonsky, L. (2018). Eta- and partial eta-squared in L2 research: A cautionary

    review and guide to more appropriate usage. *Second Language Research*, 34, 257-271.

Olshtain, E., & Celce-Murcia, M. (2001). Discourse analysis and language teaching. In. D.

    Schiffrin, D. Tannen, & H.E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 707-

    724). Oxford: Blackwell.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A

    research synthesis of college-level L2 writing. *Applied Linguistics,24,* 492-518.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics,*

    30, 590-601.

Plonsky, L., & Oswald, F. (2014). How big is "big"? Interpreting effect sizes in L2 research.

    Language Learning, 64, 878-912.

Purpura, J. 2004. *Assessing grammar*. Cambridge: Cambridge University Press.

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies*

    in Second Language Acquisition, 31, 437-470.

Révész, A., Ekiert, M., & Torgersen, E. (2016). The effects of complexity, accuracy, and fluency

    on communicative adequacy in oral task performance. *Applied Linguistics, 37*, 828-848.

Révész, A., Sachs, R., & Hama, M. (2014). The effects of task complexity and input frequency

    on the acquisition of the past counterfactual construction through recasts. *Language*

    Learning, 64, 615-650.

Roever, C. (2009). Teaching and testing pragmatics. In M. H. Long & C. J. Doughty (Eds.),

Handbook of language teaching (pp. 560–577). Malden, MA: Wiley-Blackwell.

Ross, S., & Kasper, G. (2013). *Assessing second language pragmatics*. London: Palgrave

Macmillan.

Schauer, G. (2004). 'May you speaker louder maybe?' Interlanguage pragmatic development in

requests. *EUROSLA Yearbook, 4*, 253-272.

Searle, J. (1979). *Expression and meaning: Studies in the theory of speech acts.* Cambridge:

Cambridge University Press.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P., & Foster, P. (1997). Task-type and task processing conditions as influences on

foreign language performance. *Language Teaching Research, 1*, 1-27.

Swales, J. (1990). *Genre analysis*. Cambridge: Cambridge University Press.

Taguchi, N. (2007). Task difficulty in oral speech act production. *Applied Linguistics*, *28*, 113-

135.

Taguchi, N. (2011a). Teaching pragmatics: Trends and issues. *Annual Review of Applied

Linguistics, 31,* 289-310.

Taguchi, N. (2011b). Do proficiency and study-abroad experience affect speech act production?:

Analysis of appropriateness, accuracy, and fluency. *International Review of Applied

Linguistics, 49,* 265-293.

Taguchi, N. (2011c). The effect of L2 proficiency and study-abroad experience in pragmatic

comprehension. *Language Learning*, *61,* 904-939.

Taguchi, N. (2012). *Context, individual differences, and pragmatic competence.* Clevedon:

Multilingual Matters.

Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and

should be going. *Language Teaching*, *48,* 1-50.

Takahashi, S. (2010). Assessing learnability in second language pragmatics. In A. Trosborg

(Ed.), *Handbook of pragmatics vol.7* (pp. 391-421). Berlin: Mouton de Gruyter.

Takimoto, M. (2012). Metapragmatic discussion in interlanguage pragmatics. *Journal of*

Pragmatics, *44*, 1240-1253.

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics, 4*, 91-112.

Thomas, J. (1995). *Meaning in interaction: An introduction to pragmatics*. London: Longman.

Trosborg, A. (1995). *Interlanguage pragmatics: Requests, complaints, and apologies*. Berlin:

Mouton de Gruyter.

Young, R. (2011). Interactional competence in language learning, teaching, and testing. In E.

Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 426-

443). London: Routledge.

Appendix A. Sample Task Script (Kim, 2006)

Task 1. Catering Service
In this task, you need to complain about something. Imagine you have ordered food from Party Planner's Inc. for your boss's birthday party. But there was not enough food and it was delivered late. You spent a week planning the party, but it was ruined because of the food. You were

extremely upset that it happened. Call the caterer to complain about it. You have 20 seconds to plan.

Prompt (Audio)

[phone ringing] (Answering Machine) Hi! You've reached Party Planner's Inc. We're sorry, but we're not available to take your call right now. Please leave a detailed message after the beep, and we'll get back to you as soon as possible. [Beep] Test-Taker: (45 sec response time)

Appendix B. Discourse Appropriacy Scale

5	The response is **completely discourse appropriate**. Task-relevant discourse features are used **very successfully**.
4	The response is **discourse appropriate**. Task-relevant discourse features are used **successfully**.
3	The response is **moderately discourse appropriate**. Task-relevant discourse features are used **moderately successfully**.
2	The response is **discourse inappropriate**. Task-relevant discourse features are used **unsuccessfully**.
1	The response is **completely discourse inappropriate.** Task-relevant discourse features are used **very unsuccessfully**.