Task-Based Approaches to Teaching and Assessing Pragmatics

Edited by
Naoko Taguchi
YouJin Kim

Task-Based Language Teaching

10

John Benjamins Publishing Company

Task-Based Approaches to Teaching and Assessing Pragmatics

Task-Based Language Teaching: Issues, Research and Practice (TBLT)

Task-Based Language Teaching (TBLT) is an educational framework for the theory and practice of teaching second or foreign languages. The TBLT book series is devoted to the dissemination of TBLT issues and practices, and to fostering improved understanding and communication across the various clines of TBLT work.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/tblt

Editors

Martin Bygate	John M. Norris	Kris Van den Branden
Lancaster University	Educational Testing Service	KU Leuven

Volume 10

Task-Based Approaches to Teaching and Assessing Pragmatics
Edited by Naoko Taguchi and YouJin Kim

Task-Based Approaches to Teaching and Assessing Pragmatics

Edited by

Naoko Taguchi
Carnegie Mellon University

YouJin Kim
Georgia State University

John Benjamins Publishing Company

Amsterdam / Philadelphia

Table of contents

Acknowledgement

Series editors' preface

It is with great pleasure that we introduce the tenth volume in this series, an edited collection on *Task-based approaches to teaching and assessing pragmatics*. To our knowledge, this is the first book-length treatment of the intersection between TBLT and pragmatics, and there are no scholars better suited to the task than the volume's editors – Naoko Taguchi and YouJin Kim – well-known researchers in both the TBLT and second language (L2) pragmatics domains. We are all the more pleased that their treatment of this interesting topic picks up on a theme that we highlighted in our introduction to the very first volume in the series (*Task-based language teaching: A reader*), where we observed that empirical research on tasks had, to date, overwhelmingly focused on lexicogrammatical dimensions of language performance and learning at the expense of other aspects. In this tenth volume, then, we are seeing evidence of the maturation of research and practice in TBLT, to the extent that earlier empirical gaps are gradually being addressed in work that is both theoretically motivated and simultaneously cognizant of the need to link up with educational practices.

As Taguchi and Kim note in their opening chapter, the notion of an intellectual 'shared space' between TBLT and pragmatics makes consummate sense. Both domains are fundamentally interested in how humans actually make things happen with language, in the realities of functional language use as it impacts communication success, and in the ways in which learners develop second language communicative competence that is sensitive to social/contextual factors. Clearly, all common language tasks call upon some degree of pragmatics for their successful conduct, and pragmatic 'rules' do not make much sense when stripped of the real-world context provided by tasks. Indeed, some of the seminal early voices in TBLT emphasized certain pragmatic features of language learning and explored task types that depended heavily on pragmatic competence (see, e.g., the chapter by John Swales in volume one of this series). Yet, perhaps surprisingly in light of shared concerns, pragmatics has not figured prominently in research-based publications on task-based performance or learning, nor have task-based instructional techniques been explored extensively within L2 pragmatics research. The sources for this apparent disconnect are likely both epistemological – task-based learning and L2 pragmatics have not adopted similar methodological orientations – and practical, in that a lot of what happens in the language classroom (where pragmatics no doubt receives attention) probably neither ends up in publications nor necessarily adopts a task-based perspective.

Recently, though, the perhaps inevitable intersection of pragmatics and TBLT has come to be the subject of both classroom application and empirical investigation. This volume brings together some of the key outcomes of that pioneering work, covering

three important perspectives. First, it addresses the use of tasks to intentionally teach L2 learners various aspects of pragmatics (e.g., honorifics, mitigators). Second, it examines the ways in which distinct elements of sociopragmatic and pragmalinguistic competence may be elicited via performances on communication tasks typical of the TBLT domain, and how those tasks might in turn serve as vehicles for developing L2 pragmatic competence. Third, importantly, it explores how pragmatic competence might be assessed through L2 performance tasks and what role various dimensions of pragmatics can/should play within the assessment of communicative language use. The chapters in these three sections of the volume reflect the cutting-edge, indeed innovative nature of this kind of work. In many cases, they do not reach unequivocal findings or provide empirically-grounded recommendations for practice, due to the fact that they are broaching empirical and practical topics for the first time in the field, or at least at very early stages of theorization and implementation. The volume's contribution is all the more critical, then, in that it does a very good job of both (a) sketching out the territory between TBLT and pragmatics in need of exploration, and (b) providing baseline observations and findings across a number of critical topics within this new landscape.

We congratulate the volume's editors on achieving an important new milestone in both the TBLT and L2 pragmatics domains, and we look forward to the ensuing iterations in language educational research and practice.

John Norris, Martin Bygate, Kris Van den Branden

Task-based approaches to teaching and assessing pragmatics

An overview

Naoko Taguchi & YouJin Kim
Carnegie Mellon University / Georgia State University

With the recognition that pragmatics, like grammar and lexis, should be incorporated into pedagogy, research in teaching pragmatics has grown rapidly since the 1980s (Taguchi, 2015; Taguchi & Roever, 2017). The field of task-based language teaching (TBLT) has also received significant attention within language education and applied linguistics (Bygate, 2015; Ellis, 2009; Long, 2015; Van den Branden, Bygate, & Norris, 2009). Despite common research and pedagogical interests, however, these two research domains (i.e., pragmatics and TBLT) have not been explored together. Our volume is the first book-length attempt to bring together these two fields by exploring implications of TBLT for the learning and teaching of pragmatics, as well as assessing second language (L2) pragmatic competence.

Over the last several decades, many pedagogical approaches and methods (e.g., Audiolingualism, Natural Approach, Total Physical Response, Communicative Language Teaching) have been proposed by L2 educators and researchers (for review, see Larsen-Freeman & Anderson, 2011; Richards & Rodgers, 2010). TBLT is an educational proposal and a pedagogical approach that uses tasks as a unit of instruction as well as central teaching and learning resources. Based on a holistic view of language, TBLT represents an analytic approach to syllabus design; it does not divide up the language by grammar structure or lexical topic but instead involves holistic use of language performed during communicative functions (Long, 2015). For instance, one of the earliest TBLT examples is the Communicational Teaching Project in Bangalore, India, which presented a full-scale language education program based on a multi-year task-based syllabus (Prabhu, 1987).

In TBLT, tasks are considered beneficial for language learning because they address learners' real-world needs and promote their engagement with meaningful language use (Ellis & Shintani, 2014; Long, 2015). Accordingly, researchers and educators claim that tasks should be designed based on learner needs, and thus when

designing task-based curricula, needs analyses have been considered an important starting point (see Long, 2005, 2015 for review). Tasks are also socially situated, as learners' real-world communication needs are the major considerations of task design and implementation. Although through task performance, learners are expected to use their pragmatic knowledge in a given social context, the development of pragmatic competence through task performance has not been explicitly addressed in the current TBLT literature.

The teaching of pragmatics is also concerned with socially situated language use. As Thomas (1983) originally claimed, pragmatic competence involves two knowledge dimensions: *pragmalinguistics* (knowledge of linguistic forms for performing a communicative function) and *sociopragmatics* (knowledge of contextual features, norms of interaction, and social conventions associated with a communicative situation). Because these two dimensions are congruent with the basic tenets of TBLT (e.g., situated interactions, real-world communicative needs, and communication goals), TBLT offers a framework which not only is useful for the teaching and assessing of pragmatics, but also really requires it. Our volume will contribute to this TBLT-pragmatics intersection and help advance L2 pragmatics research by articulating the relevance of TBLT as a guiding framework for task design and use in teaching and assessing pragmatics. The TBLT-pragmatics connection will be illustrated through a variety of sub-topics within various domains of applied linguistics research, including technology-enhanced learning, instructed SLA, pragmatic assessment, discursive pragmatics, and heritage language learning.

The purpose of this introductory chapter is to discuss characteristics of tasks that could be useful for those who are interested in teaching and/or learning pragmatic knowledge through tasks and to review previous instructional pragmatic studies which to some extent adopted tenets and concepts of TBLT (although virtually no study explicitly acknowledged it). In this introductory chapter, we first define what a task is and discuss characteristics of a task in TBLT. Then we provide an overview of current research in task-oriented instructed pragmatics. After these two sections, gaps in each area of literature are brought together to articulate the intersection between TBLT and pragmatics. We then provide a list of suggestions on how these two fields can complement each other to advance our knowledge of teaching and assessing pragmatics from a TBLT perspective. This chapter ends with an overview of the chapters included in the book.

Characteristics of tasks in TBLT

Over the last three decades, L2 researchers and practitioners have viewed tasks as a primary unit of language instruction and have investigated various ways to promote

task-based language learning (e.g., Bygate, Skehan, & Swain, 2001; Ellis 2003; Long 2015, 2016; Van den Branden, Bygate, & Norris, 2009). Although how one can define tasks and distinguish them from activities or exercises is still discussed in the literature, some fundamental characteristics of tasks have been recognized. For instance, Ellis's four criteria (as cited in Ellis & Shintani 2014, p. 135) for an instructional activity to qualify as a 'task' are:

1. The primary focus should be on meaning.
2. There should be some kind of gap (i.e., a need to convey information to express an opinion or to infer meaning).
3. Learners should largely rely on their own linguistic and non-linguistic resources to complete the activity, with some help from the task input.
4. There is a clearly defined outcome other than the use of language.

In addition to these four characteristics, the authenticity of tasks has been emphasized, as in Long's (2016) definition of tasks as "the real-world communicative uses to which learners will put the L2 beyond the classroom – the things they will do in and through the L2" (p. 6). In the current volume, each chapter provides a definition of 'task' that has been adopted along these or similar lines. Overall, the chapters generally define a task from a pedagogic perspective (i.e., with language learning goals), which is in line with Van den Branden's (2006) definition: "A task is an activity in which a person engages in order to attain an objective, and which necessitates the use of language" (p. 4). With regards to the characteristics of tasks, collaborative work has been implemented as a pedagogical procedure, and in order to facilitate interaction, tasks are often designed as collaborative tasks. However collaboration is not a required characteristic of a task. For instance, in Chapter 2 in this volume, Kim, Lee, and Kim designed a drama script collaborative writing task that offered students an opportunity to use various honorific expressions in Korean. Collaborative work was chosen as a pedagogic procedure in this study to elicit interaction between learners as a part of their research design. In other words, tasks can be monologic or communicative depending on how they are designed to achieve the goal.

Although the optimal goal of TBLT research is to inform syllabus design, task selection, and task sequencing within a task-based curriculum, tasks have been put to a variety of uses in the fields of applied linguistics and L2 education (Van den Branden et al., 2009). Within L2 pedagogy, task-based syllabus design has been one of the central issues. L2 practitioners and researchers have contrasted two different approaches to language teaching. Synthetic or Type A syllabuses are designed based on the elements of linguistic systems, such as sounds, morphemes, grammar, vocabulary, notions, and functions. Instructors preselect the sequence of information to be presented in a way that linguistic forms are gradually accumulated. Such approaches

have received much criticism over time. For instance, Long and Crookes (1993) state that a synthetic approach to syllabus design is based on artificial, prescriptive language samples and is not in line with recent SLA research, which highlights the importance of meaning conveyed through forms. On the other hand, analytic syllabuses, or Type B syllabuses, are organized in terms of learner performance with language as a whole. They do not split language into pieces but take functional and communicative tasks as the unit of instructional analysis. From this perspective, learners "not only learn language *in order to* make functional use of it, but also *by* making functional use of it" (Van den Branden, 2006, p. 6). According to Long (1985), the traditional distinction between *what* is taught (as in what is on the syllabus) and *how* it is taught (methodology) is not relevant, as the same unit of analysis (i.e., task) is used for both.

Building on earlier research focusing on developing educational programs (Candlin, 1987; Prahbu, 1987), research studies which focus more on the educational goals of tasks (i.e., "researched pedagogy," Van den Branden, 2015) have been of interest for teachers and learners in instructional contexts. For example, these studies explored how teachers use tasks in class (e.g., Samuda, 2015), as well as teacher education and teacher/learner perceptions towards TBLT (e.g., Carless, 2004, 2007; Van den Branden, 2015, 2016). Furthermore, from pedagogical perspectives, topics related to needs analyses (Chaudron et al., 2005), task-based curricula/syllabus development (Van den Branden, 2006), task design and implementation (Kim, 2015), teacher development (East, 2012, 2017; Van den Brandan, 2016), and program evaluation (Norris, 2015) have been explored (see also Bygate, 2015 and articles in a special issue on tasks in *Annual Review of Applied Linguistics*, 2016).

Within the field of applied linguistics, tasks also have been widely used in second language acquisition (SLA) research. Since the first publications on TBLT (Long, 1985; Prabhu, 1987), tasks have been used to elicit learner data, and many empirical studies have focused on learners' language output during task performance in terms of linguistic quality (e.g., complexity, accuracy, fluency) or interactional features (see Kim, 2015; Plonsky & Kim, 2016 for review). These studies often employ tasks to elicit learner data in order to answer theory-driven research questions, such as the role of task repetition in linguistic performance, rather than to examine the learning outcomes of task-based language teaching within a task-based course or curriculum (Van den Branden et al., 2009). However, in terms of the role of tasks in language development in intact classrooms, there has been a surge of research demonstrating the benefits of pedagogic tasks, and several theories have been put forth to account for factors that mediate linguistic development (e.g., Robinson's, 2015, Cognition Hypothesis, Skehan's, 2015, Limited Capacity Hypothesis).

Recently, in their synthesis papers, researchers (East, 2017; Ellis, 2017; Long, 2016) have identified real issues and non-issues in the field of TBLT. These researchers commonly acknowledge that whether tasks promote language learning is no longer arguable, and that collaborative tasks can facilitate learners' attention to form,

especially when implemented with other focus-on-form techniques. Furthermore, task-based research has offered useful pedagogical suggestions. For instance, the findings of previous research suggest the benefits of cognitively complex tasks in promoting interaction-driven learning opportunities (see Kim, 2015 for review) and the potential benefits of task repetition and planning in promoting better task performance in terms of complexity, accuracy, and fluency (see Bygate, in press, for a selection of studies). Some of the common research gaps in the current TBLT literature (i.e., real issues) are identified in these synthesis papers to include task sequencing, task-based learning transferability, and teacher development (Ellis, 2017; Long, 2016).

In the current volume, we acknowledge that researchers have addressed various dimensions of linguistic development, such as grammar and vocabulary, as well as complexity, accuracy, and fluency of language production during task performance. However, the current TBLT literature is still skewed towards certain language and linguistic skills. For instance, as East (2017) contends, speaking skill has been the predominant research area, and grammar and lexis have been the main focus of linguistic development research.

In a recent methodological review of task-based research focusing on studies from interactionist perspectives, Plonsky and Kim (2016) found that out of 85 studies published between 2006 and 2015, the majority focused on grammar (69%) and/or vocabulary (53%), whereas pronunciation (13%) and pragmatics (6%) were only marginally examined. A lack of pragmatics research in TBLT was also emphasized by Van den Branden, Bygate, and Norris (2009), who claimed that the impact of task performance on L2 learning had been identified mainly for lexico-grammar features. Plonsky and Kim also reported that in terms of the analysis units, task performance was often examined in light of task-based interaction (62%) or was analyzed for complexity (22%), accuracy (42%), and fluency (26%), leaving pragmatic features almost completely unexplored. In terms of task-based interaction, analyses of language-related episodes, form-focused episodes, and different types of corrective feedback were included.

In response to these research gaps, researchers have begun to expand the domain of TBLT research, going beyond morpho-syntax. For instance, a recent special issue in *Studies in Second Language Acquisition* presented studies that focus on task-based pronunciation instruction (Gurzynski-Weiss, Long, & Solon, 2017), and input-based tasks have also begun to be explored (e.g., Brunfaut & Révész, 2013; Shintani, 2016). It is surprising that although TBLT emphasizes pragmatic use of language during task performance, task-based pragmatic language teaching and learning has not been systematically examined. The goal of this volume is to address such research gaps, building on our recent work in this area.

For one example of our efforts at the intersection of TBLT and pragmatics, based on a needs analysis consisting of textbook analysis and class observation, Kim and Taguchi (2015, 2016) examined the role of collaborative writing tasks in teaching learners how to make a request in English among Korean English as a foreign

language (EFL) junior high school students. Their English textbook followed a traditional approach of introducing specific request-making expressions (e.g., "Would you mind…?") as common expressions without further information regarding when and how to use such expressions in a given social context. Building on both TBLT and L2 pragmatics literature, Kim and Taguchi (2015) examined (1) the role of collaborative writing tasks (i.e., a drama script task) in interaction-driven learning, as well as in the acquisition of request-making expressions; and (2) the effects of task complexity on the learning of pragmatics through these tasks. In a follow-up study (Kim & Taguchi, 2016), they examined the effects of both task complexity and pragmatic situational demands on the interaction-driven learning opportunities of request-making expressions.

Their findings suggest a potentially beneficial role for the task-based approach in teaching pragmatics, particularly during learner-learner interaction. Their studies also problematized how textbooks present certain grammar structures as useful expressions according to their meaning rather than highlighting the connections between sociopragmatics and pragmalinguistics involved in the expressions. Kim and Taguchi's drama script writing tasks may not meet Long's definition of tasks, which highlights authentic tasks that are designed based on learner needs of language use outside of classroom contexts. Yet, the drama script task met other general characteristics of pedagogic tasks (e.g., meaning-oriented, context-specific authenticity, and clear task outcomes). This example illustrates the importance of localized task design particularly in foreign language instruction contexts based on the needs of various stakeholders to achieve learning goals (Kim, Jung, & Tracy-Ventura, 2017). However, over the last few decades, researchers have increasingly agreed on the common characteristics of tasks. In the domain of instructed pragmatics, although the term "task" has been widely used, it typically refers to any pedagogical activity used to explicitly or implicitly teach pragmatics. Considering that how to define "task" is one of the critical issues in the TBLT literature (Ellis, 2017), such a loosely defined concept of task in the L2 pragmatics literature makes it difficult to see the transferability of research findings between the two fields. In the following section, we review existing instructional practices in L2 pragmatics and highlight how a task-based approach can inform pragmatics instruction. Please note that we used the term "activity" to refer to instructional materials used in the field of pragmatics because the pragmatics literature adopts different ways of defining tasks which are not in line with TBLT literature.

Instructional activities in L2 pragmatics

The position of pragmatics within applied linguistics and SLA has grown and been consolidated since the 1990's. This growth is evidenced in some two dozen overview

articles, monographs, edited volumes, and teachers' guides solely dedicated to pragmatics teaching and assessment (e.g., Alcón-Soler & Martínez-Flor, 2008; Ishihara & Cohen, 2010; Houck & Tatsuki, 2011; Ross & Kasper, 2014; Taguchi, 2011, 2015; Taguchi & Roever, 2017; Takahashi, 2010). This body of the literature has shown ways of teaching pragmatics and explored methods that are effective in improving L2 learners' pragmatic knowledge.

A variety of constructs and features of pragmatic knowledge have been incorporated into instruction, including speech acts (e.g., making requests, expressing opinions), reactive tokens, discourse strategies, politeness markers (e.g., modals), honorifics, interactional particles, speech styles, address forms, and hedging. The importance of these features clearly indicates that learning pragmatics involves more than just focus-on-forms activities. Pragmatics involves a close connection among linguistic forms, communicative functions, and contexts of use. For instance, sentence-final particles are salient grammatical forms in colloquial Japanese, but they are more than just grammatical forms. They are *pragmalinguistic forms* because people use particles for performing various interpersonal functions, such as expressing acknowledgment and showing alignment. At the same time, these particles and their communicative functions are shaped by the context of use. Formality of a setting, interlocutors' social positions, familiarity, and affective stance influence our decision about which particles to use and to what extent. In addition, people need to consider the impact of their particle use on the interlocutor because these particles project certain social meanings (e.g., solidarity, affect), which directly impact the interlocutor's perception of the speaker. Hence, contextual elements to consider in pragmatics are wide-ranging, including external-physical conditions, as well as intra- and inter-personal elements.

Previous instructional studies on pragmatics have primarily asked what kinds of instructional activities and procedures can best facilitate the learning of pragmatics. In response to this question, Taguchi (2015) reviewed 58 instructional intervention studies published since the 1990s. Her synthesis analyzed findings on the effectiveness of pragmatics instruction by exploring common patterns that emerge among them. In this chapter, we re-analyzed the 58 studies in Taguchi's review by coding them for instructional activities and procedures (hereafter instructional features) identified in each study. Instructional features are defined as activities and procedures used purposefully to improve learners' knowledge of target pragmatic features. For example, direct metapragmatic explanation is an explicit instructional feature (part of instructional procedures) that is used for the purpose of developing learners' knowledge of form-function-context associations. At the more implicit end of teaching approaches, a consciousness-raising task that is designed to draw learners' attention to pragmatics without direct explanation, is also coded as an instructional feature.

Our analysis presents a general trend of instructional practice in L2 pragmatics. We will discuss this trend in comparison to the definition, design, and implementation

of tasks in TBLT established in the previous section. Our discussion will focus on how characteristics of common instructional activities in pragmatics (which some researchers call 'tasks') compare to those in TBLT research, what discrepancies emerge from the comparison, and how the fields of pragmatics and TBLT can provide a complementary definition of task and task design in teaching and assessing pragmatics.

To be clear, the instructional activities and procedures we review in the next section are often called 'tasks' and 'task procedures' by L2 pragmatics researchers. However, they do not always share characteristics of tasks as established in the TBLT literature (e.g., Long, 2015; Van den Branden, 2006). Hence, the fundamental goal of the review is to problematize the definition of a task in the current L2 pragmatics literature. But let us first review the existing practice, so the problem – and the need to adapt a TBLT framework for designing tasks – becomes clear.

Coding and analysis of instructional features

Before presenting findings on common instructional features, definitions of coding terms are in order. A close examination of 58 individual studies yielded five primary features of instruction: (1) input, (2) metapragmatic information, (3) production activities, (4) consciousness-raising activities, and (5) feedback. These instructional features encompass both instructional activities and procedures, and are defined as follows:

1. Input: text-based or audio-visual materials that contain pragmatic features.
2. Metapragmatic information: explicit explanation of pragmatic features and use.
3. Production activities: activities used to elicit production of pragmatic forms.
4. Consciousness-raising activities: activities used to draw learners' attention to pragmatic features without explicit metapragmatic information.
5. Feedback: explicit or implicit feedback on learners' use of pragmatic features.

We coded the presence or absence of these instructional features in the individual studies by reading the descriptions of the teaching materials and procedures used in each study. These features are not always mutually exclusive because they can co-occur within the same instructional phase. For instance, a dialogue used to provide input of target pragmatic forms can also be used for direct explanation of those forms. Similarly, when learners practice pragmatic forms using a production activity (e.g., role play), implicit feedback can be provided via recasts. Hence, we coded these instructional features as discrete elements, rather than discrete stages of instruction. We should also note that in some cases these features appeared more than once

in a study, but we coded it as one appearance. For instance, some studies provided explicit metapragmatic information first, followed by production-based practice that also included explicit information as a review. Because we focused on the presence of the explicit information as a feature of the element, not the frequency of occurrence within or across the elements, this example was counted as one occurrence of metapragmatic information.

Trends of instructional features in L2 pragmatics

Figure 1 displays tallied occurrences of the instructional features involved in the treatment conditions in the 58 studies analyzed (see Taguchi, 2015 for the list of the primary studies). Of these 58 studies, 27 of them compared two or more treatment conditions (e.g., explicit vs. implicit teaching). For these studies, we analyzed each treatment condition for instructional features. As a result, the total number of treatment conditions we analyzed was 91 (31 conditions from studies involving one treatment condition and 60 conditions from 27 studies involving two or more conditions).
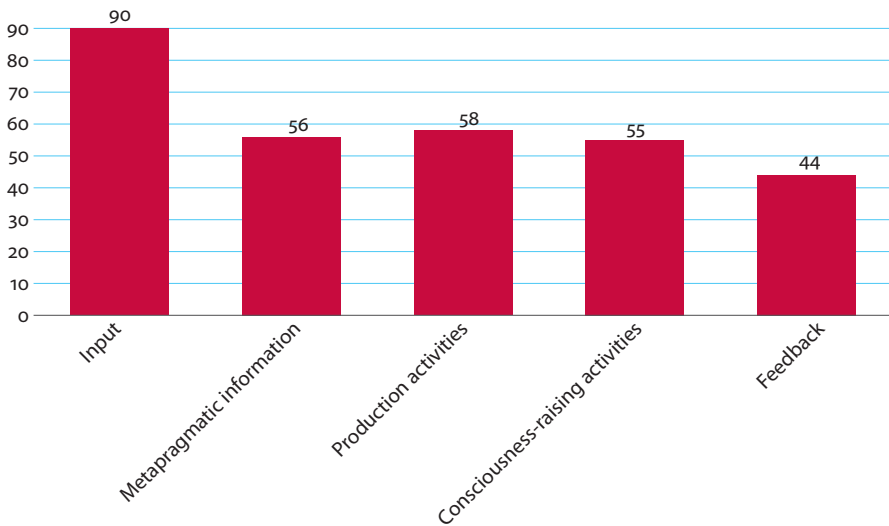
Figure 1. Frequency of instructional features (activities and procedures) in 91 treatment conditions

As shown in Figure 1, most studies involve pragmatics-focused input as one of their instructional features (appearing in 90 out of 91 treatment conditions). Direct explanation about pragmatics is another prominent feature (56 occurrences), suggesting that the majority of the studies used an explicit teaching approach. Production activities are also popular (e.g., form-focused speaking practice, role play). These production

activities usually follow direct metapragmatic information for the purpose of reinforcing understanding of the information.

While production activities are common, studies have also widely used consciousness-raising activities that do not involve production practice (55 occurrences). This type of activity is often used in combination with other instructional features. For example, some studies used video viewing, dialogue analysis and cross-linguistic comparisons for initial awareness raising, followed by a direct explanation to confirm learners' understanding of pragmatic targets (Martinez-Flor, 2008).

Additionally, unlike experimental studies in TBLT that compare different treatment conditions by manipulating specific task features, studies in pragmatics seem to take the maximum benefit approach. Instruction is designed in a way that it produces maximum learning outcomes by adding up all the activities that are considered to be potentially useful. However, this approach does not allow us to directly examine which component or feature of instruction is responsible for any positive learning outcomes, which diverges from the mainstream practice of TBLT research.

Finally, feedback is another common instructional procedure (44 occurrences), but it typically plays only a supplementary role. Aside from a few studies that directly tested the efficacy of implicit feedback (e.g., Fukuya & Martinez-Flor, 2008), feedback alone is rarely a characteristic feature of pragmatic instruction. Feedback is typically embedded in other components (e.g., production activities and metapragmatic discussion) and is often mentioned only as part of the descriptions of instruction.

In the following section we present sample instructional activities from representative studies. Our illustration reveals how these activities differ from the tasks used in the mainstream TBLT research. To clearly illustrate the contrast between tasks used in the mainstream TBLT research and L2 pragmatics research, we will focus on production activities.

Production activities span the continuum between controlled, structured practice of discrete pragmatic items, to more authentic and communicative use of pragmatics in interaction. Among the activities at the controlled and structured end are activities such as provision of pragmalinguistic forms in texts (Nguyen, Pham, & Pham, 2012), computer-delivered systematic production practice of pragmatic features (Utashiro & Kawai, 2009), and discourse completion tests in which learners read a scenario and produce a pragmalinguistic form in writing (Eslami-Rasekh, Eslami-Rasekh, & Fatahi, 2004) or speaking (Safont, 2004).

As a less controlled but still structured activity, role plays appear frequently in the existing studies. There are two types of role plays, open and closed. Kasper and Dahl's (1991) original definition distinguished between these two according to the degree of interaction elicited. In closed role plays, participants act out the set situational description by responding to the interlocutor's standardized prompt. In contrast, open role plays specify the initial situation (e.g., character roles, settings), but there are no

specified outcomes given for the situation. Because the end result of the communicative act is not predetermined, open role plays are considered to elicit a longer exchange over multiple turns and discourse phases (Kasper & Rose, 2002). This is because in open role plays "sequential organization is contingent on the interlocutor's uptake" (Kasper & Rose, 2002, p. 87), as the speaker and listener coordinate their contributions through turn taking.

In a typical role play, participants read a scenario and act out the scenario with an assigned interlocutor (usually a peer learner). The scenarios are written in a way that they elicit the target pragmatics language use (e.g., speech acts, discourse markers, reactive tokens). Learners are often presented with several scenarios involving different contextual dynamics. By role playing different roles and settings that characterize diverse communicative events, learners understand how their linguistic behaviors change corresponding to differing contextual variables (e.g., interlocutor relationship, social distance, and degree of formality). Although role plays may simulate naturalistic interaction more closely than the controlled production activities described above, the interaction elicited in role plays is still constrained because participants are asked to act out a situation while taking on imagined roles (Taguchi & Roever, 2017).

So far we have discussed structured production activities that focus on the production of discrete pragmatic items. From TBLT perspectives, some of these activities are not considered "tasks," as they do not involve authentic goal-oriented outcomes. In TBLT, authenticity can be determined based on both text authenticity (i.e., whether the task input is authentic) and task authenticity (i.e., whether the task process and outcome are authentic). However, what is unique in pragmatics is that text authenticity is part of situational authenticity because input in pragmatics is part of a social context. Although instructional materials used for teaching pragmatics are more like activities rather than tasks, these activities strive to achieve situational authenticity by incorporating realistic scenarios that learners can practice.

At the opposite end of the spectrum of production activities we can find interactive real-life simulations such as mock job interviews (Louw, Derwing, & Abbott, 2010), telecollaboration (Kakegawa, 2009), and video conferencing (Cunningham, 2016; Cunningham & Vyatkina, 2012; Sardegna & Molle, 2010). These activities are considered open activities, but they are not totally control-free. They arrange a space where learners use targeted pragmalinguistic forms in context, so learning occurs as a byproduct of goal accomplishment. For example, in Cunningham's (2016) study, L2 German learners first received explicit information about German request-making forms. Then the researcher arranged a web-conferencing session between learners and German-speaking professionals in Germany. The request-making forms served as critical linguistic resources for learners to communicate appropriately with real-life professionals. Learners also prepared discussion questions relevant to the professionals' fields of expertise, so the task had a clear non-linguistic outcome, that is, gaining

knowledge about the expertise areas. This activity shares characteristics of a task in TBLT research because it is designed to engage learners in the use of pragmatic language for authentic communicative purposes. Similarly, in Louw et al.'s (2010) study, L2 learners in Canada participated in a practice job interview in L2 English with real recruiters using authentic questions. Then, the candidates were asked to watch their interview performance and compare it with that of native speakers. They discussed similarities and differences focusing on the types of interview questions and expected responses. This activity was authentic in that it involved real-life players (i.e., recruiters) with real-life consequences (i.e., job interview outcomes). These activities meet the task characteristics in TBLT (e.g., Ellis, 2003) in terms of authenticity and goal-oriented outcomes. Although these two examples of instructional tasks are notable, to date, L2 pragmatics studies that have adopted such tasks are rare. Furthermore, although task sequencing is one of the main concerns in instructional design in TBLT, L2 pragmatics studies rarely address task sequence with pragmatics features built in as a focus, choosing instead to focus on discrete tasks as learning events.

With the exception of a few studies cited above (and also Kim & Taguchi, 2015, 2016 cited in the previous section), most studies have designed instructional activities as discrete units of language practice without an attempt to situate the activities as a part of the regular course curriculum with real-world language use. Hence, although pragmatics highlights the use of language in real-life social contexts, the instructional materials in previous studies do not always meet the requirements of real-life language use. Thus, issues of transfer of learning or relevance to learners' post-instruction real-life language use have not been addressed extensively in the literature, as TBLT researchers have noted in the field in general (Ellis, 2017; Long, 2016). More effort is necessary to link instructional contexts and authentic communicative contexts via task design and task implementation. This is critical in pragmatics because the very features of pragmatics – knowledge of sociocultural behaviors, norms of interaction, and conventions of language use – are central to everyday language use. Because pragmatic knowledge learned through instruction has the potential for great consequences during real-life interactions, the authentic community has to be configured into task design. To bring pragmatics more closely into TBLT, researchers in L2 pragmatics need to be creative in designing tasks that simulate real-life communication.

Tasks in TBLT and pragmatics teaching: What each field can offer

The previous section presented the nature of instructional activities and procedures used in the 58 instructional studies reviewed in Taguchi's (2015) synthesis of L2 pragmatics instructional research. Pragmatics instruction involves a number of distinct features and elements as listed below. These features have been implemented in unison as an instructional package in many instructional studies.

–   Pragmatic input that contains target pragmalinguistic forms and contextual factors presented in scripted dialogues and texts (e.g., conversations, film excerpts);
–   Explicit metapramatic explanations about linguistic forms, functional meanings, and relevant contextual features;
–   Consciousness-raising activities that aim at promoting learners' inductive learning of target form-function-context associations via comparison, appropriateness judgment, and metapragmatic discussion;
–   Production activities in which learners practice target pragmatic features in structured output exercises, role plays, and tasks that simulate real-life interactions.

These trends are also notable in several studies published after Taguchi's (2015) synthesis. To illustrate a few, Sydorenko (2015) implemented a computer-delivered role play designed to elicit request-making forms in L2 English. Learners watched a series of video-recorded scenarios. The computer stopped the video in pre-planned places and asked learners to provide a response to the interlocutor appearing in the video. Another example of computer-based simulation activities is found in Taguchi, Li, and Tang's (2017) study, which used a game-based virtual interactive space for teaching Chinese formulaic expressions. L2 Chinese learners interacted with a character in a video via text-based chat as they completed a goal-oriented dialogue.

When we compare these trends in pragmatics teaching with those in the TBLT research domain, we can observe how these two fields converge and diverge in their conceptualization and use of pedagogical tasks. The convergence is found most clearly in the influence of the communicative language teaching approach (CLT; e.g., Richards, 2006). Both fields have been greatly informed by CLT in the development of pedagogical tasks in that both fields emphasize the importance of functional language use for communicative purposes in language pedagogy.

CLT has been informative for pragmatics instruction because it promotes the teaching of functional language use in social interactions (Littlewood, 1981). Following the CLT principles of meaning-based language use in a collaborative context, several pragmatics-focused tasks have been developed, incorporating the key elements of pragmatics (social context, functional language use, and interaction) into task design (for example, see Cunningham, 2016; Louw et al., 2010 cited in the previous section).

Similarly, in TBLT research, various collaborative tasks have been implemented to promote L2 development. Although studies reviewed in Plonsky and Kim (2016) focused on the development of L2 grammar and vocabulary through tasks, recently, collaborative writing tasks have been used to teach pragmatics (e.g., Taguchi & Kim, 2016). Furthermore, several researchers have investigated the role of task-related variables (e.g., task complexity, the level of pragmatic demands) in teaching and learning of pragmatics (Kim & Taguchi, 2015, 2016). In addition, assessment researchers have begun to adopt the conceptual framework of TBLT in designing assessment tools that

involve authentic language use through tasks. For instance, Youn (2015) used interactive role-play tasks in assessing L2 pragmatics.

Although the fields of pragmatics and TBLT converge on the point of CLT, these fields diverge to some extent on the conceptualization and use of tasks. The divergence can be discussed in terms of what each field is lacking in their research practice and how the two fields can be complementary to fill each other's gaps. Below we will point out some of those gaps and offer recommendations as to how L2 pragmatics and TBLT can inform each other and advance the current practice.

1. Pragmatics involves a complex interplay among language, language users, and context of use. Similarly, TBLT encourages learners to internalize new information in relation to its use while understanding its pragmatic purpose (Bygate, 2015). Accordingly, both fields view language use as a social phenomenon, and language learning as socially situated, highlighting a focus on form-function-context associations. To extend the object of study from morpho-syntax to pragmatics, TBLT researchers need to attend to these multiple elements of pragmatics when designing a task and designing a lesson with tasks as the goal of instruction. The form-meaning association, which has been the primary investigation in the past TBLT literature, needs to be extended to the form-meaning-context association by incorporating contextual features (e.g., interlocutor relationships, degree of imposition) as additional layers of task features. In TBLT, tasks are currently understood as goal-oriented, meaning-oriented activities that reflect real-world language use. This definition can be advanced to reflect pragmatics considerations more closely. To do so, TBLT researchers would need to expand the implication of each task feature. For example, a task-related goal is not only about completing a task with clearly defined linguistic and non-linguistic outcomes, but also about how the task is completed and whether the communicative goal has been met in a socio-cultural context – for example, whether participants have produced the intended interpersonal effects on the interlocutor. Similarly, a task can incorporate socially-oriented meanings such as politeness, appropriateness, formality, and directness. Furthermore, when assessing authenticity in pragmatics teaching, different dimensions of authenticity – text, context, and task authenticity – need to be taken into consideration. As noted in the previous section, although researchers in L2 pragmatics have been making an effort to address context authenticity in their task design by incorporating real-life simulations, task authenticity needs to be addressed more explicitly.

2. Activities designed for teaching pragmatics have not been sufficiently authentic, goal-driven, or needs-based. This might be due to the over-reliance on explicit and implicit teaching methods in pragmatics instructional studies, resulting in a lack of attention to other methods for teaching pragmatics. TBLT principles could

help design more authentic tasks that promote the use of language in meaningful social contexts as learners work toward task completion for both linguistic and non-linguistic outcomes.

3. Instructional activities in L2 pragmatics have been developed without much careful consideration as to how characteristics of a specific activity may affect performance and learning outcomes in pragmatics. This is seen in the 'instructional package' approach described in the previous section (using all activities in one package). Without empirical studies on specific aspects of task design and implementation, implications about task impact on learners' performance and learning are limited (Ellis & Shintani, 2014; Long, 2015, 2016; Samuda & Bygate, 2008; Van den Branden et al., 2009). L2 pragmatics research can certainly benefit from the TBLT literature in this respect. Because task design and implementation factors have been widely examined in TBLT and found to affect students' task performance in terms of the quality and quantity of interaction and linguistic performance (see Kim, 2015 for review), L2 pragmatics researchers can adapt insights from TBLT findings to systematically examine which task features lead to certain performance characteristics or learning outcomes. For instance, TBLT research has revealed that task design and implementation characteristics that reflect different cognitive demands (e.g., availability of planning time and task complexity) affect accuracy, fluency, and complexity of L2 production (Ellis, 2005; Robinson, 2011). Such findings can help expand the scope of instructed pragmatics by providing guidelines and criteria to follow for task development.

4. Another shortcoming of current L2 pragmatics research is the limited scope of assessment literature, to which TBLT perspectives should be able to contribute. In many existing studies, pragmatic competence has been assessed based on the common units of speech acts, implicature, and routines by using traditional instruments such as discourse completion tests and role plays. These measures can assess learners' knowledge of pragmatics, but they are not sufficient in assessing pragmatic performance. Although alternative approaches that focus on interaction and discourse have been suggested (see Taguchi & Roever, 2017, for a review), empirical studies are still limited in this area. In addition, assessment measures are designed based on researchers' intuitions, without consideration of the degree of correspondence between task situations and real-life social situations. This is a critical limitation, as pragmatic competence refers to the ability to perform language functions in a social context. On the other hand, in the TBLT literature, task-based performance assessment is receiving increasing attention (e.g., Norris et al., 1998; Norris, 2016; Shehadeh, 2012; Youn, 2015), but compared to task design and implementation features, empirical studies on task-based assessment are still limited. This research gap can be addressed by expanding the volume of empirical investigations that address the assessment of L2 pragmatics from a task-based perspective.

5. Since pragmatics teaching has roots in the CLT approach, the use of target language is assumed to be critical when learning pragmatics. However, this assumption may not be entirely valid from a TBLT perspective. TBLT researchers, particularly those who subscribe to sociocultural theory, have claimed that any of learners' linguistic resources, including their L1, could be viewed as a tool that mediates learning (Lantolf & Thorne, 2007). Several researchers have presented evidence that learners use their L1 skillfully for task management and discussion of linguistic issues (e.g., Storch & Aldosari, 2010). So far only a few studies have explored the role of L1 in promoting the development of learners' knowledge of sociopragmatics and pragmalinguistics (see Taguchi & Kim, 2016, cited in the previous section). Since L1 can be a critical mediating tool, particularly in a foreign language (as opposed to second language) context, task-based pragmatics research can examine how use of L1 can promote L2 pragmatic knowledge.

6. Because there has been noticeable convergence between the field of instructed SLA and TBLT, many learner variables have been examined in TBLT research, such as working memory, aptitude, anxiety, proficiency, and heritage background. Some of these variables might affect the learning process as well as the outcomes of task-based pragmatics learning. For instance, since heritage language learners may have a high degree of cultural knowledge compared to foreign language learners, their task-based learning effects might be different from those of foreign language learners.

7. In addition to task and learner variables, we argue for the importance of expanding the scope of contextual variables affecting pragmatic performance in L2 pragmatics research. Although contextual variables such as power, social distance, and degree of imposition have been the primary concerns among L2 pragmatics researchers, socio-cultural, interactional variables such as face, rights, and obligations can equally influence learners' performance and thus should be examined (see González-Lloret & Ortega in this volume).

The chapters in this volume present an initial step toward addressing some of the suggestions stated above. The scope of pragmatics represented in these chapters is wide-ranging, including: speech acts (e.g., request, refusal, apology, thanking, suggestion), honorifics, formulaic expressions, and pragmatic acts in discourse (e.g., writing to persuade someone to take a specific course of action). These units of pragmatics examined in the chapters closely reflect Thomas's (1983) two dimensions of pragmatics – pragmalinguistics and sociopragmatics – introduced earlier this chapter. Pragmalinguistics refers to knowledge of linguistic forms for performing a communicative function, while sociopragmatics involves knowledge of contextual features, norms of interaction, and social conventions associated with a communicative situation. The chapters in this volume illustrate how these dimensions of pragmatic knowledge can be taught and assessed by using tasks.

Chapters in this book

The purpose of this book is to introduce empirical studies that explore how pragmatics can be addressed from various TBLT-oriented perspectives. Specific goals of the volume are as follows:

1. To explore ways in which TBLT can enable the establishment of an interdisciplinary connection within SLA by explicating and unifying fundamental principles and characteristics across research fields.
2. To exemplify ways in which the teaching and assessment of pragmatics can be integrated into TBLT.
3. To show how TBLT can expand the scope of SLA research beyond lexico-grammatical features so as to include pragmatic-sociolinguistic aspects of language use.
4. To illustrate and explore the TBLT-pragmatics connection in a variety of topics, both new and long-standing (e.g., technology in teaching, instructed SLA, different learning contexts, assessment, and discursive pragmatics).

Following this introduction, chapters in this volume are divided into three sections: (1) teaching pragmatics through tasks: the role of metapragmatic discussion (chapters 2–4); (2) using tasks to elicit pragmatics language use (chapters 5–8); and (3) task-based assessment of pragmatics (chapters 9–12).

The first group of studies (Section 1) implemented focused pedagogic tasks to teach pragmatics. Using a task targeting pragmatics features (e.g., honorifics, speech act strategies), the studies promoted L2 learners' metapragmatic discussion and collaborative talk (Swain & Lapkin, 1998) around pragmatics language use. These studies considered that increased amount of discussion on pragmalinguistics and sociopragmatics could lead to better learning of targeted pragmatic features. Chapter 2, by Minkyung Kim, Hakyoon Lee, and YouJin Kim, presents a study examining how drama script writing tasks provide opportunities for learning honorifics in L2 Korean. This study utilized collaborative writing tasks, which have received increasing attention because of their effectiveness in helping draw students' attention to linguistic features while co-constructing meaning. The unique context of this study involved university-level heritage language learners (HLL) and foreign language learners (FLL) jointly constructing a dialogue involving Korean honorifics. Results showed that both HL and FL learners developed their receptive and productive knowledge of Korean honorifics through collaborative tasks. However, HLLs in HLL-FLL dyads outperformed FLLs in FLL-FLL dyads in their productive knowledge of Korean honorifics. With regards to the amount of learning opportunities, no significant difference was found in the number of PREs between HLL-FLL dyads and FLL-FLL dyads.

Chapter 3, by Eva Alcon-Soler, also addresses pragmalinguistic forms that index formality and politeness (i.e., syntactic and lexical mitigations in the speech act of

request) in a college-level, task-supported instructional setting, but the focus is on the impact of different participatory structures on pragmatics learning. L2 learners of English in a Spanish university were randomly assigned to one of three groups: a student-students interaction group (in which the group was led by a student), a teacher-students interaction group (led by a teacher), and a control group. The first two groups (treatment groups) constructed email messages together in a classroom involving high-imposition requests (e.g., asking a professor for a recommendation letter). Both treatment groups outperformed the control group on the knowledge of request mitigations after the task-based interaction, but the participatory structure affected the degree of interactional engagement during task performance. The student-led group revealed a greater degree of engagement with more students participating in discussion and co-constructing request-making forms.

Chapter 4, by Maria Pia Gomez-Laich and Naoko Taguchi, also explores the quality of task-based interaction, but this study adapted a conversation-analytic (CA) approach to show how task complexity (induced via reasoning demands) influences L2 English learners' interaction patterns in a collaborative writing task. The study was conducted in a college-level composition class where students were studying the rhetoric of persuasion. Students formed pairs and co-constructed a persuasive essay in English in either cognitively simple or complex task condition. Analyses of video-recorded interactions revealed how students co-regulated their performance to achieve a task goal and how the performance differed between the two task conditions. The complex task condition led to more extended negotiation sequences and turn taking, frequent pauses, and hesitant ways of speaking (e.g., use of rising intonation and epistemic markers). These features of extended interaction were considered to promote the learning of persuasive writing.

The second part of the volume includes studies that highlight the use of tasks that promote pragmatic performance and pragmatics learning while completing the tasks. It starts with Chapter 5, by Derek Reagan and Caroline Payant, which explores the effects of task modality (i.e., oral and written) on the development of pragmatic competence of requests by low-intermediate learners of Spanish. Two sections of the same low-intermediate Spanish course were randomly assigned to either the oral task group or the written task group, and completed the same story completion task twice following their group condition. In both tasks, participants completed a two-way information gap in which they sequenced a series of images and created a story showcasing two characters making a request. The findings show that the story completion task facilitated the development of indirect request head acts, internal modifications and external modifications. However, no task modality effect was found.

Chapter 6, by Mayya Levkina, addresses the role of individual differences in task-based pragmatics learning, focusing on proficiency. The study examined the role of e-mail tasks in learning speech acts of apology, justification, and thanking by English

as a foreign language (EFL) learners in Spain from two different proficiency levels (B1 and C1). E-mail tasks were sequenced based on task complexity level following Robinson's SSARC model of task sequencing. Students' performance on the pretest and the posttest, which were also e-mail writing tasks was evaluated using an appropriateness scale and a grammar scale. The findings of the study suggest the benefits of task-based instruction using e-mail writing for all speech acts. However, no proficiency effect was found as both groups showed a similar pattern in their development.

Chapter 7, by Roger Gilabert and Júlia Barón, tested Robinson's task sequencing (i.e., SSARC) model and examined the role of task sequencing (e.g., simple to complex vs. random sequencing) in students' e-mail writing performance. The study further examined the validity of four levels of task complexity (simple, less simple, complex, highly complex) based on a subjective perception questionnaire and expert judgments. A total of 15 expert judges evaluated 60 EFL learners' pragmatic performance of writing four email messages at four different levels of complexity. Pragmatic performance was measured using a holistic rating scale. The findings indicated that their operationalization of task complexity was matched by teachers' perceptions in terms of difficulty and mental effort levels. However, although there was a significant difference between the simple version and the other three versions, no difference was found among the three more complex versions. In terms of task sequencing, the findings did not show a significant difference between the two task sequencing conditions in students' task performance, which did not support the benefits of sequencing tasks from simple to complex.

Chapter 8, by Marta González-Lloret and Lourdes Ortega (last chapter in this section) presents a convincing invitation to further investigate pragmatics learning in the task-based instructional framework by incorporating technology as a way of facilitating such investigation. The chapter intends to advance disciplinary knowledge of how to use technology-mediated TBLT as a guiding framework for materials design and the instruction of L2 pragmatics. The chapter first discusses how TBLT principles and their applications to technology are compatible with principles of interactional pragmatics. Based on this argument, the authors illustrate a variety of ways in which the relationship among pragmatics, tasks, and technology can be realized through pedagogical tools and instructional materials.

The third part of the volume, on task-based assessment, begins with Soo Jung Yoon's chapter on validity in assessing pragmatic competence using role plays incorporating requests and refusals (Chapter 9). The author used conversation analysis (CA) and multi-faceted Rasch measurement (MFRM) to investigate whether task-independent interactional features can be elicited from role play-based interaction and how interaction-specific rating categories can function to evaluate L2 English learners' pragmatic performance. The CA findings indicated that various interactional features emerged regardless of the role-play situations (e.g., adjacency pairs, pre-sequences,

and preference organization). These interactional features were used to develop rating criteria descriptions, and functionality of the rating criteria was assessed.

Chapters 10 and 11 deal with the assessment of pragmatic performance from the well-researched domain of L2 performance (i.e., complexity, accuracy, and fluency of spoken and written production). However, these studies added another layer for assessing L2 production: discourse appropriacy (Chapter 10) and functional adequacy (Chapter 11). In Chapter 10, Monika Ekiert, Andrea Révész, Sofia Lampropoulou, and Eivind Torgersen assessed 80 ESL learners' discourse appropriacy using five task types (complaint, refusal, narrative, advice, and summary) across four proficiency levels. The findings suggest that as the learners' general proficiency increased, ratings of discourse appropriacy also increased. Additionally the level of discourse appropriacy differed between the intermediate and advanced learners. In terms of the relationship between proficiency levels and task types on learners' task performance, although advanced learners' performance did not seem to vary depending on task types, lower level learners' discourse appropriacy differed during different tasks (e.g., refusal and summary tasks were more challenging than the other task types).

In Chapter 11, Folkert Kuiken and Ineke Vedder argue the importance of considering the functional dimension of L2 performance when evaluating task performance. They define the functional dimension of L2 performance as the appropriacy and felicity of the speaker/writer's utterances in a given context. Kuiken and Vedder tested the validity of a rating scale focusing on assessing functional adequacy in L2, which consists of four components (i.e., content, task requirements, comprehensibility, and coherence and cohesion). A total of four non-expert raters rated the oral and written argumentative texts of two groups of university students of Dutch L2 and Italian L2. The results suggested that the scale is a reliable and efficient tool for assessing the functional adequacy of written and oral task performance.

The last chapter by Veronika Timpe-Laughlin (Chapter 12) concludes this book by highlighting future directions in L2 pragmatics and task-based language assessment (TBLA). The chapter illustrates similarities between these two domains, emphasizing the possibilities for employing TBLT as a framework for designing assessment tasks for L2 pragmatics, while noting some challenges in pursuing such an intersect. The chapter first presents the fundamental concepts of task-based research and L2 pragmatics. Then, the chapter reviews current uses of TBLA for assessing L2 pragmatics by showcasing different types of task-based assessments that target pragmatics. Finally, the chapter discusses challenges related to generalizability, reliability, and validity of such assessment practice, as well as issues of task design and task difficulty. The chapter concludes with a strong note that using tasks as a foundation for assessing L2 pragmatics involves a number of benefits that, despite the challenges, are worth pursuing in the future.

The chapters in this volume present an initial effort towards establishing collaboration between TBLT and L2 pragmatics researchers. Study design, tasks, and data analysis methods used in these studies all exemplify ways in which TBLT – pragmatics connections can be realized in theory and in practice. We hope that our volume prompts future discussion regarding how to study, teach, and assess pragmatics from the TBLT perspective.

References

Alcón-Soler, E., & Martínez-Flor, A. (2008). *Investigating pragmatics in foreign language learning, teaching and testing*. Bristol: Multilingual matters.

Brunfaut, T., & Révész, A. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, 35, 31–65. doi:10.1017/S0272263112000678

Bygate, M. (Ed.) (2015). *Domains and directions in the development of TBLT: A decade of plenaries from the international conference*. Amsterdam: John Benjamins. doi:10.1075/tblt.8

Bygate, M. (2015). Introduction. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (pp. xv–xxiv). Amsterdam: John Benjamins. doi:10.1075/tblt.8.004int

Bygate, M. (Ed.) (in press). *Learning language through task repetition*. Amsterdam: John Benjamins.

Bygate, M., Skehan, P., & Swain, M. (Eds.). (2001). *Researching pedagogic tasks: Second language learning, teaching and testing*. London: Longman.

Candlin, C. N. (1987). Towards task-based language learning. In C. N. Candlin & D. F. Murphy (Eds.), *Language learning tasks* (pp. 5–22). London: Prentice Hall.

Carless, D. (2004). Issues in teachers' reinterpretation of a task-based innovation in primary schools. *TESOL Quarterly*, 38, 639–662.

Carless, D. (2007). The suitability of task-based approaches for secondary schools: Perspectives from Hong Kong. *System*, 35, 595–608.

Chaudron, C., Doughty, C. J., Kim, Y., Kong, D., Lee, J., Lee, Y., Long, M., Rivers, R., & Urano, K. (2005). A task-based needs analysis of a tertiary Korean as a foreign language program. In M. Long (Ed.), *Second language needs analysis* (pp. 225–261). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667299.009

Cunningham, J. (2016). Request modification in synchronous computer-mediated communication: The role of focused instruction. *Modern Language Journal*, 100, 484–507.

Cunningham, J., & Vyatkina, N. (2012). Telecollaboration for professional purposes: Towards developing a formal register in the foreign language classroom. *The Canadian Modern Language Review*, 68, 422–450. doi:10.3138/cmlr.1279

East, M. (2012). *Task-based language teaching from the teachers' perspective*. Amsterdam: John Benjamins. doi:10.1075/tblt.3

East, M. (2017). Research intro practice: The task-based approach to instructed second language acquisition. *Language Teaching*, 50, 412–424.

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

Ellis, R. (Ed.). (2005). *Planning and task performance in a second language*. Amsterdam: John Benjamins. doi:10.1075/lllt.11

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19, 221–246. doi:10.1111/j.1473-4192.2009.00231.x

Ellis, R. (2017). Moving task-based language teaching forward. *Language Teaching*, 50, 507–526.

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. London: Routledge.

Eslami-Rasekh, Z., Eslami-Rasekh, A., & Fatahi, A. (2004). The effect of explicit metapragmatic Instruction on speech act awareness of advanced EFL students. *TESL EJ*, 8, 1–12.

Fukuya, Y., & Martinez-Flor, A. (2008). The interactive effects of pragmatic-eliciting tasks and pragmatics instruction. *Foreign Language Annuals*, 41, 478–500.
doi:10.1111/j.1944-9720.2008.tb03308.x

Gurzynski-Weiss, L., Long, A., & Solon, M. (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? *Studies in Second Language Acquisition*, 39, 213–224.

Houck, N., & Tatsuki, D. (2011). *Pragmatics from research to practice: New directions*. Alexandria, VA: TESOL.

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

Kakegawa, T. (2009). Development of the use of Japanese sentence-final particles through email correspondence. In N. Taguchi (Ed.), *Pragmatic competence* (pp. 301–333). Berlin: Mouton de Gruyter.

Kasper, G., & Dahl, M. (1991). Research methods in interlanguage pragmatics. *Studies in Second Language Acquisition*, 13, 215–247. doi:10.1017/S0272263100009955

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

Kim, Y. (2015). The role of tasks as vehicles for learning in classroom interaction. In N. Markee (Ed.), *Handbook of classroom discourse and interaction* (pp. 163–181). Malden, MA: Wiley-Blackwell.

Kim, Y., Jung, Y., & Tracy-Ventura, N. (2017). Implementation of a localized task-based course in an EFL context: A study of students' evolving perceptions. *TESOL Quarterly*, 51, 632–660.

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom context: The role of task complexity. *Modern Language Journal*, 99, 656–677.
doi:10.1111/modl.12273

Kim, Y., & Taguchi, N. (2016). Learner-learner interaction during collaborative pragmatic tasks: The role of cognitive and pragmatic task demands. *Foreign Language Annals*, 49, 42–57.

Lantolf, J., & Thorne, S. L. (2007). Sociocultural theory and second language learning. In. B. Van Patten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 201–224). Mahwah, NJ: Lawrence Erlbaum Associates.

Larsen-Freeman, D., & Anderson, M. (2011). *Techniques and principles in language teaching*. Oxford: Oxford University Press.

Littlewood, W. (1981). *Communicative language teaching: An introduction*. Cambridge: Cambridge University Press.

Long, M. (1985). A role of instruction in second language acquisition: Task based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modeling and assessing second language development* (pp. 77–99). Clevedon: Multilingual Matters.

Long, M. (Ed.) (2005). *Second language needs analysis*. Cambridge: Cambridge University Press.
doi:10.1017/CBO9780511667299

Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley Blackwell.

Long, M. (2016). In defense of tasks and TBLT: Nonissues and real issues. *Annual Review of Applied Linguistics*, 36, 5–33.

Long, M., & Crookes, G. (1993). Units of analysis in syllabus design: The case for task. In G. Crookes & S. Gass (Eds.), *Tasks in pedagogical contexts: Integrating theory and practice* (pp. 9–54). Clevedon: Multilingual Matters.

Louw, K. J., Derwing, T. M., & Abbott, M. L. (2010). Teaching pragmatics to L2 learners for the workplace: The job interview. *The Canadian Modern Language Review*, 66, 739–758. doi:10.3138/cmlr.66.5.739

Martínez-Flor, A. (2008). The effect of an inductive-deductive teaching approach to develop learners' use of request modifiers in the EFL classroom. In E. Alcon-Soler (Ed.), *Learning how to request in an instructed language learning context* (pp. 191–225). Bern: Peter Lang.

Nguyen, T., Pham, T., & Pham, M. (2012). The relative effects of explicit and implicit form-focused instruction on the development of L2 pragmatic competence. *Journal of pragmatics*, 44, 416–434. doi:10.1016/j.pragma.2012.01.003

Norris, J. (2015). Thinking and acting programmatically in task-based language teaching: Essential roles for program evaluation. In M. Bygate (Ed.), *Domain and directions in the development of TBLT* (pp. 27–57). Amsterdam: John Benjamins. doi:10.1075/tblt.8.02nor

Norris, J. M. (2016). Current uses for task-based language assessment. *Annual Review of Applied Linguistics*, 36, 230–244.

Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments* (Technical Report #18). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.

Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics*, 36, 73–97.

Prabhu, N. (1987). *Second language pedagogy*. Oxford: Oxford University Press.

Richards, J. C. (2006). *Communicative language teaching today*. Cambridge: Cambridge University Press.

Richards, J., & Rodgers, T. (2010). *Approaches and methods in language teaching* (2nd ed.) Cambridge: Cambridge University Press.

Ross, S., & Kasper, G. (2013). *Assessing second language pragmatics*. Houndmills: Palgrave MacMillan. doi:10.1057/9781137003522

Robinson, P. (2015). The cognition hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed.), *Domain and directions in the development of TBLT* (pp. 87–122). Amsterdam: John Benjamins. doi:10.1075/tblt.8.04rob

Safont, M. P. (2004). An analysis of EAP learners' pragmatic production: A focus on request forms. *Ibérica*, 8, 23–39.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Houndmills: Palgrave.

Samuda, V. (2015). Tasks, design, and the architecture of pedagogical spaces. In M. Bygate (Ed.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp. 271–302). Amsterdam: John Benjamins. doi:10.1075/tblt.8.10sam

Sardegna, V. G., & Molle, D. (2010). Videoconferencing with strangers: Teaching Japanese EFL students verbal backchannel signals and reactive expressions. *Intercultural Pragmatics*, 7–2, 279–310. doi:10.1515/iprg.2010.013

Shehadeh, A. (2012). Task-based language assessment: Components, development and implementation. In C. Coombe, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to Second Language Assessment* (pp. 156–163). Cambridge: Cambridge University Press.

Shintani, N. (2016). *Input-based tasks in foreign language instruction for young learners*. Amsterdam: John Benjamins.

Skehan, P. (2015). Limited attention capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (Ed.), *Domain and directions in the development of TBLT* (pp. 123–156). Amsterdam: John Benjamins. doi:10.1075/tblt.8.05ske

Storch, N., & Aldosari, A. (2010). Learners' use of first language (Arabic) in pair work in an EFL class. *Language Teaching Research*, 14, 355–375. doi:10.1177/1362168810375362

Sydorenko, T. (2015). The use of computer-delivered structured tasks in pragmatic instruction: An exploratory study. *Intercultural Pragmatics*, 12, 333–362. doi:10.1515/ip-2015-0017

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal*, 82, 320–337 doi:10.1111/j.1540-4781.1998.tb01209.x

Taguchi, N. (2011). Teaching pragmatics: Trends and issues. *Annual Review of Applied Linguistics*, 31, 289–310. doi:10.1017/S0267190511000018

Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. State-of-the-art article. *Language Teaching*, 48, 1–50 doi:10.1017/S0261444814000263

Taguchi, N., & Kim, Y. (2016). Collaborative dialogue in learning pragmatics: Pragmatic-related episodes as an opportunity for learning request-making. *Applied Linguistics*, 26, 416–437.

Taguchi, N., Li, Q., & Tang, X. (2017). Learning Chinese formulaic expressions in a scenario-based interactive environment. *Foreign Language Annals*, 50, 641–660.

Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford: Oxford University Press.

Takahashi, S. (2010). Assessing learnability in second language pragmatics. In A. Trosborg (Ed.), *Handbook of pragmatics VII* (pp. 391–421). Berlin: Mouton de Gruyter.

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4, 91–111. doi:10.1093/applin/4.2.91

Utashiro, T., & Kawai, G. (2009). Blended learning for Japanese reactive tokens: Effects of computer-led, instructor-led, and peer-based instruction. In N. Taguchi (Ed.), *Pragmatic competence* (pp.275–299). Berlin: Mouton de Gruyter.

Van den Branden, K. (Ed.) (2006). *Task-based language teaching in practice*. Cambridge: Cambridge University Press.

Van den Branden, K. (2015). Task-based language education: From theory to practice… and back again. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (pp. 303–320). Amsterdam: John Benjamins. doi:10.1075/tblt.8.11van

Van den Branden, K. (2016). The role of teachers in task-based language education. *Annual Review of Applied Linguistics*, 36, 164–181.

Van den Branden, K., Bygate, M., & Norris, J. M. (Eds.) (2009). *Task-based language teaching: A reader*. Amsterdam: John Benjamins. doi:10.1075/tblt.1

Youn, S. -J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, 32, 199–225. doi:10.1177/0265532214557113

Teaching pragmatics through tasks: The role of metapragmatic discussion

Learning of Korean honorifics through collaborative tasks

Comparing heritage and foreign language learners

Minkyung Kim, Hakyoon Lee & YouJin Kim
Georgia State University

This study examines the effects of collaborative writing tasks on the development of Korean honorifics among heritage language (HL) and foreign language (FL) learners. Participants were 14 HL learners (HLLs) and 32 FL learners (FLLs) of Korean (i.e., 14 HLL-FLL and 9 FLL-FLL dyads) in beginning-level language classrooms at a university in the U.S.A. They completed a pretest, two collaborative writing tasks, immediate posttests, delayed posttests (administered three weeks after immediate posttests), and a post-interaction questionnaire in their regular classes. Learning outcomes were measured by discourse completion tests and acceptability judgment tests. Interactions of HLL-FLL and FLL-FLL dyads were analyzed for the occurrence of pragmatic related episodes (PREs; Taguchi & Kim, 2016). Findings indicated that both HLLs and FLLs developed their receptive and productive knowledge of Korean honorifics over time. It was also found that HLLs in HLL-FLL dyads had an advantage over FLLs in FLL-FLL dyads in improving their productive knowledge of Korean honorifics in the immediate posttests, but not in the delayed posttests. In terms of PRE occurrences, both HLL-FLL and FLL-FLL dyads focused more on honorific nouns and honorific verb suffixes than honorific verbs and honorific subject particles. In addition, HLLs and FLLs alike had positive attitudes towards their partners in completing collaborative tasks. This study highlights the benefits of collaborative writing tasks for learning pragmatics in FL classrooms where HLLs and FLLs coexist.

Introduction

Although much research on task-based language teaching (TBLT) has focused on students' learning of grammar and vocabulary, relatively little attention has been drawn to the effectiveness of TBLT on learning of pragmatics (Plonsky & Kim, 2016; Taguchi, 2015). Furthermore, little research has examined learning Korean pragmatics in classroom contexts from TBLT perspectives. In the current study, we focus on the benefits of collaborative writing tasks for learning Korean honorifics as target pragmatic

features. Korean honorifics, which convey politeness and a formal register of speech, are important and challenging pragmatics targets for Korean language learners (Sohn, 1999). In teaching Korean honorifics, one of the learner variables that are of interest is cultural and heritage background because cultural familiarity may impact learning of honorifics in instructional contexts. The purpose of this study is threefold. First, we examine the effects of collaborative writing tasks in learning Korean honorifics by comparing learning outcomes between heritage language learners (HLLs) and foreign language learners (FLLs). Second, we explore interactional focuses of HLL-FLL and FLL-FLL dyads on Korean honorifics during collaborative tasks. Finally, we compare HLLs and FLLs' perceptions of their partners and collaborative tasks.

Background

Pragmatics and Task-Based Language Teaching

Pragmatics is the study of how language users' linguistic choices affect people's relationships, actions and beliefs (Crystal, 1997). According to Leech (1983), pragmatic competence consists of *pragmalinguistics* (i.e., knowledge and ability to use appropriate linguistic forms needed to convey intended meanings) and *sociopragmatics* (i.e., the broader knowledge of social rules, social norms, and appropriateness and politeness necessary in a social context). To be pragmatically competent, people need to have both types of knowledge. In second language (L2) learning contexts, learners often face difficulties in learning pragmatics because pragmatics involves more than knowing form and meaning and involves culture-specific knowledge of the form-meaning connections (e.g., Taguchi, 2015). Over the last few decades, an increasing amount of research has shown that pragmatics is teachable and indeed should be taught in L2 classrooms (e.g., Kasper & Rose, 2002; Taguchi, 2015).

Although different instructional techniques have been explored in the field of instructed pragmatics, the use of interactive tasks in teaching pragmatics has not been systematically investigated (Taguchi, 2011, 2015). Taguchi (2015) synthesizes previous instructional studies in L2 pragmatics and explains the benefits of explicit instruction over non-instructional contexts in that explicit instruction can lead to the gaining of pragmatic knowledge and the use of learned pragmatic forms. Previous meta-analyses (e.g., Jeon & Kaya, 2006; Takahashi, 2010) similarly concluded that the students who received explicit instruction tend to outperform their non-instructed counterparts, especially when the length of treatment is short. For instance, Alcón-Soler (2007) compared the effectiveness of explicit and implicit treatment on Spanish learners of English in their learning of English request expressions. The explicit group received metapragmatic information on target pragmatic forms and was asked to find examples

of requests in scripts they were given and to justify their choices. The implicit group on the other hand performed awareness-raising activities using input enhancement techniques but did not receive metapragmatic explanations. In the posttest, although both treatment groups outperformed the control group, there was no significant group difference in pragmatic awareness. However, only the explicit group demonstrated long term learning effects measured at the delayed posttest which was provided three weeks after the treatment.

Although explicit teaching has been found to be effective, several studies suggest that pragmatic targets often mediate the benefits of different instructional approaches. Rose and Ng (2001) compared the deductive and inductive method on the effects of teaching English compliments and compliment responses. They found that the inductive approach (which withheld explicit information) was more effective for teaching complimenting, while the deductive approach with explicit explanations more positively affected teaching the sociopragmatic rules of compliment responses (e.g., accepting or rejecting compliments). Critically, although previous studies in L2 pragmatics have implemented different types of instructional activities (which the authors also refer to as 'tasks'), task design factors have not been widely investigated beyond the issues of explicit or implicit instruction.

Extending and expanding the study of task design factors in relation to the learning of L2 pragmatics is of obvious potential relevance to TBLT. One of the main tenets in TBLT is that L2 learners can learn language in the process of completing authentic and communicative tasks (e.g., Long, 2015). While engaging in tasks that have interactional authenticity, L2 learners have opportunities to use target linguistic features in meaningful contexts (Van den Branden, Bygate, & Norris, 2009). Despite various definitions of task, researchers generally agree that "a task is a goal-oriented activity that people undertake and that involves the meaningful use of language" (Van den Branden, 2016, p. 240). In TBLT, a considerable body of research indicates that collaborative tasks encourage learners to negotiate meaning and form, engage them more deeply in language, and thus facilitate their language learning (cf. Swain & Lapkin, 2001).

There has been increasing attention paid towards TBLT approaches to teaching pragmatics on the assumption that teaching of L2 pragmatics can be put into practice by adopting TBLT as a guideline for instructions (e.g., Kim & Taguchi, 2015, 2016; Taguchi & Kim, 2016). Specifically, through the use of authentic, goal-oriented, and meaningful tasks, L2 students can in principle be provided with opportunities to use target L2 pragmatics in meaningful contexts. That is, during communicative tasks L2 students can attend to and perform target L2 pragmatics, which can lead to their learning of pragmalinguistic features and sociopragmatic factors. Taguchi and Kim (2016) suggest that identifying pragmatic related episodes (PREs) is useful to investigate students' learning of pragmatics during collaborative tasks. The authors define

RPE as "any part of language production where learners talk about pragmalinguistics (request-making forms) and sociopragmatics (contextual factors) they are attending, question their pragmatic language use, or correct themselves or others" (p. 424).

A series of recent studies by these authors (Kim & Taguchi, 2015, 2016; Taguchi & Kim, 2016) examined the effects of task-based instructions in learning pragmatics (i.e., request-making expressions) with junior high school students learning English in Korea. In particular, they focused on various task design factors such as collaborative vs. individual work, task complexity, and pragmatic situational demands. Instructional materials included explicit information about target features and collaborative writing tasks (i.e., writing drama scripts based on given scenarios as drama scriptwriters). The effects of pragmatic instruction were measured by using discourse completion tasks for target request-making expressions in a pretest, an immediate posttest, and a delayed posttest that was taken four weeks after the immediate posttest. PRE occurrences during interaction were also analyzed. Findings indicated that collaborative tasks are more beneficial than individual tasks because of negotiation opportunities arising during collaborative tasks (Taguchi & Kim, 2016); that complex tasks promote more productions of PREs during interaction and longer-lasting learning effects over time than simple tasks (Kim & Taguchi, 2015); and that complex tasks facilitate more productions of PREs around sociopragmatic factors, but not around pragmalinguistic forms during interaction than simple tasks (Kim & Taguchi, 2016). These findings support the benefits of collaborative tasks in encouraging learners to discuss pragmatics elements (particularly sociopragmatic elements) and hence promoting learning of pragmatics.

With respect to target languages, much research in the fields of instructed pragmatics and TBLT has focused on English (Plonsky & Kim, 2016; Taguchi, 2015). Because pragmatics essentially involves attending to distinctive contextual, social and cultural aspects of individual languages, the importance of "pragmatic-specific-to-languages" has been emphasized (Taguchi, 2015, p. 38). For instance, when it comes to making polite inquires, in English, modal verbs (e.g., *may* or *can*) are often used to achieve a formal register, while in Korean, a suffix (e.g., *-(u)si*) is added to verbs or an honorific verb (e.g., *yeoccubta*; an honorific verb of 'to ask') is used to indicate the speaker's politeness. Thus, when English speakers learn Korean, they need to learn Korean-specific pragmatics to perform speech acts appropriately. In this respect, investigating learning of pragmatics in an underrepresented language, namely Korean in the current study, can contribute to our understanding of instructed pragmatics.

Heritage language learners during collaborative tasks

In the TBLT literature, the nature and effectiveness of collaborative tasks have been widely investigated in terms of many factors, including task implementation (e.g.,

Kim, 2011), interaction patterns (e.g., Storch, 2002), task modes (e.g., Baralt, 2013), and learner characteristics (e.g., Shin, Lidster, Sabraw, & Yeager, 2015). A relatively less commonly investigated factor in learner characteristics is heritage language (HL) background, in particular, the ways in which heritage language learners (HLLs) carry out collaborative tasks with foreign language learners (FLLs). A HLL is defined as "a student who is raised in a home where a non-English language is spoken, who speaks or merely understands the heritage language, and who is to some degree bilingual in English and the heritage language" (Valdés, 2001, p. 37). HLLs encompass a variety of linguistic, social, cultural, and historical backgrounds (Kondo-Brown, 2005). Additionally, in general they differ from FLLs in that they are more likely to be exposed to their HL outside of institutional learning contexts. Thus they may well have various advantages compared with FLLs in a language class, potentially due to the naturalistic linguistic and cultural backgrounds and the sheer amount of target language they have been exposed to (see Valdés, 2001).

In response to the increasing number of HLLs enrolled in foreign language programs (Montrul, 2010), several studies have compared HL and FL learners' performances during collaborative tasks (Bowles, 2011; Bowles, Adams, & Toth, 2014; Henshaw, 2015). In the context of learning Spanish in US postsecondary education, Bowles et al. (2014) explored interaction patterns between pairs of two Spanish FLLs and pairs of one Spanish HLL and one Spanish FLL during collaborative tasks. Participants were 26 learner dyads in intermediate-level language classrooms at a university in the U.S.A., and the collaborative task was a two-way information exchange. Each dyad's focus on form during interaction was analyzed and their perceptions of their partners were assessed through a post-interaction survey. Results indicated that FLL-HLL pairs were more successful than FLL pairs in solving linguistic problems during interaction. It was also found that both FLLs and HLLs in FLL-HLL pairs thought that FLLs benefited more from the interactions than the other way around. Based on the findings, the authors pointed out that different pedagogical approaches are needed to meet HL and FL learners' different learning needs.

Henshaw (2015) examined not only interaction patterns between Spanish FL and HL learners but also learning outcomes during a collaborative writing task (i.e., writing a narrative in Spanish based on a wordless picture story). Participants were eight HLL-FLL pairs in their fifth-semester course at a university in the U.S.A. Interactions were analyzed by counting the frequency of form-focused episodes (FFEs). Learning outcomes were measured by counting the incorporation of linguistic features from successfully-resolved FFEs in an immediate writing task and a delayed writing task that was taken two weeks after the immediate task. Students' self- and peer-perceptions were also assessed through a post-interaction survey. Results indicated that FLLs produced more FFEs than HLLs, but HLLs resolved FFEs more successfully than FLLs. In terms of learning outcomes, FLLs correctly used linguistic information provided by

their HL partners more than the other way around in the immediate post-task, but not in the delayed post-task. Results from the post-interaction survey suggested that FLLs respected HLLs' expertise to some degree, while some HLLs expressed that they felt uncomfortable when acting as experts. These findings indicate that collaborative tasks between FLLs and HLLs may be more beneficial to FLLs than HLLs.

In summary, despite growing attention to instructed pragmatics and the use of collaborative tasks in language classes, no research has been conducted to examine effects of collaborative tasks in learning pragmatics in Korean as a FL. More generally, little attention has been drawn to how HLLs learn pragmatics in FL classrooms. In addition, although previous research has examined learning outcomes of HLL-FLL interactions (Henshaw, 2015), no studies have compared learning outcomes of HLL-FLL and FLL-FLL interactions.

In the current study, the topic of interest is the learning of Korean honorifics. Korean honorifics are linguistic features that mark a speaker's respect and deference to those in a higher position in hierarchical relations of age, power, social status, and generation (Sohn, 1999). Such honorifics are crucial in practicing the social and cultural value of politeness in Korean-speaking communities. In this respect, in FL classrooms, Korean honorifics are introduced at the beginning levels (Lee, 2011). By the same token, in immigrant families from Korea living in the U.S.A., parents often want their children (i.e., Korean HLLs) to use appropriate Korean honorifics. However, the immigrant parents face challenges in teaching their children honorifics because the children lack the context for their use and may not be exposed to honorifics at home unless they live with their grandparents (Park, 2006). Hence, given their importance in using Korean, our focus is on HL and FL learners' learning Korean honorifics in FL classrooms in the U.S.A.

Purpose of the study

The purpose of the current study then was to examine the development of receptive and productive knowledge of Korean honorifics by comparing learning outcomes of HLLs and FLLs on collaborative writing tasks. Learning outcomes were measured by using two different tests: discourse completion tests (DCTs) and acceptability judgment test (AJTs). DCTs were chosen to track changes in students' use of target forms in output (i.e., productive knowledge), while AJTs were chosen to track changes in students' awareness of pragmatic forms in input (i.e., receptive knowledge; Taguchi, 2015). The study further analyzed interaction data from HLL-FLL and FLL-FLL dyads to track how each dyad used and discussed target Korean honorifics while completing collaborative tasks. Students' learning opportunities through collaborative tasks were operationalized as occurrences of PREs. In addition, learners' perceptions of their partners and collaborative tasks were assessed using a post-interaction survey. Learner perceptions were investigated to supplement interaction and outcome data

for the purpose of data triangulation (Bowles et al., 2014; Kim & McDonough, 2008). Generally, our expected outcomes were that both HLLs and FLLs would develop both of their productive and receptive knowledge of Korean honorifics; that HLLs would develop their knowledge in a greater degree and produce more PREs during collaborative tasks than FLLs potentially due to their exposure to Korean at home from childhood (Valdés, 2001); and that FLLs would benefit from interactions with HLLs (Bowles et al., 2014). The current study was guided by four research questions:

1. Do collaborative writing tasks affect HLLs and FLLs' development of productive knowledge of Korean honorifics? If so, are there any differences in the development of productive knowledge between HLLs and FLLs?
2. Do collaborative writing tasks affect HLLs and FLLs' development of receptive knowledge of Korean honorifics? If so, are there any differences in the development of receptive knowledge between HLLs and FLLs?
3. Are there any differences in the occurrence of PREs between HLL-FLL pairs and FLL-FLL pairs during collaborative writing tasks?
4. Are there any differences in HLLs and FLLs' perceptions of their partners and collaborative tasks?

Methods

Participants

The study included a total of 58 learners of Korean who were enrolled in two second-semester Korean courses at a university in the U.S.A. using convenience sampling. At the university where the data were collected, one or two sections of second-semester Korean courses are open every semester. The number of students enrolled in these courses ranged from 20 to 30 on average with a maximum cap of 30. The two classes were taught by one of the authors of this study. The class met three times a week (each class lasting 50 minutes). Most of the participants had completed the first-semester Korean course, while several HLLs were exempted from the first-semester Korean course. Twelve students were absent at the time when the data were collected, which left us with 46 students for analysis.

On a language background questionnaire, students were asked to identify whether they considered themselves as a Korean HLL, which was defined as someone who grew up in a home where the Korean language was spoken by family members (Valdés, 2001). Among 46 students, 14 students were HLLs and 32 students were FLLs. Although most of the HLLs were exposed to naturalistic Korean input at home and had good listening comprehension skills, they reported that they had not received formal instruction in Korean until they took Korean courses at college. Forty-two

students indicated that their first language (L1) was English, while two students' L1 was Chinese, one student's L1 was Polish, and one student's L1 was Spanish. Participants' average age was 19.56 (SD = 1.04), ranging from 18 to 25. Thirteen students were male and 33 students were female.

Participants' Korean proficiency was assessed using the Test of Proficiency in Korean (TOPIK) for beginning and intermediate learners. Fifteen vocabulary items and five reading items were used with a maximum possible score of 20. HLLs ranged in proficiency scores from six to 20 with a mean of 16.71 (SD = 4.25). FLLs ranged in proficiency scores from three to 16 with a mean of 10.88 (SD = 3.50). There was a significant mean difference between HLLs and FLLs in terms of Korean proficiency level, $t(44)$ = 4.88, $p < .001$. For the collaborative tasks, two different types of dyads were formed: fourteen HLL-FLL dyads and nine FLL-FLL dyads. There was no significant difference in mean Korean proficiency scores between the FLLs in the two types of group, $t(30)$ = .993, p = .329.

Target pragmalinguistic forms: Korean honorifics

Korean honorifics are part of pragmatics because the use of honorifics reflects social distances in age, power, and status, social hierarchy, and/or social roles among the speaker, the hearer, and the referent (Mueller & Jiang, 2013; Sohn, 1999). To use Korean honorifics appropriately, Korean learners need to know how to use pragmalinguistic features and understand sociopragmatic factors, that is, which honorifics forms to use in what context. For example, when students talk with their teacher in Korean, they should use Korean honorifics appropriate to that context. Accordingly, Korean learners are expected to learn Korean honorific expressions from the beginning stages of their learning (Lee, 2011). Korean honorifics include many linguistic features, such as nominal suffixes, honorific case particles, an honorific verb suffix, and honorific lexical verbs and nouns.

In this study, four different types of Korean honorifics were chosen as target forms: (a) an honorific subject particle, -*kkeyse*, (b) an honorific suffix on a verb, -*(u)si*, (c) three honorific nouns, and (d) three honorific verbs (see Table 1 for the target forms). The students had learned the honorific suffix on a verb, -*(u)si*, in their first-semester Korean course, but they had not learned the other target honorific forms. The other target honorific forms (i.e., the subject honorific particle, honorific nouns, and honorific verbs) were taken from the textbook used in the third-semester Korean course, "*Integrated Korean: Beginning 2*" (Cho et al., 2010). Because the target pragmatic forms were beyond participants' regular curriculum in the second-semester course, we were able to minimize extraneous variables (e.g., participants' pre-existing knowledge of honorifics) that might confound the effects of the tasks on the students' learning.

Table 1. Target Korean honorifics

Types of Korean honorifics	Target honorific forms	Basic forms
Honorific suffix on a verb	~(으)시 (~(u)si)	Ø
Honorific subject particle	께서 (kkeyse)	이/가 (i/ka)
Honorific verbs	계시다 (kyesita) "to stay" 주무시다 (cwumwusita) "to sleep" 드시다/잡수시다 (dusita/cabswusita) "to eat"	있다 (itta) 자다 (cada) 먹다 (mekta)
Honorific nouns	연세 (yeonse) "age" 성함 (sengham) "name" 생신 (sayngsin) "birthday"	나이 (nai) 이름 (irum) 생일 (sayngill)

Note. The Yale romanization of Korean was used.

Instructional materials: Collaborative writing tasks

Prior to collaborative task sessions, the instructor introduced Korean honorifics for about 20 minutes. The purpose of the instruction was to make sure that students had linguistic resources needed to complete honorifics-related tasks (Mochizuki & Ortega, 2008). The instructional materials were created based on the third-semester Korean course by Cho et al. (2010). The instruction consisted of explicit teaching of sociopragmatic factors and pragmalinguistic forms of Korean honorifics. A handout was made based on the instruction and was given to the students as a resource for use during the collaborative writing tasks.

Students carried out two drama script tasks with the same partner in two class days. Since the learners in the current study were beginners, collaborative work was considered to be more beneficial for them to complete the writing tasks. Furthermore, since the focus of the study is to examine interaction-driven learning opportunities, collaborative writing tasks were chosen over individual writing tasks. Each task lasted about 30 minutes. The two scenarios were presented to students in English. One scenario illustrated a situation involving a short conversation between a Korean learner and an elderly Korean person. The other scenario illustrated a situation involving a short conversation between a Korean learner and a Korean professor. Following Van den Branden's (2006) definition of tasks (i.e., "an activity in which a person engages in order to attain an objective, and which necessitates the use of language," p. 4), our collaborative tasks were "goal-oriented" because students assumed that they were drama script writers who were to complete the scripting of the given scenes with the purpose of learning and Korean honorifics. Additionally, while completing the scripts, students were involved in "the meaningful use of language" because the two scenarios were plausible encounters in their everyday life and required students to meaningfully use honorific expressions according to the given contexts. One example scenario, in which Susan is asking questions to her Korean friend's grandmother, was:

> In pairs, collaboratively write a Korean drama script. The scene is about Susan running into her Korean friend's (Minsoo) grandmother on the way home. Susan and Minsoo's grandmother met once at Minsoo's birthday party last month. Susan is interested in Minsoo and wants to have a small talk with Minsoo's grandmother. In your script, you must include the following topics: Minsoo's grandmother's name, age, birthday, where she lives, what she eats for breakfast, what time she goes to bed, where her husband is, how often she sees Minsoo, and where she is going now.

While the tasks were intended to be goal-oriented and involve the meaningful use of the target pragmatic forms, we admit that these scenarios included some personal questions (e.g., age and what time she goes to bed) that might not be appropriate in other cultures. However, given the limited number of honorific nouns (e.g., age and name) and honorific verbs (e.g., to sleep and to be), it was inevitable to include some personal questions in order to encourage students to use honorific forms in conversation. Furthermore, in the Korean culture, asking age is considered as an acceptable topic to talk about even between people who meet each other for the first time. This is primarily because language use, including using honorific forms, among Korean people is influenced by age differences.

In their regular Korean class, students often carried out collaborative writing tasks similar to those used in this study. As for grouping in this study, most participants were paired with one of their regular group members to strengthen the ecological validity of the research (Kim & Taguchi, 2015). Hence, the tasks were in line with the participants' regular curriculum, and furthermore the students were accustomed to completing collaborative tasks as a pair activity. During the tasks, students were encouraged to discuss the content and language use with their partners as much as they could and were required to write lines by taking turns (i.e., if one student wrote one line, the other was required to write the following line). Considering that the participants were beginning learners with limited linguistic resources, some Korean words that were not related to Korean honorifics but were needed to complete the tasks were provided as a linguistic support in the tasks. When students asked a question about appropriate Korean honorifics, the instructor did not answer the question directly, but encouraged them to refer to the handout for the answer in order to minimize the effect of teacher involvement. Students' interactions during task performance were audio-recorded and transcribed.

Assessment of learning outcome: Discourse completion test and acceptability judgment test

Students' development of productive and receptive knowledge of Korean honorifics was measured by written discourse completion tests (DCT) and acceptability judgment tests (AJT), each given as pre-, immediate post-, and delayed posttests. Students

were allowed about 10 minutes to complete each test. Previous studies have shown that using a written DCT is appropriate to measure students' knowledge of pragmatics after collaborative writing tasks because the treatment tasks have the same modality of writing as the DCT (Kim & Taguchi, 2015, 2016). We used both AJT and written DCT to examine whether collaborative writing tasks (i.e., productive tasks) impacted the development of both of productive and receptive knowledge.

The DCT contained 15 items: three items for each of the four target honorific elements (i.e., *kkeyse, -(u)si*, honorific nouns, and honorific verbs; 12 items total) and three filler items that were not related to Korean honorifics. Each DCT scenario was written in English to ensure participants' understanding. DCT scenarios presented social distances between two interlocutors similar to those used in the collaborative tasks (e.g., talking to a professor and talking to a grandmother). To minimize the practice effects, two versions of DCTs were created with different characters but in similar conversation contexts. One version was used for the pretest, and the other version was used for the two posttests. The items in each test were randomly ordered. See the example DCT item:

Direction: Imagine that you are having a conversation described below. Complete a conversation in Korean.
Megan is asking Steven what his Korean professor's name is. His professor's name is Kim, Minsung.

Megan: 한국어 교수님의 _____?
hankwuke   kyosunim-ui
Korean   professor-of

Steven: 김민성이에요.
Kimminseong-ieyo
Kim Minsung-is

In this item, when Megan is asking Steven his Korean professor's name, she is required to use an honorific noun, *sengham* ("name"), instead of the corresponding non-honorific noun, *ilum*. This is because when addressing the name of someone whose social position is higher than that of the speaker or someone who is older than the speaker, the speaker needs to use the honorific noun to show respect to that person. With respect to the scoring of the DCT, we focused on the participants' appropriate use of target honorific forms. Two points were given if the honorific forms were appropriately used and grammatically accurate. One point was given if the target honorific forms were attempted, but with the incorrect use or/and misspellings. No point was given if non-target forms were used or the answer was left blank. Thus, for the 12 target items, the possible maximum score of the DCT was 24.

The AJT contained 36 target items and four filler items. The target items were shown in two conditions (i.e., appropriate use and inappropriate use) for each of the four target honorific elements (i.e., *kkeyse, -(u)si*, honorific nouns, and honorific verbs). The appropriate use conditions contained three items for each target honorific element ($k = 12$), while the inappropriate use conditions contained six items for each target honorific element ($k = 24$). Table 2 shows example items for the two conditions. The appropriate use conditions described situations where honorific forms are correctly used (see Example (1) in Table 2). The inappropriate use conditions described situations either where an honorific form was incorrectly used (e.g., using an honorific form to a younger person; see Example (2) in Table 2; $k = 12$) or where an honorific form was not used when it was expected (e.g., not using an honorific form to an older person; see Example (3) in Table 2; $k = 12$). For each AJT item, participants were asked to decide whether the target sentence was appropriate by marking *yes* or *no*. For appropriate sentences ($k = 12$), the number of *yes* answers was counted as correct, while for inappropriate sentences ($k = 24$), the number of *no* answers was counted as correct. The possible maximum score of the AJT was 36. To avoid practice effects, two versions of AJTs were constructed with the same basic sentence patterns but with a different subject for each sentence. One version was used for the pretest, and the other version was used for the two posttests. The items in each test were randomly ordered.

Table 2. Examples for AJT items

Condition		Example item			
Appropriate use	(1)	할머니**께서**	집에서	편지를	쓰**세**요.
		halmeni-kkeyse	*cib-eyse*	*pyenci-lul*	*ssu-sey-yo.*
		Grandmother-NOM(HON)	house-at	letter-ACC	write-HON-POL
Inappropriate use	(2)	여동생의	**생신**은		7월이에요.
		yetongsayng-uy	***sayngsin**-un*		*7wel-i-eyyo.*
		Younger sister-of	birthday(HON)-NOM		July-be-POL
	(3)	할머니**께서**		저녁을	먹어요.
		Halmeni-kkeyse		*cenyek-ul*	*mek-eyo.*
		Grandmother-NOM(HON)		dinner-ACC	eat-POL

Note. Honorific expressions are in bold font. Abbreviations: HON = honorifics; NOM = normative; ACC = accusative; POL = polite ending

Exploration of student perception: Post-interaction questionnaire

To explore participants' perceptions of interactions, a post-interaction questionnaire ($k = 7$) was administered. Each item had a Likert-scale item, followed by an open-ended question. Four items asked how helpful collaborative tasks were in terms of (a) learning Korean in general (e.g., Item 1: "How helpful were drama-script writing tasks with your partner in terms of learning Korean in general? Explain why."),

(b) practicing writing in Korean, (c) practicing speaking in Korean, and (d) learning Korean honorifics. Answer choices for these items ranged from *very helpful* to *not helpful at all* on a five-point scale. One item asked how much students liked working with their partners (ranging from *like it a lot* to *did not like it at all* on a five-point scale). There was another item asking about task difficulty (ranging from *very easy* to *very difficult* on a five-point scale). We also included one item asking whether students thought they had enough time to complete the tasks (ranging from *very enough* to *very insufficient* on a five-point scale).

Data collection procedure

As shown in Figure 1, the data were collected over four weeks during participants' regular Korean class sessions. The data collection started on the eleventh week of the sixteen-week long semester. In the first session, students filled out the background information survey and took the proficiency test. They also received the explicit instructions on target Korean honorifics. In the second session, students took pretests, and then completed the first collaborative writing task. In the third session, students completed the second collaborative writing task, and then took immediate posttests. In the fourth session, students filled out a post-interaction perception questionnaire. Three weeks after the immediate posttests, students took delayed posttests. For test implementation, the DCT preceded the AJT in order to prevent the potential effects of the written input provided in the AJT on completing the DCT.

Session 1 (Day 1)	Background information survey (10 minutes)
	Proficiency test (TOPIK) (20 minutes)
	Explicit instruction on target forms (20 minutes)

↓

| Session 2 (Day 3) | Pretests (DCT 1 & AJT 1; each for 10 min) |
| | Collaborative task 1 (30 minutes) |

↓

| Session 3 (Day 5) | Collaborative task 2 (30 minutes) |
| | Immediate posttests (DCT 2 & AJT 2; each for 10 minutes) |

↓

| Session 4 (Day 8) | Post-interaction perception questionnaire (5 minutes) |

↓

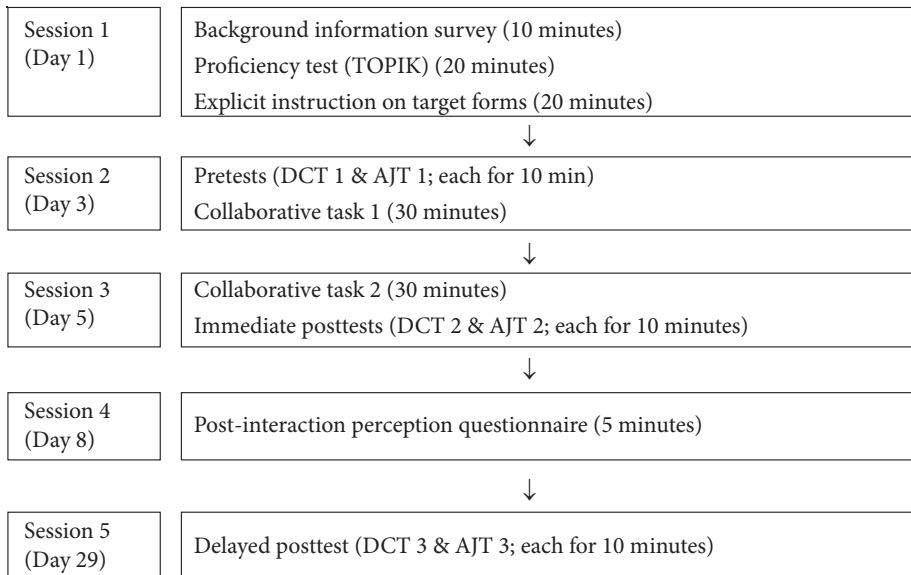| Session 5 (Day 29) | Delayed posttest (DCT 3 & AJT 3; each for 10 minutes) |

Figure 1. Summary of data collection procedures

Pragmatics-Related Episode (PRE) coding

Learners' interaction data during the collaborative tasks were coded and analyzed for occurrences of PREs (Kim & Taguchi, 2015, 2016). PREs were coded for target pragmatics: sociopragmatic factors (e.g., social contexts, speech levels, and character relationships), and pragmalinguistic forms (i.e., honorific verbs and nouns, the subject honorific suffix, and the subject honorific particle). For PREs targeting sociopragmatic factors, the number of PREs, and the person who initiated each sociopragmatic PRE (i.e., triggers) were coded. We did not count who resolved sociopragmatic PREs because attending to sociopragmatic factors was generally expressed through statements (e.g., figuring out the contexts in which honorifics should be used), rather than questions, and thus it did not necessarily involve resolvers of sociopragmatic PREs.

For PREs focusing on pragmalinguistic forms, four pieces of information were coded along with frequency counts of who initiated pragmalinguistic PREs (i.e., triggers), who resolved the pragmalinguistic PREs emerging during interactions (i.e., resolvers), target pragmalinguistic features (i.e., *kkeyse, -(u)si*, honorific nouns, and honorific verbs), and resolution outcomes (i.e., correct or incorrect resolutions).

Example 1 illustrates how students focused on sociopragmatic factors while completing collaborative tasks. In this example, two FLLs discussed how to create a line for the scene (i.e., the grandmother says her husband is at home), considering the formality level of speech.

Example 1. PRE focusing on sociopragmatic factors

    1   FLL1:   Her husband is at home right now. How do you say right now?
    2   FLL2:   지금
                  cikum
                  now
    3   FLL1:   지금   집에
                  cikum   *cib-e*
                  now   house-at
    4   FLL2:   and then existence is 계세요 but that's an honorific form.
                                      kyeseyo
                                      is(HON)-POL
                  You don't say honorific to your husband. You can just say 있어요
                                                          isseoyo
                                                           is-POL
                  that is less formal.

Example 2 illustrates how students discussed pragmalinguistic forms while carrying out collaborative tasks. The HLL explained to the FLL the difference between a non-honorific noun for age, *nai*, and an honorific noun for age, *yeonse* in line 1 below.

Example 2. PRE focusing on pragmalinguistic forms

   1   HLL: You have to say 연세 instead of 나이 because 나이 is for younger
                              yeonse   *nai*   *nai*
                              age(HON)   age   age
                people and 연세 is for older people. I will ask how old you are.
                              yeonse
                              age(HON)

   2   FLL: 연세?
                yeonse
                age(HON)

   3   HLL: You can say, 내 나이는 나이 means age.
                              nae naine-un nai
                              my age-NOM age

   4   FLL: Okay.

For coding PREs, after two authors independently coded around 20% of the transcribed data, they discussed all of the disagreements until they reached agreement. The two authors then coded PREs for the rest of the data.

Statistical analysis

Regarding the first and second research questions (i.e., HLLs and FLLs' development of productive and receptive knowledge of Korean honorifics) and the fourth research question (i.e., HLLs and FLLs' perceptions of their tasks and partners), we first separated participants into three groups: HLLs in HLL-FLL dyads, FLLs in HLL-FLL dyads, and FLLs in FLL-FLL dyads. The results of HLLs and FLLs in HLL-FLL dyads were separately examined because HLLs would differ in completing pragmatic-related tasks and tests from FLLs because HLLs likely have considerable exposure to linguistic and cultural input from their family, which is generally not available to FLLs. In addition, the results of FLLs in HLL-FLL dyads and FLLs in FLL-FLL dyads were separately examined on the assumption that during collaborative tasks, characteristics of interlocutors (HLLs vs. FLLs) would influence the nature of interactions and learning outcomes.

    With respect to statistical analysis related to the first and second research questions, to examine time effects, a Friedman test (i.e., a non-parametric test for data with a within-subject factor and with more than two repeated measures on a single group) was used. If significant time effects were found across pre-, immediate post- and delayed post-tests, post-hoc analysis of Wilcoxon tests (i.e., a non-parametric test for data with two repeated measures on a single group) was conducted. Group differences were examined using a Kruskal-Wallis H test (i.e., a non-parametric test for data with a between-subject factor and with more than two independent groups). If group

differences were found across HLLs in HLL-FLL dyads, FLLs in HLL-FLL dyads, and FLLs in FLL-FLL dyads, post-hoc analysis of Mann-Whitney U tests (i.e., a non-parametric test for data with two independent groups) was conducted.

To answer the third research question (i.e., differences in occurrences of PREs between HLL-FLL and FLL-FLL pairs), we first calculated the number of occurrences of sociopragmatic PREs and pragmalinguistic PREs, respectively. Mann-Whitney U tests were used to examine group differences between HLL-FLL and FLL-FLL pairs in occurrences of sociopragmatic and pragmalinguistic PREs. In addition, to examine interactional patterns in HLL-FLL pairs, we first coded utterances that contained initiations of sociopragmatic PREs, and initiations and resolutions of pragmalinguistic PREs, considering who (either HLLs or FLLs) contributed to each of these coded utterances. Then, chi-square one-way goodness of fit tests were performed to examine whether there were differing distributions of triggers of sociopragmatic PREs (i.e., who initiated PREs), and triggers and resolvers of pragmalinguistic PREs between HLLs and FLLs.

Results

Development of productive knowledge of Korean honorifics

The first research question asked whether HLLs and FLLs developed productive knowledge of Korean honorifics. Data from the fourteen HLL-FLL dyads (i.e., 14 HLLs who worked with FLLs and 14 FLLs who worked with HLLs) and the nine FLL-FLL dyads (i.e., 18 FLLs who worked with FLLs) were included in the analysis. Table 3 shows the descriptive statistics of written DCT results.

Table 3. Descriptive statistics of DCT results

	Pretest			Immediate posttest			Delayed posttest		
	Mean (SD)	Range	%	Mean (SD)	Range	%	Mean (SD)	Range	%
HLL in HLL-FLL (*n* = 14)	3.71 (5.47)	0–17	12.50	11.29 (5.14)	1–18	41.67	7.29 (5.64)	0–18	27.38
FLL in HLL-FLL (*n* = 14)	1.86 (2.18)	0–6	10.71	7.64 (3.20)	3–13	37.20	4.64 (3.99)	0–11	33.32
FLL in FLL-FLL (*n* = 18)	2.89 (3.43)	0–13	12.04	5.56 (5.51)	0–24	23.15	5.89 (5.23)	0–23	34.54
Total	**2.83 (3.87)**	**0–17**	**11.78**	**7.93 (5.28)**	**0–24**	**33.06**	**5.93 (5.02)**	**0–23**	**24.73**

Note. % indicates the average percentage of correct responses (i.e., ratio of students' scores to the possible maximum scores, 24, on average).

Prior to the examination of time effects, we found that there was no difference among the three learner groups in pre-existing knowledge that was measured through the pretest, using Kruskal-Wallis H test ($\chi^2(2)$ = .507, p = .776). A Friedman test revealed a significant effect of time on DCT results ($\chi^2(2)$ = 48.802, p < .001), suggesting significant differences on DCT scores over time. Follow-up pairwise comparisons were conducted to evaluate the time effects. In order to avoid Type I error, alpha level was adjusted to .016 (.05/3). Post-hoc analysis of Wilcoxon tests revealed that immediate posttest scores were significantly higher than pretest scores (Z = −5.682, p < .016); and delayed posttest scores were also significantly higher than pretest scores (Z = −4.613, p < .016) but significantly lower than immediate posttest scores (Z = −2.822, p < .016).

In terms of group differences, a Kruskal-Wallis H test revealed a significant group difference in DCT scores in immediate posttests ($\chi^2(2)$ = 11.858, p < .005), but not in delayed posttests ($\chi^2(2)$ = 1.598, p = .45). Post-hoc analyses of Mann-Whitney U tests revealed a significant difference in immediate posttest scores between HLLs in HLL-FLL and FLLs in FLL-FLL (U = 45, p < .016), but not between HLLs in HLL-FLL and FLLs in HLL-FLL (U = 53, p > .016) or between FLLs in HLL-FLL and FLLs in FLL-FLL (U = 74, p > .016).

In sum, the DCT results showed both immediate and delayed learning effects, such that both HLLs and FLLs improved their productive knowledge of Korean honorifics over time. A group difference was found only in the immediate posttest between HLLs in HLL-FLL and FLLs in FLL-FLL, such that HLLs in HLL-FLL outperformed FLLs in FLL-FLL in the immediate posttest.

Development of receptive knowledge of Korean honorifics

The second research question asked whether HLLs and FLLs developed receptive knowledge of Korean honorifics after the task-based interaction. Table 4 shows the descriptive statistics of AJT results. Similar to the DCT analyses, learning effects over time were examined using a Friedman test, and a Kruskal-Wallis H test was used to examine group differences. A Kruskal-Wallis H test revealed no significant AJT score difference in pretest ($\chi^2(2)$ = .176, p = .916), suggesting no pre-existing significant differences in students' receptive knowledge. In terms of learning effects over time, a Friedman test revealed a significant effect of time on AJT results ($\chi^2(2)$ = 48.356, p < .001). Post-hoc analysis of Wilcoxon tests revealed that immediate posttest scores were significantly higher than pretest scores (Z = −5.353, p < .016) and delayed posttest scores were also significantly higher than pretest scores (Z = −5.657, p < .016), but there was no difference between immediate and delayed posttest scores (Z = −2.204, p > .016). In terms of group effects, Kruskal-Wallis H tests indicated no group

differences in immediate posttests ($\chi^2(2)$ = .680, p = .712) or in delayed posttests ($\chi^2(2)$ = 1.463, p = .481).

Table 4. Descriptive statistics of AJT results

	Pretest			Immediate posttest			Delayed posttest		
	Mean (*SD*)	Range	%	Mean (*SD*)	Range	%	Mean (*SD*)	Range	%
HLL in HLL-FLL (*n* = 14)	15.14 (9.31)	5–30	42.66	22.71 (6.09)	13–31	61.31	23.00 (5.92)	15–33	65.67
FLL in HLL-FLL (*n* = 14)	15.43 (6.84)	4–28	42.26	20.86 (6.92)	9–33	59.72	22.14 (4.99)	12–31	59.72
FLL in FLL-FLL (*n* = 18)	16.00 (7.72)	4–32	44.44	20.11 (6.93)	7–32	55.86	23.56 (5.99)	14–35	65.43
Total	**15.57 (7.75)**	**4–32**	**43.24**	**21.13 (6.63)**	**7–33**	**58.70**	**22.96 (5.59)**	**12–35**	**63.77**

Note. % indicates the average percentage of correct responses (i.e., ratio of students' scores to the possible maximum scores, 36, on average).

In sum, the AJT results showed that both HLLs and FLLs improved their receptive knowledge of Korean honorifics over time, which was in line with the DCT results. The AJT results also indicated no group effect, which means that HLLs and FLLs alike developed their receptive knowledge of Korean honorifics. In comparison between DCT and AJT results, each average percentage of correct responses in AJTs across HLLs and FLLs was higher than that in DCTs. The average percentages of correct responses in pretests, immediate posttests, and delayed posttests in AJTs were 43.24%, 58.70%, and 63.77%, respectively. Meanwhile, the average percentages of correct responses in pretests, immediate posttests, and delayed posttests in DCTs were 11.78%, 33.06%, and 24.73%, respectively.

Occurrence of PREs during collaborative tasks

The third research question asked whether there were differences in occurrences of PREs between HLL-FLL and FLL-FLL pairs during collaborative tasks. Table 5 displays descriptive statistics of PREs. On average, HLL-FLL dyads produced 3.36 sociopragmatic PREs (*SD* = 2.06) and 14.07 pragmalinguistic PREs (*SD* = 6.22), while FLL-FLL dyads produced 2.24 sociopragmatic PREs (*SD* = 1.26) and 10.89 pragmalinguistic PREs (*SD* = 6.24). Mann-Whitney U tests showed that there was no statistically significant difference between HLL-FLL and FLL-FLL pairs in occurrence of sociopragmatic PREs, U = 50, p = .397 or pragmalinguistic PREs, U = 42, p = .185.

Table 5. Mean and standard deviation of frequency of PREs

	Socio-pragmatic factors	Pragmalinguistic forms					PRE total
		HON verb	HON noun	HON verb suffix	HON subject particle	Total	
HLL-FLL pairs ($n = 14$)	3.36 (2.06)	3.50 (2.10)	4.43 (2.38)	5.57 (3.16)	.57 (1.40)	14.07 (6.22)	17.43 (7.45)
FLL-FLL pairs ($n = 9$)	2.44 (1.26)	1.89 (1.79)	2.89 (1.91)	4.11 (1.79)	2.00 (2.79)	10.89 (6.24)	13.33 (6.45)

Note. HON = honorific

As for four types of pragmalinguistic PREs, in general, both HLL-FLL and FLL-FLL dyads focused on the honorific verb suffix the most with a mean occurrence of 5.57 for HLL-FLL and 4.11 for FLL-FLL, and the honorific nouns the second most with a mean occurrence of 4.43 for HLL-FLL and 2.89 for FLL-FLL.

Mann-Whitney U tests showed a marginally significant difference between HLL-FLL and FLL-FLL in occurrence of PREs related to honorific subject particles, $U = 38$, $p = .053$, such that FLL-FLL dyads produced more PREs related to honorific subject particles than HLL-FLL dyads. In addition, the results showed a marginally significant difference in occurrence of PREs related to honorific verbs, $U = 36.5$, $p = .09$, such that HLL-FLL dyads produced more PREs related to honorific verbs than FLL-FLL dyads. There was no significant difference between HLL-FLL and FLL-FLL pairs in occurrence of PREs related to honorific nouns, $U = 41.5$, $p = .171$, or PREs related to the honorific verb suffix, $U = 48$, $p = .342$.

In terms of correct resolutions for pragmalinguistic PREs, the ratio of correct resolutions in HLL-FLL dyads was 87.31% (172 correct resolutions out of 197 occurrences of pragmalinguistic PREs), while the ratio of correct resolutions in FLL-FLL dyads was 77.55% (76 correct resolutions out of 98 occurrences of pragmalinguistic PREs).

As a further analysis on interactional patterns in HLL-FLL pairs, Table 6 shows distributions of triggers of sociopragmatic PREs, and triggers and resolvers of pragmalinguistic PREs between HLLs and FLLs. Results of chi-square one-way goodness of fit tests showed that triggers of sociopragmatic PREs were equally distributed, $\chi^2(1) = 3.60$, $n = 47$, $p > .05$; that triggers of pragmalinguistic PREs were equally distributed, $\chi^2(1) = .86$, $n = 197$, $p > .05$; and that resolvers of pragmalinguistic PREs were not equally distributed, $\chi^2(1) = 20.18$, $n = 197$, $p < .001$, such that HLLs resolved significantly more pragmalinguistic PREs than FLLs did.

Table 6. Total number of occurrences of PREs in the 14 HLL-FLL dyads

	Sociopragmatic factors	Pragmalinguistic forms	
	Triggers	Triggers	Resolvers
HLL ($n = 14$)	17 (1.21)	105 (7.50)	130 (9.29)
FLL ($n = 14$)	30 (2.14)	92 (6.57)	67 (4.79)

Note. Average frequencies of PREs in HLL-FLL dyads are shown in parentheses.

Student perceptions of interactions

The fourth research question asked whether HLLs and FLLs differed in their perceptions of their partners and collaborative tasks. Each student's post-interaction survey data was analyzed. Table 7 displays descriptive statistics of the survey results. In general, in terms of students' perception of helpfulness of their partners, students from all three groups found their partners helpful the most in learning writing in Korean, with a mean of 4.21 for HLLs in HLL-FLL dyads (on a five-point rating scale), 4.43 for FLLs in HLL-FLL dyads, and 4.17 for FLLs in FLL-FLL dyads. Learners from all three groups also liked working with their partner, with a mean of 4.29 for HLLs in HLL-FLL dyads, 4.21 for FLLs in HLL-FLL dyads, and 4.05 for FLLs in FLL-FLL dyads.

Among seven survey items, Kruskal-Wallis H tests showed that there was a statistically significant difference among HLLs in HLL-FLL, FLLs in HLL-FLL, and FLLs in FLL-FLL in students' perception of whether enough time was provided, $\chi^2(2) = 9.767$, $p < .01$. Post-hoc analysis of Mann-Whitney U tests detected the difference between HLLs in HLL-FLL and FLLs in FLL-FLL ($U = 45$, $p < .016$), but not between HLLs in HLL-FLL and FLLs in HLL-FLL ($U = 89$, $p > .016$) or between FLLs in HLL-FLL and FLLs in FLL-FLL ($U = 67$, $p > .016$). This means that HLLs in HLL-FLL dyads thought they had enough time when completing the collaborative tasks in a significantly greater degree than FLLs in FLL-FLL did, while there were no differences in perceptions of time allocation between HLLs in HLL-FLL and FLLs in HLL-FLL or between FLLs in HLL-FLL and FLLs in FLL-FLL.

In terms of the other six survey items, Kruskal-Wallis H tests showed no significant differences in students' perceptions among the three groups: helpfulness of their partners in learning Korean, $\chi^2(2) = 2.64$, $p = .876$; helpfulness of their partners in practicing writing in Korean, $\chi^2(2) = 1.112$, $p = .573$; helpfulness of their partners in practicing speaking in Korean, $\chi^2(2) = 2.873$, $p = .238$; helpfulness of their partners in learning Korean honorifics, $\chi^2(2) = 1.892$, $p = .388$; preferences of their partners, $\chi^2(2) = .175$, $p = .916$; and task easiness, $\chi^2(2) = 4.318$, $p = .115$.

Table 7. Learners' perceptions on their partners and collaborative tasks

	Mean	SD	Min	Max
HLLs in HLL-FLL dyads (*n* = 14)				
Learning Korean	4.07	.47	3	5
Practicing writing in Korean	4.21	.70	3	5
Practicing speaking in Korean	3.93	.62	3	5
Learning Korean honorifics	3.93	.92	2	5
Appreciation of their partner	4.29	.73	3	5
Task easiness	3.21	.58	2	4
Enough time provided	4.00	.88	2	5
FLLs in HLL-FLL dyads (*n* = 14)				
Learning Korean	4.14	.66	3	5
Practicing writing in Korean	4.43	.51	4	5
Practicing speaking in Korean	3.86	.77	3	5
Learning Korean honorifics	3.50	.76	2	5
Appreciation of their partner	4.21	.80	3	5
Task easiness	3.14	.95	2	5
Enough time provided	3.71	1.20	2	5
FLLs in FLL-FLL dyads (*n* = 18)				
Learning Korean	4.11	.76	2	5
Practicing writing in Korean	4.17	.71	3	5
Practicing speaking in Korean	3.33	1.14	1	5
Learning Korean honorifics	3.67	1.14	1	5
Appreciation of their partner	4.05	1.06	1	5
Task easiness	2.59	.93	1	4
Enough time provided	2.76	1.03	1	5

Discussion

The purpose of the current study was to examine the effects of collaborative tasks in learning Korean honorifics in a US foreign language classroom context by comparing HLLs and FLLs. The main findings of this study indicated that both HLLs and FLLs improved their productive and receptive knowledge of Korean honorifics over time. All students showed improvements in both DCT and AJT scores not only from pretests to immediate posttests but also from pretests to delayed posttests, which indicated both immediate and sustained learning effects. A group difference was found

only in the immediate posttest for productive knowledge: HLLs in HLL-FLL dyads outperformed FLLs in FLL-FLL dyads in the immediate DCT. These findings suggest that HLLs and FLLs in HLL-FLL pairs learned Korean honorifics to a similar degree. The findings also suggest that FLLs, regardless of whether they worked with HLLs or FLLs, alike learned Korean honorifics, implying that working with HLLs does not necessarily provide more learning opportunities to FLLs. Thus, in the context of this study (i.e., learning Korean honorifics at the beginning level), overall, the treatment itself (i.e., collaborative tasks) seems to play a more important role than the group variable (i.e., HLL-FLL or FLL-FLL dyads) or the individual variable (i.e., HLLs or FLLs in HLL-FLL pairs). In all, our findings support the benefits of collaborative writing in learning L2 in general (Sato & Ballinger, 2016; Swain & Watanabe, 2013) and in learning L2 pragmatics in particular (Kim & Taguchi, 2015, 2016; Taguchi & Kim, 2016).

In comparing the AJT and DCT results, we found different levels of difficulty between receptive and productive tasks (Bardovi-Harlig, 2009). Specifically, we found a noticeable tendency that proportions of correct responses in the AJTs (i.e., 43.24% in pretests, 58.70% in immediate posttests, and 63.77% in delayed posttests) were higher than those in the DCTs (i.e., 11.78% in pretest, 33.06% in immediate posttest, and 24.73% in delayed posttests). This finding may be explained by different levels of cognitive demands required by recognition and production tasks. In the AJTs, the response format was relatively easy because what students needed to do was to process the given sentences, mainly considering the semantic and pragmatic content of each sentence, and then to decide whether each sentence was appropriate or inappropriate. However, the DCTs required precise orthographic, morphological, and syntactic processing of honorific forms (i.e., retrieving appropriate honorific forms from memory and writing them accurately), which may have been more demanding for beginning learners than the AJTs.

Compared to FLLs, HLLs' advantages in learning pragmatics (i.e., higher proficiency scores and more exposure to naturalistic input at home) were confirmed in two aspects of this study. First, HLLs in HLL-FLL dyads received higher immediate posttest scores than FLLs in FLL-FLL dyads for productive knowledge, but not for receptive knowledge. This finding may be accounted for by Skill Acquisition Theory (DeKeyser, 2007), which suggests the skill-specific nature of L2 learning and transfer appropriate processing, especially during the process of gaining procedualized knowledge. That is, HLLs' higher level of Korean proficiency and greater exposure to naturalistic input may have served as an advantage only when it comes to productive knowledge because the collaborative tasks involved using productive knowledge rather than receptive knowledge (i.e., the productive knowledge test and the treatment tasks had the same modality of writing). Also, reception tests may have been equally manageable to both HLLs and FLLs because of their simple format (i.e., binary choice

of *yes* or *no*), while the production tests may have been more difficult for the FLLs than the HLLs.

Another aspect in which HLLs in HLL-FLL dyads showed an advantage over FLLs in FLL-FLL dyads was found in the post-interaction survey. The HLLs in HLL-FLL dyads reported that they had enough time when completing the tasks to a greater degree than the FLLs in FLL-FLL dyads did. Again, this finding may be attributed to differences in proficiency levels and linguistic and cultural backgrounds: HLLs were likely to have more linguistic and cultural resources that could assist completing the tasks within the time limit than FLLs.

With reference to PRE occurrence during collaborative tasks, HLL-FLL and FLL-FLL pairs did not show statistically significant differences in the occurrence of socio-pragmatic and pragmalinguistic PREs, indicating the lack of group effects on students' focus on sociopragmatic factors and pragmalinguistic features. Interestingly, this lack of group differences in the occurrence of sociopragmatic and pragmalinguistic PREs seems to be in line with the AJT and DCT results that also did not show group effects, except for the immediate posttest for productive knowledge. In terms of occurrence of pragmalinguistic PREs related to each target feature, both HLL-FLL and FLL-FLL dyads focused on honorific verb suffix (*-(u)si*) most, and honorific nouns the second most. These results might be due to students' previous learning experiences (e.g., learning the verb suffix *-(u)si* in their first-semester Korean course) or saliency effects (e.g., honorific nouns might be more salient to beginning learners than other honorific forms, such as subject particles that can be omitted in natural conversation).

On the other hand, HLL-FLL pairs focused on honorific verbs more than FLL-FLL pairs did at the marginally significant level (p = .09). One possible explanation for this difference may lie in the fact that using honorific verbs could be particularly difficult for FLLs. In order to appropriately use honorific verbs, students need to know not only specific honorific verb forms but also inflectional morphological rules for appropriate conjugation. Indeed, several FLL-FLL pairs did not use honorific verbs at all, but instead added *-(u)si* to non-honorific verbs when honorific verbs had to be used (i.e., overgeneralization). However, in HLL-FLL pairs, since HLLs might have been exposed to some honorific verbs at home, they might have used honorific verbs without difficulty. As shown in Example 3, when a FLL did not use an honorific verb for 'to sleep', *cumusida*, a HLL immediately corrected the FLLs' utterance, using the honorific verb with an appropriate conjugation first, and then explaining to the FLL that *cumusida* is an honorific verb for 'to sleep.'

Example 3. PRE focusing on a honorific verb produced by a FLL-HLL dyad

1	FLL:	몇시에	잤어요?
		myeot-si-e	*casseo-yo*
		what-time-at	sleep-POL

2   HLL: 주무셨어요.   몇시에   주무셨어요.
         cwumwus-yeoss-eoyo.   *myeot-si-e*   *cwumwus-yeoss-eoyo*.
         sleep(HON)-Past-POL   what-time-at   sleep(HON)-Past-POL

         주무시다   for   자다.
         cwumwusita   *cata*.
         sleep(HON)   sleep

On the other hand, FLL-FLL pairs focused on the honorific subject particle, *-kkeyse*, more than HLL-FLL pairs did at the marginally significant level ($p = .05$). In fact, while only 14.29% of HLL-FLL pairs (i.e., two out of 14) paid attention to honorific subject particles, 55.56% of FLL-FLL pairs (i.e., five out of nine) did. Perhaps, this difference might be because honorific subject particles were salient particularly to FL learners. One of the most noticeable differences between Korean and English is that Korean has particles to indicate relations of words, such as subjects and objects, within a sentence, while English does not. Due to this cross-linguistic difference, FLLs whose exposure to Korean was limited to classroom contexts might have paid more attention to subject particles than HLLs who might have been accustomed to being exposed to subject particles at home.

With reference to PRE occurrences within HLL-FLL pairs, HLLs and FLLs initiated sociopragmatic and pragmalinguistic PREs at a similar rate, which may indicate that HLLs and FLLs alike paid attention to target pragmatic features. These findings do not seem to support previous research (Henshaw, 2015) which found that FLLs initiated form-focused episodes more often than HLLs. Perhaps these differences might be due to different instructional focuses: while Henshaw's (2015) study asked students to focus on vocabulary and grammar, our study's instructional focus was on pragmatics (i.e., Korean honorifics) that might have appealed to HLLs and drawn their attention to target form-function-context mappings. HLLs might have been motivated to learn Korean honorifics, expecting to be able to use them in talking to their parents or older people in their everyday lives.

In terms of PRE resolutions within HLL-FLL pairs, HLLs resolved pragmalinguistic PREs significantly more than FLLs did. In addition, the ratio of correct resolutions in HLL-FLL dyads (i.e., 87.31%) was higher than the ratio of correct resolutions in FLL-FLL dyads (i.e., 77.55%). Although FLLs in HLL-FLL and FLL-FLL dyads did not show significant differences in learning outcomes, the different ratio of correct resolutions in these two groups may indicate HLLs' contributions to resolving pragmalinguistic PREs. HLLs probably served as "resourceful interlocutors" (Bowles et al., 2014) or "suppliers of information" (Henshaw, 2015) to FLLs during collaborative tasks. Also, findings corroborate previous studies (Blake & Zyzik, 2003; Bowles et al., 2014) in that HLL-FLL dyads produced more target-like outcomes than FLL-FLL dyads. Such findings are also supported by learner perception data. In the open-ended

section of the post-interaction survey, many FLLs who worked with HLLs respected HLLs' expertise, which supports previous studies (Bowles et al., 2014; Henshaw, 2015). For example, from HLL-FLL dyads, one FLL wrote, "My partner knew more than me, so I learned from him," and another FLL wrote, "Luckily, my partner was very understanding and helpful when it comes to things I had trouble with."

With respect to students' perception of their partner, HLLs and FLLs did not show any differences in their perceptions of helpfulness of, and preferences for, their partners. HLLs and FLLs alike considered their partner helpful in learning Korean, practicing speaking and writing in Korean, and learning Korean honorifics with each of mean points above 3.3 (cf. 3 points = *neutral*; and 4 points = *helpful*). Furthermore, HLLs and FLLs alike reported that they liked working with their partners with a mean of above 4 points (cf. 4 points = *liked it*). These findings were complemented by qualitative data in response to the open-ended questions in the post-interaction survey: One HLL wrote, "Learning Korean honorifics helped improve my speaking with adults"; one FLL from the HLL-FLL dyads wrote, "[The task] does help with integrating what we learned into real life situations"; and one FLL from the FLL-FLL dyads wrote, "I learned a lot of new words and expressions, as well as how to use honorifics when speaking." This lack of difference in students' perception of their partner is not in line with previous studies (Bowles et al., 2014; Henshaw, 2015) which found that HLLs and FLLs alike considered their interaction more helpful for the FLLs' learning than for the HLLs' learning. One possible explanation for these different findings might be the fact that, unlike previous studies (Bowles et al., 2014; Henshaw, 2015), in this study, students were paired with one of their regular group members, which ensured high familiarity with their partners. This arrangement probably led the students' more positive attitude towards their partner. Indeed, one student wrote in the survey, "My group is a good group and very easy to work with."

Conclusion

As the first attempt to compare learning outcomes of HLLs and FLLs in learning pragmatics during collaborative tasks, this study found positive evidence of task-based interaction on learning pragmatics for both HLLs and FLLs in a foreign language classroom. Focusing on learning pragmatics from TBLT perspectives, this study contributed to expanding the scope of TBLT approaches beyond the focus on grammar and vocabulary towards pragmatics – language use in social contexts (Kim & Taguchi, 2015, 2016; Plonsky & Kim, 2016). This study also contributed to adding a new dimension to the investigation of pragmatic instructions in languages other than English (Taguchi, 2015) by focusing on Korean, which is a less commonly taught foreign language in the U.S.A.

Many instructors have addressed the concern of mixing HLLs with FLLs in the same language class due to the different linguistic and cultural backgrounds they bring into the classroom (Montrul & Perpiñán, 2011; Potowski, 2002). However, having a separate language class or program for HLLs may not be always a feasible option. Considering that not all language institutions can provide separate courses for HLLs, it is important to address how to incorporate learning needs of these two different types of learners in language classrooms. Based on our findings that HLLs and FLLs alike learned target pragmatics except for the immediate posttest for productive knowledge, this study lends support to accommodating HLLs and FLLs in the same classroom. Furthermore, our findings highlight HLLs' strengths in learner-learner interaction in that they can serve as linguistic and cultural resources to FLLs in learning pragmatics.

This study has some limitations. First, it did not include HLL-HLL dyads. Investigating how HLLs interact with another HLL during collaborative tasks would provide helpful information on their interactional patterns as well as their perspectives on their partners' contributions to the tasks. Second, individual differences, such as motivation and the amount of exposure to Korean culture including social media, were not taken into account in this study. Considering learner characteristics that may influence interaction patterns during collaborative tasks would merit future research. Finally, as the study hinted at the potential for teaching pragmatics to lower level learners with an aid of collaborative tasks, future studies could design a variety of pragmatics-focused tasks targeting beginning-level learners.

References

Alcón-Soler, E. (2007). Fostering EFL learners' awareness of requesting through explicit and implicit consciousness-raising tasks. In M. P. García-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 221–241). Clevedon: Multilingual Matters.

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(4), 689–725. doi: 10.1017/S0272263113000429

Bardovi-Harlig, K. (2009). Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics. *Language Learning*, 59, 755–795. doi: 10.1111/j.1467-9922.2009.00525.x

Blake, R., & Zyzik, E. (2003). Who's helping whom? Learner/heritage speakers' networked discussions in Spanish. *Applied Linguistics*, 24(4), 519–544. doi: 10.1093/applin/24.4.519

Bowles, M. (2011). Exploring the role of modality: L2-heritage learner interactions in the Spanish language classroom. *Heritage Language Journal*, 8(1), 30–65.

Bowles, M. A., Adams, R. J., & Toth, P. D. (2014). A comparison of L2-L2 and L2-heritage learner interactions in Spanish language classrooms. *Modern Language Journal*, 98(2), 497–517. doi: 10.1111/modl.12086

Cho, Y. -M., Lee, H. -S., Schulz, C., Sohn, H. -M., & Sohn, S. -O. (2010). *Integrated Korean: Beginning 2* (2nd ed.). Hawaii, HI: University of Hawaii Press. Cambridge: Cambridge University Press.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–113). Mahwah, NJ: Lawrence Erlbaum Associates.

Henshaw, F. (2015). Learning outcomes of L2-heritage learner interaction: The proof is in the post-tests. *Heritage Language Journal*, 12(3), 245–270.

Jeon, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development. In J. M. Norris, & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 165–211). Amsterdam: John Benjamins. doi:10.1075/lllt.13.10jeo

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

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 61, 100–140. doi:10.1111/j.1467-9922.2011.00644.x

Kim, Y., & McDonough, K. (2008). The effect of interlocutor proficiency on the collaborative dialogue between Korean as a second language learners. *Language Teaching Research*, 12, 211–234. doi:10.1177/1362168807086288

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. doi:10.1111/modl.12273

Kim, Y., & Taguchi, N. (2016). Learner-learner interaction during collaborative pragmatic tasks: The role of cognitive and pragmatic task demands. *Foreign Language Annals*, 49, 42–57.

Kondo-Brown, K. (2005). Differences in language skills: Heritage language learner subgroups and foreign language learners. *Modern Language Journal*, 89(4), 563–581. doi:10.1111/j.1540-4781.2005.00330.x

Lee, E. (2011). A study of instructional content of honorific expressions in Korean language education. *Korean Grammar Education*, 13, 281–315.

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

Long, M. (2015). *Second language acquisition and task-based language teaching*. Chichester: Wiley-Blackwell.

Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in beginning-level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research*, 12(1), 11–37. doi:10.1177/1362168807084492

Montrul, S. (2010). How similar are L2 learners and heritage speakers? Spanish clitics and word order. *Applied Psycholinguistics*, 31, 167–207. doi:10.1017/S014271640999021X

Montrul, S., & Perpiñán, S. (2011). Assessing differences and similarities between instructed heritage language learners and L2 learners in their knowledge of Spanish tense-aspect and mood (TAM) morphology. *Heritage Language Journal*, 8(1), 90–133.

Mueller, J., & Jiang, N. (2013). The acquisition of the Korean honorific affix *(u)si* by advanced L2 learners. *Modern Language Journal*, 97(2), 318–339. doi:10.1111/j.1540-4781.2013.12005.x

Park, E. (2006). Grandparents, grandchildren, and heritage language use in Korean. In K. Kondo-Brown (Ed.), *Heritage language development: Focus on East Asian immigrants* (pp. 57–87). Amsterdam: John Benjamins. doi:10.1075/sibil.32.07par

Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics*, 36, 73–97.

Potowski, K. (2002). Experiences of Spanish heritage speakers in university foreign language courses and implications for teacher training. *ADFL Bulletin*, 33, 35–42. doi:10.1632/adfl.33.3.35

Rose, K. R., & Ng, K. F. (2001). Inductive and deductive teaching of compliments and compliment responses. In K. R. Rose, & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 145–169). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524797.013

Sato, M., & Ballinger, S. (Eds.). (2016). *Peer interaction and second language learning: Pedagogical potential and research agenda*. Amsterdam: John Benjamins.

Shin, S. -Y., Lidster, R., Sabraw, S., & Yeager, R. (2015). The effects of L2 proficiency differences in pairs on idea units in a collaborative text reconstruction task. *Language Teaching Research*, 20(3), 366–386.

Sohn, H. -M. (1999). *The Korean language*. Cambridge: Cambridge University Press.

Storch, N. (2002). Patterns of interaction in ESL pair work. *Language Learning*, 52, 119–158. doi:10.1111/1467-9922.00179

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning and testing* (pp. 99–118). Harlow: Longman.

Swain, M., & Watanabe, Y. (2013). Collaborative dialogue as a source of learning. In C. Chapelle (Ed.), *The encyclopedia of applied linguistic* (pp. 3218–3225). Hoboken, NJ: Wiley-Blackwell.

Taguchi, N. (2011). Teaching pragmatics: Trends and issues. *Annual Review of Applied Linguistics*, 31, 289–310. doi:10.1017/S0267190511000018

Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. *Language Teaching*, 48(1), 1–50. doi:10.1017/S0261444814000263

Taguchi, N., & Kim, Y. (2016). Collaborative dialogue in learning pragmatics: Pragmatic-related episodes as an opportunity for learning request-making. *Applied Linguistics*, 37, 416–437.

Takahashi, S. (2010). Assesing learnability in second language pragmatics. In A. Trosborg (Ed.), *Handbook of pragmatics VII* (pp. 391–421). Berlin: Mouton de Gruyter.

Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. Peyton, J. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 37–80). McHenry: The Center for Applied Linguistics and Delta Systems.

Van den Branden, K. (2006). *Task-based language education: From theory to practice*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667282

Van den Branden, K. (2016). Task-based language teaching. In G. Hall (Ed.), *The Routledge handbook of English language teaching* (pp. 238–251). New York, NY: Routledge.

Van den Branden, K., Bygate, M., & Norris, J. M. (Eds.). (2009). *Task-based language teaching: A reader*. Amsterdam: John Benjamins.

CHAPTER 3

Effects of task supported language teaching on learners' use and knowledge of email request mitigators

Eva Alcón-Soler
Universitat Jaume I

The present study examines whether task supported language teaching (TSLT) has an impact on L2 English learners' use and knowledge of request mitigators, assessing the impact of student-students vs. teacher-students interactions on students' attention to pragmatics during task-based interaction. Forty-eight students of English at a Spanish university participated in the study. Participants were randomly assigned to one of three groups: the student-students interaction group ($N = 16$), the teacher-students interaction group ($N = 16$), and the control group ($N = 16$). Data were collected in pre-test, post-test and delayed post-test email tasks and analysed for frequency of use of request mitigators. Participants' self-evaluations of email appropriateness were also used to examine whether TSLT facilitated knowledge of request mitigators during student-students and teacher-students tasks performance. In addition, interactions during the TSLT treatment were recorded and analysed for pragmatic related episodes (Taguchi & Kim, 2016) on request-making expressions. Findings from the study showed positive effects of TSLT on learners' use of request mitigators. In addition, differences were found in the impact of the participatory structure on students' level of interactional engagement during task performance, which seems to have an impact on pragmatic learning outcomes. More specifically, in teacher-students interaction students hardly ever paid attention towards pragmatics, but if they did, it had an impact on the students' knowledge of request mitigators. On the contrary, metapragmatic discussion in student-students interaction seemed to trigger attention towards pragmatics and enhanced students' awareness of how to mitigate email requests.

Introduction

In the area of instructional pragmatics, previous studies have examined various peda-gogical tasks and instructional methods that can facilitate the learning of pragmat-ics, mainly in the area of speech acts. These studies have typically used discourse completion tasks, role plays and situational scenarios to analyse learners' gains in the

knowledge of speech acts, with the primary interest of comparing explicit and implicit teaching approaches (see Taguchi, 2015, and Takahashi, 2010, for a review on instructional pragmatics). More recently, Nguyen (2013), Taguchi and Kim (2016), and Takimoto (2012) have examined the effect of task-based language teaching (TBLT) on learning pragmatics. In these studies the effects of a peer-to-peer interaction generated in a collaborative task have been examined in relation to pragmatics learning outcomes, but the question remains as to whether the effect can be extended to whole group interaction in a classroom setting.

To answer this question, we investigated the effect of task-supported language teaching and the impact of classroom participatory structure (student-students vs. teacher-students interaction) on learners' use and knowledge of request mitigators in email. We used a collaborative writing task in which L2 English learners co-constructed an email message in group in order to promote negotiation around targeted pragmatic features (i.e., request mitigators), and subsequent noticing and learning of the features.

This chapter is organised as follows: first, we review existing research on instructional pragmatics, together with the guiding theoretical frameworks in most instructional studies, pointing out that our understanding of the applicability of TBLT in the area of pragmatics learning is still limited. Second, we refer to previous research on task performance, focusing on different types of classroom participatory structure, and the benefits of task-based collaborative learning. Third, we report on the study in this paper, including its methodological aspects, the results and discussion of findings. Finally, the conclusions, limitations and pedagogical implications of the study are presented.

Background

Pragmatic instruction

Pragmatics is important regardless of the language learning context, but in formal language learning settings pragmatic instruction becomes a key issue. This may explain the growing interest in research on pragmatics in the classroom after Rose and Kasper (2001) published their collection of studies on the teachability of pragmatics. Since then, classroom-oriented research on instructional pragmatics has provided both teachers and researchers with the possibility of working together and examining whether various pedagogical activities can facilitate the learning of pragmatics. As pointed out by Rose (2005), instructional pragmatics research has addressed two main issues: (a) opportunities for pragmatic learning in the classroom, and (b) the teachability of pragmatics, focusing on the most effective way of teaching pragmatics. In relation to the opportunities found in the classroom for learning pragmatics, Vellenga

(2004) revealed that teachers' talk and textbook materials do not provide the conditions needed for authentic pragmatic input. Looking at the way in which conversations are presented in the textbooks, Bardovi-Harlig (2015) reached a similar conclusion. Consistent with these findings, in the area of language pedagogy, it has been widely acknowledged that simple exposure to input is not enough for learning pragmatics, and cross-cultural comparisons, explicit information, awareness-raising tasks, focused practice and different types of feedback have been suggested as ways to draw learners' attention to pragmatics (see Ishihara & Cohen, 2010, for activities to teach pragmatics). More recently, the use of technology and semi-authentic tasks to enhance learners' exposure to pragmatics (see González-Lloret & Ortega, 2014; Taguchi & Sykes, 2013) has received growing attention, which in turn has advanced our understanding of how pragmatics can be taught.

Regarding the teachability of pragmatics, Schmidt's (1993, 2001) noticing hypothesis has been the primary guiding theoretical framework in most instructional studies (see Jeon & Kaya, 2006; Taguchi, 2011, 2015, and Takahashi, 2010, for an overview). Research findings show that instruction is beneficial for pragmatics learning when learning is assessed immediately after the instruction, but there seems to be no general agreement on the delayed effects of pragmatic instruction (Codina, 2008; Salazar, 2003). In addition, existing findings show that explicit teaching is generally more effective than implicit teaching, although both approaches are effective if implicit teaching involves activities with a focus on noticing and processing (see Taguchi, 2015, for a review of relevant comparative studies).

Although Schmidt's (1993, 2001) noticing hypothesis has been the primary reference point, more recently new frameworks such as input processing (Takimoto, 2012), skill acquisition theory (Li, 2012), and collaborative dialogue (Taguchi & Kim, 2016) have been incorporated as frameworks for instructional pragmatics research. In an attempt to examine the teachability of pragmatics from a new theoretical standpoint, the present study examines whether task-supported language teaching in a classroom can work as a means to develop L2 English learners' knowledge of how to mitigate email requests. Specifically, this study examines a task implementation factor: the impact of teacher-students vs. student-students collaborative dialogue during email writing on developing learners' pragmatic knowledge in request mitigations.

We focus on request mitigations in this study because request modifiers are important in email writing, both to increase politeness and to decrease the potential threatening condition of making a high imposition email request. Previous studies examining email requests during academic consultations revealed that L2 learners do not mitigate sufficiently (Alcón-Soler, 2013; Biesenbach-Lucas, 2006, 2007; Economidou-Kogetsidis, 2009, 2011; Félix-Brasdefer; 2012; Hartford & Bardovi-Harlig, 1996), suggesting the need for teaching learners how to use mitigators to soften the imposition of the request. As described in Blum-Kulka et al. (1989), requests are made up

of two main parts: the head of the request and its peripheral modifiers. The former performs the function of requesting, while the latter mitigate the force of the requests either externally or internally. This study focused on the latter (i.e., mitigators), including softeners (lexical devices that soften the tone, such as *possibly, perhaps, just, maybe, just, kind of…*) and two syntactic mitigators "*Could you…*" (*Could you* please give me some extra days…), and "*I was wondering if…*" (*I was wondering if* you had a tutorial). In addition, we focused on two external mitigators: grounders (*I have to go to the dentist… Could we have a tutorial on Monday?*) and preparators (*I really need to talk to you,* Could we meet another day?).

Task implementation, collaborative dialogue and pragmatic knowledge

Task-Based Language Teaching (TBLT) has received much attention in the area of instructed second language acquisition (SLA) and language pedagogy (Ellis, 2003; García Mayo, 2015; Long, 2015; Kim, 2015; Skehan, 2014; Van den Braden, Bygate & Norris, 2009). Starting with the concept of task, several definitions have been suggested (see Ellis & Shintani, 2014 for an overview of definitions of tasks and task types), but most researchers and language educators would agree that characteristics of tasks involve (1) a focus on meaning, (2) language use in real language use contexts, and (3) goal-oriented outcomes. While completing a task, teachers and students are also encouraged to interact and collaborate on an equal basis to promote the creation of meaning and language learning. Besides, tasks can be conducted individually or collaboratively. Both types of task implementation have been widely examined in the fields of TBLT and instructed SLA. The current study focuses on collaborative task performance in teacher-learners interactions as well as in student-students interactions. Both types of task implementation involve learners' engagement in real-time interaction and collaborative dialogue.

The benefits of task-based collaboration have been addressed widely. Ellis (2003) addresses the benefits of task-based interaction from two research perspectives: the psycholinguistic and sociocultural. The psycholinguistic perspective has focused on how different task types and task conditions may influence learners' performance (Robinson, 2001; Skehan 1998). On the other hand, the sociocultural approach (Vygotsky, 1978) views learning as a mediated process (see Lantoff, 2000, for a review of second language learning as a mediated process), where task performance may trigger knowledge constructed through social interaction.

The present study uses email request tasks, based on the tenet that language is used to achieve communication purposes. Indeed, email writing is considered an authentic task with real life communicative purposes. In line with García Mayo (2014), who dealt with task-based interaction in foreign language learning contexts, this study examined whether task implementation factors, namely student-students vs. teacher-students

interactions, has an impact on constructing pragmatic knowledge of request mitigations during a task-based interaction.

As language educators, we are also aware that teachers' and students' contributions to the performance of a task during classroom interaction are critical for learning. For instance, Prabhu (1987) suggested that the teacher should lead the task with the whole class, while Willis and Willis (2007) were in favour of task performance in small groups. Ellis and Shintani (2014) present several task implementation options: the students can perform the task in groups or pairs, the teacher can guide task performance, that is, whole-class implementation of the task, or one student can play the role of the teacher and perform the task with the rest of the class. Following Ellis and Shintani's options, in the present study, the collaborative task selected, writing an email request to a professor, is guided by the teacher or by one student that takes on the role of the teacher.

Regardless of the type of participatory structure, scholars make an additional distinction in terms of the use of task, that is, task-based language teaching (TBLT) versus task-supported language teaching (TSLT). TBLT is an approach in which the task is central in structuring the syllabus and the lesson, whereas in TSLT the syllabus and generally the lesson too is linguistically defined. Both approaches however accept the use of explicit instruction at some point (see for instance Long, 2015). The present study uses TSLT as a framework for teaching how to mitigate email requests in L2 English (within a linguistically structured programme), with explicit information about how to mitigate email requests being given before engaging learners in email request tasks.

Finally, pedagogy-oriented research argues for the benefits of tasks that require learners to work collaboratively, providing evidence that the collaborative dialogue generated during task performance triggers language use and learning simultaneously (Kim, 2008; Nassaji & Tian, 2010; Storch, 2007). To date, most of these studies have focused on describing how learners engage in collaborative dialogue and whether their engagement leads to learning of grammar and vocabulary (Alegría de la Colina & García Mayo, 2007; Basterrechea & García Mayo, 2013; García Mayo and Azkarai, 2016; Swain & Lapkin, 1995, 2001, 2002). In the area of pragmatics, only a few studies have explored the impact of collaborative dialogue (in a form of metapragmatic discussion) in building pragmatic knowledge.

Following Swain and Lapkin (1995), metapragmatic discussion can be understood as a type of collaborative dialogue that triggers output used for discussing pragmalinguistic forms, as well as for establishing the link between pragmalinguistic and sociopragmatic aspects. For example, studies by Kubota (1995), Alcón (2007) and Nguyen (2013) implemented metapragmatic discussion, in combination with other teaching techniques, as part of the instructional treatment. Takimoto (2012) investigated the effects of two types of consciousness-raising instruction: consciousness-raising with

and without metapragmatic discussion for teaching request mitigators in L2 English. Both groups compared different request forms with contextual features (e.g., interlocutor relationship) and rated the appropriateness of the target request-making forms. Then they came up with a list of ways in which the request could be made more appropriate. One group created the list individually, while the other group did so in collaboration with their peers through metapragmatic discussion. The results showed there were no group differences in the appropriateness judgement of requests, but the group with metapragmatic discussion outperformed the other group on the production of requests.

As Taguchi (2015) noted, Takimoto's work can be interpreted as a study conducted within the theoretical concept of collaborative dialogue, but the author did not examine the nature of collaborative dialogue that occurred during metapragmatic discussion. Taguchi and Kim's (2014) study, on the other hand, examined the nature of collaborative dialogue affecting pragmatics learning. The authors operationalized collaborative dialogue as "pragmatic-related episodes" (PREs) and examined the effect of PREs on learning two aspects of requests: head acts and request modifiers. Participants were divided into three groups: the collaborative group who received metapragmatic instruction on requests, followed by a dialogue construction task (involving request-making) completed in pairs; the individual group who received metapragmatic instruction, followed by the same task completed individually; and the control group. Results showed that the collaborative group outperformed the individual group on the production of the request head acts in an immediate post-test, but no group difference was observed for request mitigators.

Because Taguchi and Kim's (2016) study analysed peer-to-peer interactions only, it is questionable whether the effect of task-based interaction can be extended to a whole group interaction in a classroom setting. A question also remains as to whether the participatory structure of interaction (e.g., student-to-student vs. student-to-teacher) affects instructional outcomes. Hence, the present study investigated a whole class interaction arising from a task involving a classroom instructor and students. Specifically, we compared two types of classroom participatory structures: an instructor interacting with the class and a peer learner interacting with the class. We examined whether students' attention to pragmatic features (i.e., request mitigations in emails) during task performance differs between these two types of classroom interaction, and whether those differences affect learning of request mitigators differently.

To sum up, although TBLT has been influential in many formal instructional contexts, classroom-based studies examining the applicability of TBLT to pragmatics learning are still limited. Thus, in line with the recent call for expanding the theoretical scope of instructed pragmatics (Taguchi, 2015), the present study examines whether task supported language teaching (TSLT) has a positive impact on L2 English learners' knowledge of request mitigators. In doing so, the study also assesses the impact

of different task participatory structure (teacher-students vs. student-students) and students' attention to pragmatic issues during task-based interaction on L2 learning of request mitigators. The investigation of the impact of teacher-led versus student-led classes has not been explicated investigated in the area of TBLT, or in the area of teaching and learning pragmatics.

Research questions

The following research questions are addressed:

1. Is task-supported language teaching effective for learning how to mitigate high-imposition email requests?
2. Does the type of task participatory structure (teacher-students vs. student-students interaction) make a difference in students' attention to pragmatic features during students' task performance and students' subsequent learning of request mitigators?

Method

Participants

Forty-eight students of English as a foreign language, thirty-four females and fourteen males, participated in the study. Their average age was 20.5. Their English proficiency was judged to be upper intermediate based on the standardized Quick Oxford Placement test (U.C.L.E.S., 2001), which is equivalent to Common European Framework level B2 (Verhelst et al., 2009). Participants were all enrolled in the first year of the Degree in Translation at a Valencia University. They were instructed in the minority (Catalan, also referred to as Valencian) and the majority (Spanish) language. Participants in the present study were instructed in English. In English language sessions teachers followed a communicative language teaching approach, which involved using language to achieve communicative outcomes. Students performed different tasks in pairs, small groups, or in lockstep (i.e., teacher-students interaction) during these sessions.

Participants were randomly divided into three groups: the student-students interaction group ($N = 16$), the teacher-students interaction group ($N = 16$), and the control group ($N = 16$). Dividing students in groups for language teaching sessions is a common procedure at university level in Spain. During these sessions teachers encourage classroom interaction, guided by the teacher, or by one student that takes the role of the teacher. The present study was conducted during English language sessions where the three groups performed email requests that were sent to a professor who agreed

to participate as a recipient of the emails. The student-students interaction group first received metapragmatic information on how to produce email requests to their teachers, followed by three task-supported language teaching (TSLT) sessions, where one student took on the role of the teacher and led the whole group's task completion on writing high-imposition email requests (see below for the procedure). In contrast, the teacher-students interaction group received the same metapragmatic information, but the email request tasks were completed in a format of teacher-students classroom interaction, that is, the teacher leading the performance of the email-writing task. The control group did not receive any metapragmatic information or task treatment sessions.

All treatment sessions took place during regular classroom sessions (see information on TSLT sessions below). A female Spanish teacher conducted all the treatment sessions in English. This teacher completed a course on teaching pragmatics, including a session on how to mitigate email requests.

This study took place as part of a typical university English course with a focus on writing. The course involved various units including writing application letters, letters of complains, book reviews, and emails. The classroom instructors carried out the tasks designed for this study (to be described in the methods section) without indicating the aim of the study. Participants were familiar with the whole class interaction led by a teacher or by another student.

Materials

Following Ellis and Shintani's (2014, p. 135) criteria for an activity to be described as a "task", the task selected in the present study, writing a high imposition email request, satisfies the requirements of a task: (i) there is an information gap, students write an email to be read by the professor, and (ii) there is clearly an outcome, asking a specific request to the professor. In addition, students are familiar with addressing high imposition email requests to their teachers during academic consultation in their university. According to Brown and Levinson's (1987) sociopragmatic variables, this task situation was described as involving higher social distance, higher social power, and higher degree of imposition.

In order to make the task authentic, we selected three types of email requests found in Alcon's (2013) corpus-based analysis of requests occurring in academic cyber-consultations in an international high school setting. In her study, participants (60 students studying English in an immersion setting) were asked to indicate the degree of imposition of different types of naturalistic email requests on a five-point Likert scale, ranging from 1 (no imposition) to 5 (very high imposition). The mean imposition score was 4.1 for asking for an extension of a deadline; 4.6 for asking someone to provide a recommendation letter within a day; 4.3 for asking to submit an essay after the

deadline; and 4.1 for asking for an extra tutorial session. Because these requests were perceived as high-imposition requests, they were used in the treatment sessions in this study. The following section presents the sequence of instruction and task-supported language teaching sessions (TSLT).

Procedure

The study lasted eight weeks (January – April 2015) and included a pre-test (week 1), information on email request mitigators (week 2), TSLT treatment sessions (week 3), a post-test (week 4), and a delayed post-test (week 8).

Week 1. *Pre-test*
Learners sent an email to their teachers asking for an extension of a deadline, knowing that the teacher was reluctant to accept the request. Immediately after sending the email, they received an automatic email response asking them to judge the appropriateness of their own email on a Likert scale (1–5) and to give reasons for their choice.

Week 2. *Day 1 and Day 2*
These sessions did not involve actual task performance. Metapragmatic information was provided on the use of internal mitigators including softeners (*possibly, perhaps, just, maybe, just, kind of…*) and two syntactic mitigators "*Could you…*" (*Could you please give me some extra days…?*), and "*I was wondering if*" (*I was wondering if you had a tutorial this week*). In addition, information was provided on two types of external mitigators: grounders (*I have to go to the dentist… Could we have a tutorial on Monday?*) and preparators (*I really need to talk to you. Could we meet another day?*). Then, students were provided with examples of inappropriate emails, followed by the teacher's reconstruction of the emails.

Week 3. *Day 1 (TSLT session 1), Day 2 (TSLT session 2) and Day 3 (TSLT session 3)*
The teacher guided task performance in the teacher-students interaction group. In the student-students group, one student took on the role of the teacher and guided task performance. In both conditions the whole class completed an email request collaboratively.

TSLT sessions
Students wrote collaboratively three emails and sent them to a professor, who agreed to participate as a recipient of the emails. Students were asked to send an email to one of the most strict professors asking for a recommendation letter within one day (TSLT session 1); asking to accept an essay after the deadline; and asking for an extra tutorial session (TSLT session 3). The student-students and the teacher-students interaction

groups collaboratively constructed the emails. Immediately after sending each email, students were asked to judge the appropriateness of the email on a Likert scale (1–5) and to give reasons for their choice. The control group did not take part in the TSLT sessions, but performed the pre-test, post-test and delayed post-test.

Week 4. *Post-test*
Students were asked to perform the same pre-test task. After completing the task, they were asked to judge the appropriateness of their email on a Likert scale (1–5) and to give reasons for their choice.

Week 8. *Delayed post-test*
Students were asked to perform the same pre-test task. After task completion, they were asked to judge the appropriateness of their email on a Likert scale (1–5) and to give reasons for their choice.

Data analysis

To answer Research Question 1, that is, whether TSLT works for learning how to mitigate requests, data were collected from the pre-test, post-test and delayed post-test email tasks, in which participants asked the teacher for an extension of a deadline. A total of 144 email request tasks were analysed (3 emails written by 16 participants in each of the three groups). The frequencies of the softeners, syntactic mitigators, grounders and preparators in the emails were calculated individually by the researcher and a trained research assistant. Each modifier was counted. When two or more modifiers were found in one email request, all of them were coded and counted. Both the researcher and a research assistant coded all the data. They discussed cases of discrepancy and reached an agreement on 95% of the data. The frequency of mitigation devices was calculated to analyze quantitative differences in terms of performance of email request mitigators. Since the data did not confirm a normal distribution, a non-parametric test (Kruskal-Wallis) was used to compare frequency of request mitigators before and after participants completed the tasks (pre-test and post-test), as well as to analyse whether the effects of TSLT were sustained (delayed post-test). The Mann-Whitney U test was used to compare frequency of request mitigators across groups (student-students, teacher-students and control).

To answer Research Question 2, which addressed the extent to which students paid attention to pragmatics in student-students and teacher-students task-based interaction and their subsequent learning of request mitigators, the three TSLT sessions (whole class interactions) were tape-recorded, transcribed and analysed. Following Taguchi and Kim (2016), students' attention to pragmatic issues were operationalized

as Pragmatic-Related Episodes (PREs). The authors (2016) define PREs as "any part of language production where learners talk about the pragmalinguistic forms they are producing and the sociopragmatic factors they are attending to (e.g. setting and inter-locutor relationship), question their pragmatic language use, or correct themselves or others." (p. 4) Instances where PREs were resolved successfully were identified. See Excerpt 1 below for an example. Here in line 7 S1 initiates a PRE pointing out the need to modify the request by giving a reason. This is accepted by the other interlocutor, in line 8, who provides the reason. Then, S1, in line 9, suggests the use of "just", which triggers some negotiation during the next four turns. Finally, in line 14, S2 provides the reason that has been jointly constructed.

Excerpt 1.

1	S1:	We can say that…
2	S2:	I've just received a phone call… and… and asking for it
3	S1:	They need the letter for tomorrow….
4	S2:	No contractions right?
5	S1:	No, maybe we can now say….
6	S2:	Yes
7	S1:	… for you… **we have to give a reason**
8	S2:	Yes, **they have told today** but we…
9	S1:	OK why not with **just**?
10	S2:	Just?
11	S1:	Yes, we have just told **because he is a teacher and I want to be polite**
12	S2:	would….
13	S1:	We have been told
14	S2:	**We have just been told…**

In addition to the occurrences of PREs, we also assigned scores on students' level of engagement during PREs. Engagement was defined as the degree of students' atten-tion to pragmatics. Following Philip and Duchesne (2016), cognitive engagement was operationalized as students' participation in initiating PREs, asking and answering questions, or attempting to resolve PREs. Following this definition and operational-ization, two points were given when one of the class participants initiated the PRE; one point was given when another class participant reacted to the PRE by co-constructing pragmatic knowledge. There were some instances in which more than one student participated in constructing knowledge during the PREs, and in those cases scores were assigned to each individual student. See Excerpt 2 below for an example. Here, S3, who takes on the role of the teacher, initiates the PRE in line 1, and S1, S2 and S6 react by constructing knowledge during the PRE (lines 3, 5, 12, 14 and 16). S1 made five attempts to resolve the PRE triggered by the S3, and thus he received 5 points; S2 formulates a question in line 4 and makes a suggestion in line 7, and thus received

2 points. S6 answers the S3's prompt in line 9, and the S3's question in line 11. This student also interrupts S1, in line 17, by using an extra grounder ("… and really want to write an excellent paper"), obtaining a total of 3 points.

Excerpt 2.

1	S3:	Our aim is to ask for a deadline… so… how can we ask in a polite way?
2		(Silence)
3	S1:	We should give a reason first?
4	S2:	Do we start like I am writing to…?
5	S1:	I am writing to you in order to
6	S3:	Ok, I'll write in the backboard
7	S2:	Could give us an extra week
8	S3:	In order to…
9	S6:	Give us an extra week
10	S3:	You are asking?
11	S6:	Yes, in order to ask you to have an extra week
12	S1:	But we do not meet the deadline, why not I was wondering if…
13	S3:	Ok
14	S1:	If you could give an extra week
15	S3:	If you could give us an extra week
16	S1:	Because we have had a lot of exams this month…
17	S6:	… Yes a lot of exams and we really want to write an excellent paper
18	S3:	Ok I wonder if you could give us an extra week because of the exams we have had and because we are also interested in writing a good paper.
19	S7:	(change topic, asking about how to say a lexical word)

Before data coding, the researcher and a research assistant practised coding together on the data from the pilot study to ensure consistency. Then, the researcher and the research assistant independently coded the data in the main study (all transcripts of classroom interaction data from the email-writing tasks). The agreement rate (based on 30% of the data) was 91% for the successfully resolved PREs and 89% for level of engagement during the email writing tasks.

In addition, in this study, evidence of knowledge of request mitigators generated during PREs was operationalized as explicit reference to the use of request mitigators in students' self-evaluation of the appropriateness of their emails. After constructing an email, students were asked to self-evaluate email appropriateness. See Excerpt 3 below for an example. Here, the student makes the request asking for a recommendation letter by using one intensifier ("really"), a preparatory ("I know you are busy"), one grounder ("because I have just been told about the letter"), one preparator ("I know that I should have written in advance"), and an apology ("Sorry for any inconvenience…"). However, in his self-evaluation of email appropriateness he only makes explicit reference to the use of the grounder ("… I have explained the reason why I did not write before").

Excerpt 3. *Example of an email asking for a recommendation letter*

> Dear XXX (name of the teacher)
>
> I am writing to you because I really need your help. I know you are busy, but I would like to know if you would be willing to write me a recommendation letter for tomorrow. I need to ask you this favour. I know that I should have written in advance, but I have just been told about the letter. Sorry for any inconvenience this may cause you. I would really appreciate if you could do it.
>
> I am looking forward to hearing from you.
>
> Best regards, XXX (name of the student)
>
> *Example of self-evaluation of email appropriateness*
>
> Do you think that this email is appropriate?
>
> 1. not at all
> 2. a little
> 3. so-so
> 4. **quite**
> 5. very
>
> Please explain your choice by referring to the language used in the email.
>
> "In this occasion, I have thought carefully about the language and expressions I was going to use because it was a sensitive issue. I use their names, because this is how I address to them, it is accepted. I wanted to be very polite in order to achieve my goal. Moreover, I have explained the reason why I did not write before".

Two points were given if a request mitigator was used in the email and was mentioned to explain the self-assessed degree of email appropriateness (e.g., the use of a grounder, "because I have just been told about the letter", in the above example). One point was given if a mitigator was used in the email, but it was not mentioned in the evaluation of email appropriateness (the case of 'just' in the example above); and no points were given if mitigators were neither used nor mentioned in students' evaluation of email appropriateness. The researcher and a research assistant coded the data (students' reported knowledge of request mitigators), achieving 93% of agreement rate (based on 30% of the data).

Finally, total scores of reported knowledge of request mitigators and level of engagement in PREs were compared across groups (student-students and teacher-students) by using the Levene test.

Results

The first research question asked whether task-supported language teaching is effective for learning how to mitigate high-imposition email requests. Table 1 and 2 display

descriptive statistics of the frequency of softeners and the syntactic mitigators in the pre, post, and delayed post-test emails, considering type of participatory structure during task performance.

Table 1. Frequency of softeners in pre-test, post-test and delayed post-test emails

Group		Mean	SD	Minimum	Maximum
Control group (n = 16)	Pre-test	0.19	0.40	0	1
	Post-test	0.25	0.45	0	1
	Delayed	0.17	0.40	0	1
Student-students group (n = 16)	Pre-test	0.13	0.34	0	1
	Post-test	1.38	0.72	0	2
	Delayed	0.19	0.48	0	1
Teacher-students group (n = 16)	Pre-test	0.19	0.40	0	1
	Post-test	0.94	0.57	0	2
	Delayed	0.11	0.48	0	1

Note. Post-test was given at week 4, while delayed post-test was given at week 8.

Table 2. Frequency of syntactical mitigators in pre-test, post-test and delayed post-test emails

Group		Mean	SD	Minimum	Maximum
Control group (n = 16)	Pre-test	0.19	0.40	0	1
	Post-test	0.38	0.62	0	2
	Delayed	0.16	0.40	0	1
Student-students group (n = 16)	Pre-test	0.31	0.48	0	1
	Post-test	1.69	0.48	1	2
	Delayed	0.81	0.54	0	2
Teacher-students group (n = 16)	Pre-test	0.31	0.48	0	1
	Post-test	1.06	0.44	0	2
	Delayed	0.69	0.70	0	2

As a general pattern, the internal mitigators (i.e., softeners and the syntactic mitigators *"Could you…"* and *"I was wondering if…"*) hardly ever appeared in the pre-test, but they showed an increase in the two treatment groups in the post-test, followed by a decrease in the delayed post-test.

With regard to the use of softeners, the Kruskal-Wallis test revealed no group difference in the pre-test, $\chi^2(2) = .294$, $p = .863$, found a significant group difference in the post-test, $\chi^2(2) = 18.711$, $p < .001$, but no difference was observed in the delayed post-test, $\chi^2(2) = 8.871$, $p = .110$. In addition, the Mann-Whitney test showed that both treatment groups outperformed the control group at post-test: $z = -3.88$, $p = <.001$ (student-student group) and $z = -3.21$, $p < .001$ (teacher-student group). However, when the two treatment groups (student-students and teacher-students) were compared, there was no significant difference ($z = -1.92$, $p = 0.06$) at the post-test, but there was a significant group difference in the delayed post-test ($z = -4.31$, $p = <.001$) in favour of the student-student group.

In the case of the syntactic structures "*Could you…*", and "*I was wondering if* ", the Kruskal-Wallis test revealed no group difference at pre-test, $\chi^2(2) = .826$, $p = .662$, a significant group difference at immediate post-test, $\chi^2(2) = 24.870$, $p < .001$ and also at delayed post-test $\chi^2(2) = 9.989$, $p < .010$. The Mann-Whitney U test found that both treatment groups (student-students and teacher-students) outperformed the control group at post-test: $z = -4.31$, $p < .001$ (student-student group) and $z = -3.27$, $p < .001$ (teacher-student group) at post-test. The same statistical test revealed a significant difference between the two treatment groups in the post-test ($z = -3.23$, $p < .001$) in favour of the student-student group, but not in the delayed post-test ($z = 070$, $p = .48$).

In relation to the target external mitigators (i.e., preparators and grounders), Table 3 and 4 display descriptive statistics of their frequency of use.

Table 3. Frequency of preparators in pre-test, post-test and delayed post-test emails

Group		Mean	SD	Minimum	Maximum
Control group ($n = 16$)	Pre-test	0.13	0.34	0	1
	Post-test	0.31	0.48	0	1
	Delayed	0.21	0.48	0	1
Student-students group ($n = 16$)	Pre-test	0.25	0.45	0	1
	Post-test	1.44	0.73	0	2
	Delayed	0.81	0.54	0	2
Teacher-students group ($n = 16$)	Pre-test	0.25	0.45	0	1
	Post-test	1.38	0.72	0	2
	Delayed	0.56	0.63	0	2

Table 4. Frequency of grounders in pre-test, post-test and delayed post-test emails

Group		Mean	SD	Minimum	Maximum
Control group ($n = 16$)	Pre-test grounders	1.63	1.03	0	3
	Post-test grounders	1.56	1.03	0	3
	Delayed grounders	1.44	0.96	0	3
Student-students group ($n = 16$)	Pre-test grounders	1.88	0.81	0	3
	Post-test grounders	1.94	0.57	1	3
	Delayed grounders	1.56	0.81	0	3
Teacher-students group ($n = 16$)	Pre-test grounders	1.63	0.81	1	3
	Post-test grounders	2.00	0.63	1	3
	Delayed grounders	1.81	0.83	0	3

Regarding preparators, they almost never appeared at pre-test, showed an increase in the two treatment groups in the post-test, and a decrease in the delayed post-test. The Kruskal-Wallis test revealed no group difference at pre-test, $\chi^2(2) = .989$, $p = .61$, found a significant group difference at immediate post-test, $\chi^2(2) = 18.776$, $p < .001$, and also at delayed post-test, $\chi^2(2) = 6.159$, $p < .05$. In addition, pair-wise comparisons using the Mann-Whitney U test showed that the student-students group outperformed the control group ($z = -2.51$, $p = <.001$), but the teacher-student group did not ($z = -1.16$, $p = .25$) (both at post-test). No differences were found between the two treatments groups in the post-test ($z = -29$, $p = .77$) or at delayed post-test ($z = -1.29$, $p = .20$).

With regard to grounders, the Kruskal-Wallis test revealed no group difference at pre-test, $\chi^2(2) = 1.143$, $p = .565$, no significant group difference at immediate post-test, $\chi^2(2) = 2.619$, $p = .270$ or at delayed post-test, $\chi^2(2) = 1.538$, $p = .464$. The Mann-Whitney U test showed that both treatment groups did not outperform the control group: $z = -1.29$, $p = .201$ (student-student group) and $z = -114$, $p = .26$ (teacher-student group). In addition, the Mann-Whitney U test did not reveal a significant group difference between treatments groups at immediate post-test ($z = -291$, $p = .77$) or at delayed post-test ($z = -948$, $p = .34$).

So to summarise, in response to the first research question, results showed that TSLT is effective for learning how to mitigate high-imposition email requests. A comparison of the frequency of request mitigators in the pre-test, immediate post-test and delayed post-test shows that TSLT makes a difference in short-term learning of internal request mitigators, softeners and the syntactic structures: "*Could you…*" and "*I was wondering if…*" However, considering the results in the delayed post-test, the effects of TSLT seem to have disappeared with regard to softeners. In relation to external mitigators (preparators and grounders), TSLT was effective for developing the knowledge of preparators. However, there was no effect on the knowledge of grounders potentially due to a ceiling effect: students were able to use grounders at pre-test.

In addition, a comparison of the frequency of request mitigators between the treatment groups and the control group shows that, although request mitigators appeared more frequently in the student-students email test tasks, both types of participatory structures equally worked for improving learners' knowledge on how to use the internal mitigators (i.e., softeners and the syntactic mitigators). No difference between the participatory structures was found for external mitigators, grounders and preparators.

The second research question addresses the extent to which the type of task participatory structure made a difference in students' attention to pragmatic features during task performance, and subsequent learning of request mitigators. Students' attention to pragmatic features was operationalized as the occurrence of PREs. Table 5 shows descriptive statistics for frequency of successfully resolved PREs, scores on students' level of engagement during PREs, and scores on students' reported pragmatic knowledge on request mitigators (as appeared in students' self-assessment of post-test emails).

As illustrated in Table 5, over the three email task sessions, the student-students group produced 5.75 PREs on average, while the mean for the teacher-students group was 4.44. With regard to the reported pragmatic knowledge (learned request mitigations) at post-test, the mean for the student-students group was 14.38, while the mean for the teacher-students group was 15.19. Results of t-test revealed no group differences on the frequency of PREs, $t(30) = 1.26$; $p = .21$ or reported pragmatic knowledge of request mitigators, $t(30) = -.268$; $p = .79$. Group difference was observed on the students' level of engagement during PREs, $t(18.3) = 3.97$; $p = < .001$ in favour of the student-students group, which is worth exploring further.

As shown in Table 6, the Kendall's Tau correlation revealed a significant relationship between frequency of PREs and reported pragmatic knowledge of request mitigators at post-test in the student-students group (Kendall's Tau correlation = .71, $p < .001$), while no such relationship was found in the teacher-students group (Kendall's Tau correlation = −.03, $p = .48$). In contrast, in the teacher-students group, students' level of engagement had a significant correlation with students' reported knowledge of request mitigators (Kendall's Tau correlation = .80 $p < .001$), while such correlation was absent in the student-students group (Kendall's Tau correlation = .06, $p = .74$).

Table 6. Kendall's Tau correlation PRE, level of engagement and pragmatic knowledge

Group	Kendall's Tau	Frequency of PREs	Engagement level	Reported pragmatic knowledge
Student-Students	PREs			
	Engagement level	.14		
	Reported pragmatic knowledge	.71[**]	.063	
Teacher-Students	PREs			
	Engagement level	−.02		
	Reported pragmatic knowledge	−.03	.80[**]	

*$p < .05$;
**$p < .001$

These findings suggest that the type of participatory structure (student-students vs. teacher-students) had an impact on the nature of collaborative dialogue during task performance, which, in turn, affected learners' self-reported pragmatic knowledge (request mitigators) at post-test. This was observed when we examined the nature of student-students vs. teacher-students PREs and its impact on drawing participants' attention to pragmatics.

In spite of individual differences, some general patterns did emerge in the data. See the following excerpts for illustration. In the student-students group, although the student who takes on the role of the teacher allocates turns (line 5), other students initiate and provide information and feedback to each other while jointly performing the email construction task during the TSLT sessions. In line 6, S4 makes a suggestion in response to S1 (providing an appropriate grounder to use). Then, S3 questions S4's contribution in line 7, which is not understood by S4 in line 8, but it is indirectly clarified by S5 in line 9. At line 11, S4 incorporates other students' suggestions to provide an excuse, with the agreement of S1 (line 12) and S2 (line 13). The excuse is introduced by S5, in line 14, and accepted by S2 (line 15) and S1 (line 16). In line 16, S1 questions the degree of politeness of the request, S2 (line 17) and S6 (line 18) give reasons why to be polite in making the request to the teacher. These reasons are accepted by S2, who perform the request in line 19, and by S1 in line 20.

Excerpt 4. (S: student).

1	S1:	Ok let's see…
2	S2:	Avui eres el profe? (Today are you the teacher?)
3	S1:	We have to think… have to contact Josep (name of the teacher) for a tutorial
4	S3:	I'll write
5	S1:	David you start. How can we ask and be polite?
6	S4:	I am writing to you because I need to talk to you
7	S3:	An excuse?
8	S4:	What?
9	S5:	Do you want to pass the exam?
10		(all laugh)
11	S4:	Ok we can explain why we need to see him say…
12	S1:	Right we find an excuse
13	S2:	OK an excuse
14	S5:	We had an interview for our internship
15	S2:	Ok and we need an extra tutorial
16	S1:	OK. Is that polite?
17	S2:	Well, he has tutorials
18	S6:	Ya sabes que poca broma (no jokes with him)
19	S2:	Ok would you mind if I go to your office because I have some doubts
20	S1:	OK (they change topic)

This interaction pattern that emerged during student-students TSLT sessions (i.e., PREs jointly constructed among participants) had an impact on learners' building of sociopragmatic and pragmalinguistic information. More specifically, extended metapragmatic discussions about how to make the request reflected learners' engagement in establishing a link between language use and the social factors that influence

language choice. These metapragmatic discussions probably had an impact on self-reported knowledge of request mitigators at post-test, independently of students' degree of engagement in the PREs.

See the post-test example below from Student 6 (S6), who participated in the PRE described above (Excerpt 5). This student, in spite of only participating, in line 18, with a joke in Spanish, seems to benefit from the discussion generated during the PRE. The discussion about finding an excuse or the fact that they were addressing to a professor, which was jointly constructed though interaction, probably had an impact on S6's evaluation of her email appropriateness at post-test.

Excerpt 5.

> Good morning (name of the teacher)
> I'm just writing to you because I wasn't able to attend your tutorials because I was coming back from my holidays in Ireland. Could you please arrange a meeting in order to talk about the course and its main contents, and also about the final assignment?
> Thank you very much (name of the student)
> *Example of S6's evaluation of email appropriateness at post-test*
> "I wanted a tutorial and I knew that I had to find an excuse to be polite (he explains in Spanish how difficult it is for him to talk to this teacher), but I need to ask for a tutorial."

In contrast, as shown in Excerpt 6 from the teacher-students' task performance, it was the teacher who is frequently in charge of classroom interaction. The students simply responded to the teacher's questions. This pattern probably resulted in the lower engagement level of this group compared with the student-students interaction group. The teacher leads the interaction (lines 1 and 6), and he frequently corrects students (line 3). Because of this corrective-oriented approach, students merely incorporate the correct request mitigator supplied by the teacher into their utterance (line 4). Because of this approach, students were probably not able to apply the knowledge of request mitigators to their emails at post-test.

Excerpt 6. (T: teacher; S: student).

> 1   T: Our aim is to ask for an extension of the deadline. How could we say it?
> 2   S1: We need an extension because…
> 3   T: We should be more polite. Could you…?
> 4   S1: Could we have an extension?
> 5   S2: Because we have not had time
> 6   T: OK. Why haven't you had time?

In the teacher-students interaction, students hardly ever paid attention to pragmatics, but when they did, it had an impact on the student that initiated the PRE. This may

explain why in teacher-students' interactions, it was the level of engagement within PREs that correlated with the students' reported pragmatic knowledge at post-test. In Excerpt 7 taken from the teacher-students group, S4 initiates a PRE in line 4, asking whether the use of "*can*" is polite in the given context. The teacher answers the question in line 5, giving information about how social factors, in this case addressing to a professor, may influence the use of "*could*".

Excerpt 7. (T: teacher; S: student).

1	T:	So we have Magdalena (local festivity)
2	S:	(all laugh)
3	T:	We need an extension because Magdalena
4	S4:	Is it polite to use 'can'?
5	T:	No. Better if you use 'could' more polite if you make a request to a professor
6	S4:	Could we have a deadline extension?

In Excerpt 8, we can see that the teacher's information provided during this PRE (Excerpt 7) had impact on pragmatic knowledge of S4, who initiated the PRE. Here S4, in his self-evaluation of email appropriateness, makes explicit reference to the use of "*could*".

Excerpt 8. Dear XXX (name of the teacher)

> I'm writing to you because I have had several health problems for the last weeks (I can provide medical evidence and letters), and I won't be able to submit my papers within the established deadline. Please, could I hand in the papers the following month?
> Looking forward to hearing from you (name of the student)
> *Example of S4's self-evaluation of email appropriateness at post-test*
> "I feel the request is appropriate because before asking for the extension of the deadline, I give an explanation, the reason why I am not able to hand in the papers on the deadline. I know that I need to introduce the request in an indirect way and that she is a teacher, so I use 'could' as formal and polite language. I follow the typical structure of an email."

To sum up, differences were observed on the nature of collaborative dialogue between two types of task participatory structure, which might have affected pragmatic learning outcomes. During teacher-students interaction students hardly engaged in metapragmatic discussions nor initiated PREs, but when they did initiate PREs, it contributed to their explicit knowledge of request mitigators, as measured in their self-evaluation of email appropriateness. In contrast, in student-students interaction, students were engaged in a greater degree of metapragmatic discussion. These more involved discussions, independent from who initiated the PRE, or students' level of engagement in the PRE, probably had an impact on students' knowledge of request mitigators in email assessed at post-test.

Discussion

To date, most instructional studies in pragmatics have been conducted within the framework of Schmidt's (1993, 2001) noticing hypothesis. The present study intended to advance the current practice by assessing the effectiveness of TSLT in teaching pragmatics. With regard to the first research question, which explored whether TSLT has an impact on L2 English learners' production of request mitigators, the frequency analysis of request mitigators in the pre-test, post-test and delayed post-test indicated that TSLT was effective as a means to teach how to mitigate high-imposition email requests. Thus, in line with previous studies on the effect of pragmatic instruction (see Taguchi, 2011, 2015, and Takahashi, 2010, for a summary of findings from instructional pragmatics-based studies), the present study provides evidence on the usefulness of TSLT for learning pragmatics.

Our results also provide evidence that the effects of TSLT may differ according to the type of request mitigators taught. In the case of grounders, no difference was found among the pre-test, post-test and delayed post-test. One possible explanation for the no effect is that learners were probably familiar with the use of these mitigators in their L1, and they transferred their L1 pragmatic knowledge to L2. These findings are in line with those reported in Taguchi and Kim (2016), who claimed that, due to the ceiling effect, there was no instructional effect for grounders. However, TSLT was effective for improving learners' knowledge on softeners, the syntactic structures *"Could you…",* and *"I was wondering if ",* and preparators. Thus, TSLT was beneficial to increase L2 learners' use of internal request mitigators.

Our findings also revealed the immediate and delayed effects of TSLT. However, contrary to Codina's (2008) study reporting delayed effects of pragmatic instruction, in our study, the delayed effects of TSLT were only partially proved. In fact, the effects of TSLT were sustained with regard to the syntactic mitigators and preparators, but not with softeners. One reason for these contradictory findings is that instruction may be influenced by the type of request mitigators taught. Another possible interpretation is that, while TSLT works for short-term gains, its long-term effects may be influenced by other factors such as input exposure and learners' motivation, among others. We acknowledge that we have to be cautious with this interpretation because frequency of input was not addressed in the present study. Still, a tentative hypothesis is that learners were not exposed to softeners as frequently as they were to the syntactic mitigators or preparators, and thus they were not able to maintain their learned knowledge of softeners. This hypothesis needs to be further explored in future investigations.

Moreover, in the present study, both treatment groups (student-students vs. teacher-students) outperformed the control group with regards the use of internal request modifiers (softeners and the syntactic modifiers). These findings suggest that, although the effect of task-based interaction on pragmatics learning can be observed

during peer-to-peer interaction (Taguchi & Kim, 2016), such effect also occurs when learners participate in a whole group classroom interaction. Large class size and the fact that students share an L1 make it difficult to promote a one-to-one interaction in L2 at the university level in Spain. However, the two task implementation options examined in the present study (i.e. the instructor interacting with the whole class and a peer learner taking on the role of the teacher and interacting with the class) were found to be effective for teaching request mitigators in a regular classroom where the study was conducted. These findings are encouraging in that task-based learning can usefully accommodate the existing curricula conditions and produce intended learning.

Findings related to the second research question showed that type of task participatory structure (student-students vs. teacher-students interaction) does not make a difference in the attention learners paid to pragmatics. PREs occurred in both types of participatory structure and positively affected students' knowledge of target pragmatic features, as revealed in their self-evaluations of email appropriateness at post-test. However, differences were observed with regard to the students' level of engagement during PREs. Metapragmatic discussions about how to make a high-imposition request in an email occurred more frequently during student-students interaction than teacher-students interaction. It seems that, in the student-students group, participants actively engaged in metapragmatic discussions by initiating PREs, asking questions, or attempting to resolve PREs, which might have influenced both their knowledge of request mitigators and other learners' knowledge of the importance of request mitigators to write appropriate emails. These findings support previous studies (Takimoto, 2012; Taguchi & Kim, 2016) on the benefits of metapragmatic discussion or collaborative dialogue for pragmatics learning. Takimoto (2012) reported that metapragmatic discussion led to a greater production of request mitigators, while Taguchi and Kim (2016) illustrated how collaborative dialogue, operationalized as PREs, led to a greater production of request head acts.

In contrast to student-students interaction, in teacher-students interaction the teacher was often in charge of classroom interaction. She initiated PREs and provided information about request mitigators, while students merely incorporated the information or responded to the teachers' questions. This corrective approach may explain why not all students in the teacher-students group had the same number of opportunities to participate in interaction. Thus, it is possible that those who had a greater amount of participation may benefit more in terms of building pragmatic knowledge, as indicated in the correlation between the level of students' engagement and their reported pragmatic knowledge at post-test. This hypothesis, stated by Taguchi and Kim (2016), was also supported in the present study. However, further empirical studies are needed to confirm the generalizability of these findings.

Conclusion, limitations and pedagogical implications

Findings from this study revealed positive effects of TSLT on L2 English learners' use of request mitigators, although the effects were not observed for all types of request mitigators. The study also revealed the impact of participatory structure during task-based interaction on learning outcomes of request mitigators. Despite individual differences, in student-students interaction, metapragmatic discussion seemed to direct students' attention towards pragmatics and enhanced their pragmalinguistic and sociopragmatic knowledge of request mitigators. On the other hand, in teacher-students interaction, teachers provided information about pragmatics, and students rarely initiated PREs. However, when they did pay attention to pragmatics, it had an impact on their explicit knowledge of request mitigators. Thus, the present investigation goes beyond previous studies suggesting the benefits of metapragmatic discussion during task completion on pragmatics learning. Our findings suggest that the type of classroom participatory structure may also influence students' level of engagement during PREs, which, in turn, have an impact on pragmatics learning outcomes.

Several considerations are in order when interpreting our findings as evidence of the effectiveness of TSLT on pragmatics learning. First, the small sample size and the rather restricted learning context where we conducted the study do not allow the results to be generalizable to other contexts. In addition, we focused on frequency of use of request mitigators rather than appropriateness of their use. Further studies need to combine frequency analysis with measures of appropriateness. We also need to be cautious in interpreting gains in pragmatic knowledge in this study. We dealt with explicit knowledge (measured in students' self-evaluation of the appropriateness of their emails at post-test), but gains in implicit knowledge may be different from explicit knowledge reported in the present study. Second, this study did not consider learners' individual differences, which might have affected the impact of the participatory structure on learning outcomes. Third, this study examined the nature of PREs qualitatively to supplement quantitative findings, i.e., group difference on the students' level of engagement during PREs. However, because the qualitative analysis was relatively small, future studies need to analyse more data to reveal quality of PREs and its impact on pragmatics learning outcomes. Despite these limitations, the present study fills a research gap by directly testing the efficacy of TSLT in a whole group classroom interaction for pragmatics learning.

Our findings also suggest some pedagogical implications. First, the impact of student-students PREs on building pragmatic knowledge observed in this study highlights the benefits of collaborative dialogue as a tool for pragmatics learning, and suggests the use of collaborative tasks for creating conditions for metapragmatic discussion in a formal classroom setting. Second, since students' level of engagement

during teacher-students interaction had impact on their knowledge of request mitigators, teacher training courses should encourage teachers to think about different classroom interaction patterns and their potential impact on students' pragmatic knowledge.

Acknowledgements

References

Alcón-Soler, E. (2007). Fostering EFL learners' awareness of requesting through explicit and implicit consciousness-raising tasks. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 221–241). Clevedon: Multilingual Matters.

Alcón-Soler, E. (2013). Mitigating e-mail requests in teenagers' first and second language academic cyber-consultation. *Multilingua*, 32(6), 779–799. doi:10.1515/multi-2013-0037

Alegría de la Colina, A., & García Mayo, M. P. (2007). Attention to form across collaborative tasks by low-proficiency learners in an EFL setting. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 91–116). Clevedon: Multilingual Matters.

Bardovi-Harlig, K. (2015). Operationalizing conversation in studies of instructional effect in L2 pragmatics. In E. Alcón & L. Yates (Eds.), *Pragmatic learning across contexts. System*, 48 (special issue) 21–34.

Basterrechea, M., & García Mayo, M. P. (2013). Language-related episodes during collaborative tasks: A comparison of CLIL and EFL learners. In K. McDonough and A. Mackey (Eds.), *Second language interaction in diverse educational contexts* (pp. 25–43). Amsterdam: John Benjamins. doi:10.1075/lllt.34.05ch2

Biesenbach-Lucas, S. (2006). Making requests in E-mail. Do cyber-consultation entail directness? Towards convention in a new medium. In K. Bardovi-Harlig, C. Félix-Brasdefer, & A. S. Omar (Eds.), *Pragmatics language learning* (Vol. 11, pp. 81–107). Honolulu, HI: University of Hawai'i Press.

Biesenbach-Lucas, S. (2007). Student writing emails to faculty: An examination of e-politeness among native and non-native speakers of English. *Language Learning and Technology*, 11(2), 59–81.

Blum-Kulka, S., Kasper, G., & House, J. (1989). *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex.

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

Codina-Espurz, V. (2008). The immediate vs. delayed effect of instruction on mitigators in relation to learner' language proficiency in English. In E. Alcón-Soler (Ed.), *Learning how to request in an instructed language learning context* (pp. 227–256). Bern: Peter Lang.

Economidou-Kogetsidis, M. (2009). Interlanguage request modification: The use of lexical phrasal downgraders and mitigating supportive moves. *Multilingua*, 28(1), 79–112. doi:10.1515/mult.2009.004

Economidou-Kogetsidis, M. (2011). "Please answer me as soon as possible": Pragmatic failure in non-native speakers' e-mail requests to faculty. *Journal of Pragmatics*, 43, 3193–3215. doi:10.1016/j.pragma.2011.06.006

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

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition*. New York, NY: Routledge.

Félix-Brasdefer, C. (2012). E-mail requests to faculty: E-politeness and internal modification. In M. Economidou-Kogetsidis & H. Woodfield (Eds.), *Interlanguage request modification* (p. 87–118). Amsterdam: John Benjamins. doi:10.1075/pbns.217.04fel

García Mayo, M. P. (2014). Collaborative tasks and their potential for grammar instruction in second/foreign language contexts. In A. Benati, M. C. Laval, & M. Arche (Eds.), *The grammar dimension in instructed second language learning: Theory, research and practice* (pp. 82–102). London: Continuum.

García Mayo, M. P. (Ed.) (2015). *Learning language and content through tasks: Exploring the interfaces, System*, 54 (special issue).

García Mayo, M. P., & Azkarai, A. (2016). EFL task-based interaction: Does task modality impact on language-related episodes. In M. Sato & S. Ballinger (Eds.), *Peer interaction and second language learning: Research agenda and pedagogical potential* (pp. 241–266). Amsterdam: John Benjamins.

González-Lloret, M. & Ortega, L. (2014). *Technology-mediated TBLT*. Amsterdam: John Benjamins.

Hartford, B. S., & Bardovi-Harlig, K. (1996). At your earliest convenience: Written student requests to faculty. In L. F. Bouton (Ed.), *Pragmatics and language learning* (pp. 55–69). Urbana-Champaign, IL: University of Illinois, Division of English as an International Language.

Ishihara, N., & Cohen, A. (2010). *Teaching and learning pragmatics. Where language and culture meet*. London: Longman.

Jeon, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development: A meta-analysis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 165–211). Amsterdam: John Benjamins. doi:10.1075/lllt.13.10jeo

Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of second language vocabulary. *Modern Language Journal*, 92, 114–130. doi:10.1111/j.1540-4781.2008.00690.x

Kim, Y. (2015). The role of tasks as vehicles for language learning in classroom interaction. In N. Markee (Ed.), *The handbook of classroom discourse and interaction* (pp. 163–181). Oxford: Wiley-Blackwell.

Kubota, M. (1995). Teachability of conversational implicatures to Japanese EFL learners. *IRLT Bulletin*, 9, 35–67.

Li, S. (2012) The effect of input-based practice on pragmatic development in L2 Chinese. *Language Learning*, 62, 403–438. doi:10.1111/j.1467-9922.2011.00629.x

Lantolf, J. (Ed.). (2000). *Sociocultural theory and second language learning*. Oxford: Oxford University Press.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Oxford: Wiley-Blackwell.

Nasaji, H., & Tian, J. (2010). Collaborative and individual output task and their effects on learning English phrasal verbs. *Language Teaching Research*, 14, 164–187.

Nguyen, T. T. M. (2013). Instructional effects on the acquisition of modifiers in constructive criticisms by EFL learners. *Language awareness*, 22, 76–94.
doi:10.1080/09658416.2012.658810

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50–72.

Prabhu, N. S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.

Robinson, P. (2001). Task complexity, task difficulty and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. doi:10.1093/applin/22.1.27

Rose, K. R. (2005). On the effects of instruction in second language pragmatics. *System*, 33, 385–399. doi:10.1016/j.system.2005.06.003

Rose, K. R. & Kasper, G. (Eds.) (2001). *Pragmatics in language teaching*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524797

Salazar, P. (2003). Pragmatic instruction in the EFL context. In A. Martínez-Flor, E. Usó-Juan, & A. Fernández-Guerra (Eds.), *Pragmatic competence and foreign language teaching*. Castelló: Servei de Publicacions.

Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 21–42). Oxford: Oxford University Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–33). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524780.003

Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*, 19, 268–286.
doi:10.1017/S0267190500003585

Skehan, P. (Ed.). (2014). *Processing perspectives on task performance*. Amsterdam: John Benjamins. doi:10.1075/tblt.5

Storch, N. (2007). Investigating the merits of pair work on text editing tasks in ESL classes. *Language Teaching Research*, 11, 143–159. doi:10.1177/1362168807074600

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16(3), 370–391.
doi:10.1093/applin/16.3.371

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks. Second language learning, teaching and testing* (pp. 99–117). London: Longman.

Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learner's response to reformulation. *International Journal of Educational Research*, 37, 285–304.
doi:10.1016/S0883-0355(03)00006-5

Taguchi, N. (2011). Teaching pragmatics: Trends and issues. *Annual Review of Applied Linguistics*, 31, 289–310. doi:10.1017/S0267190511000018

Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. State-of-the-art article. *Language Teaching*, 48, 1–50.
doi:10.1017/S0261444814000263

Taguchi, N., & Kim, Y. (2014). Collaborative dialogue in learning pragmatics: Pragmatics-related episodes as an opportunity for learning request-making. *Applied Linguistics*, 1–23.

Taguchi, N., & Sykes, J. (2013). *Technology in interlanguage pragmatics research and teaching*. Amsterdam: John Benjamins. doi:10.1075/lllt.36

Takahashi, S. (2010). Assessing learnability in second language pragmatics. In A. Trosborg (Ed.), *Handbook of pragmatics VII* (pp. 391–421). Berlin: Mouton de Gruyter.

Takimoto, M. (2012). Metapragmatic discussion in interlanguage pragmatics. *Journal of Pragmatics*, 44, 1240–1253. doi:10.1016/j.pragma.2012.05.007

University of Cambridge Local Examinations Syndicate (UCLES). (2001). CAMBRIDGE IELTS 2 – *Examination Papers from the University of Cambridge Local Examinations Syndicate*. Cambridge: Cambridge University Press.

Van den Braden, K., Bygate, M., & Norris, J. (2009). *Task-based language teaching: A reader*. Amsterdam: John Benjamins.

Vellenga, H. (2004). Learning pragmatics from ESL & EFL textbooks: How likely? *TESOL-EJ*, 8(2), 1–18.

Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N., & North, B. (2009). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Vygostky, L. (1978). *Mind in society*. Cambridge, MA: The MIT Press.

Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford: Oxford University Press.

Task complexity effects on interaction during a collaborative persuasive writing task

A conversation analytic perspective

María Pía Gomez-Laich & Naoko Taguchi
Carnegie Mellon University

This study examined whether task complexity (Robinson, 2011a, b), induced through reasoning demands, affects L2 learners' interaction patterns during a collaborative writing task that involved the pragmatic act of persuasion. We analyzed interaction of two pairs of students when they co-constructed a persuasive essay in English based on a prompt. One pair completed a 'simple' task, which provided explicit information about the arguments, macro-structure of the essay, and linguistic devices to use in a persuasive essay, while the other pair completed a 'complex' task in which such information was withheld, and thus they needed to use reasoning skills to figure out the structure of the persuasive essay. Using a conversation analysis-inspired approach, we examined how students co-constructed an essay. Results revealed differences between pairs completing a complex and simple task in terms of (1) pre-writing negotiation over the essay's structure and (2) during-writing negotiation over sources of trouble. The complex task condition prompted participants to use more reasoning processes to accomplish the task goal, as shown in more extended negotiation sequences and turn taking, frequent pauses, and hesitant ways of speaking (e.g., use of rising intonation and epistemic markers).

Introduction

The field of pragmatics studies a variety of units, including speech acts (e.g., requesting, apologizing, persuading), politeness expressions in interaction, linguistic characteristics of speech registers, conversation mechanisms, the control of presupposition, and the creation of coherent discourse (Cutting, 2015). These linguistic, discoursal and interactional units serve as venues to study appropriate and effective communicative acts in spoken or written communication. Among these units, the present study focuses on the conventions involved in the act of persuasion. The act of persuasion involves a clear connection among forms, communicative goals, and effects of the act on the other, and thus constitutes an appropriate object of pragmatics analysis.

Users of any given language have persuasive strategies at their disposal (i.e., a range of options from which a speaker can choose in situations where persuasion is necessary) (Johnstone, 1989). Some of these strategies include using logic, telling stories, employing displays of emotion, threats or bribes, or simply repeating what we want until the interlocutor(s) gives in. Successful persuasion in writing, however, depends on having explicit knowledge of the linguistic and rhetorical features of this genre (Miller, Mitchell, & Pessoa, 2014). This knowledge is critical because convincing an audience of the veracity of one's arguments involves both rational exposition and the manipulation of rhetorical and language features (Hyland, 1998). People need to know the culturally conventionalized linguistic forms and text structures to present their position in a persuasive manner. People also need to be mindful about the projected effects of their message on the audience.

This study investigated L2 English learners' ability to produce a persuasive essay as a genre in a college composition program. The study used a task-based approach, in which a task was conceptualized as a goal-oriented meaningful activity (Long, 2015). Our task was oriented to the goals of persuading the reader to accept the writers' opinion on a specific topic using rhetorical moves and linguistic forms appropriately and effectively. To achieve this goal, participants interacted with each other and collaboratively constructed a persuasive essay. We used collaborative writing as a task procedure to promote shared decision-making and responsibility among L2 learners as they jointly constructed an essay. A coordinated effort among learners to complete a task is considered to generate more negotiation and interaction around the linguistic and discourse features targeted in the task, which leads to greater noticing and learning of the features.

In order to design a collaborative writing task, we adopted Robinson's (2011a, b) Cognition Hypothesis. The hypothesis contends that a cognitively complex task, due to its increased communicative demands, leads to greater interaction and negotiation of task-induced linguistic features. Adopting this hypothesis, we investigated how task complexity, manipulated along the [±reasoning demands] resource-directing variable, affected participants' interaction patterns when performing a collaborative writing task that involved writing a persuasive essay. We used a conversation analysis-inspired approach to explore any similarities and differences during participants' collaborative task performance.

Background

L2 pragmatics and genre

Teaching pragmatics involves a wide range of learning objects such as "conversational structure, conversational implicature, conversational management, discourse organization, and socio-linguistic aspects of language use" (Bardovi-Harlig &

Mahan-Taylor, 2003, p. 37). However, existing instructional studies have predominantly focused on teaching isolated speech act strategies and other utterance-level forms along with the sociocultural norms of their use (e.g., address terms, discourse markers, response tokens, hedging) (Taguchi, 2015). Studies that explore the teaching of written discourse and genres are rare. This is a serious neglect because pragmatics attends to the effects of our language use on the interlocutor (Crystal, 1997), and such effects are typically achieved not only by using utterance-level forms, but also by using discourse-level conventions. In fact, bridging pragmatics and genre-related research, Unger (2006) claimed that, when interpreting an utterance, we draw not only on the properties of a particular utterance (e.g., linguistic forms) but also on the properties of the type of text or discourse that contains the utterances. Utterance-level and discourse-level properties together impact the audience's understanding of a text, contributing to their judgments of discourse appropriateness and rhetorical effectiveness. For example, when writing a message to persuade college students to volunteer abroad, linguistic forms such as "You should", "I recommend", and "We suggest" can serve as useful resources. However, these forms alone do not lead to effective persuasion. Equally important is the knowledge of how to structure the message at the discourse-level by using rhetorical moves of persuasion (e.g., stating the main argument up front, presenting supporting evidence in a logical manner, emphasizing the key information throughout). Hence, the ability to control a written genre and produce intended effects on the audience deserves more attention as part of pragmatic competence.

Genres are modes of speaking or writing that involve socially accepted conventions that people learn to adopt in a particular community (Fairclough, 2003). When studying genres, a number of scholars have focused on communicative purposes for achieving socially recognized goals (Martin, 1992; Swales, 1990). Communicative purposes range from a localized goal such as telling a personal anecdote to amuse family members, to a broad goal such as giving a political speech to persuade a general audience. Depending on which communicative purposes they pursue, speakers and writers need to use specific linguistic and rhetorical conventions in a way that suits their goals, settings, and audience.

The concept of communicative purposes contrasts with the concept of illocutionary force (i.e., a speaker's intention in producing an utterance) specified in speech act theory (Austin, 1962; Searle, 1969). Speech act theory posits that illocutionary force is solely in the speaker's mind and can be expressed in a single utterance. Thus, the theory ignores the listener's contributions to meaning construction or the contribution of the cumulative effects of multiple utterances to meaning. Since Austin and Searle's decade, there has been a consistent development towards seeing a pragmatic act as interactively negotiated among speakers. With the surge of interactional pragmatics (Linell, 1998; Clark, 1996), a speech act is currently viewed as a dynamic, co-constructed entity in discourse.

The trend of interactional pragmatics has started to spread to L2 pragmatics research. For example, recent studies have analyzed how L2 learners co-construct a speech act with their interlocutors sequentially turn-by-turn (e.g., Dippold, 2011; Huth, 2006). However, the traditional, utterance-level perspective of speech acts is still predominant because most existing studies have focused on utterance-level, lexico-grammatical features in speech acts (e.g., syntactic and lexical mitigations, semantic strategies) (for a review, see Taguchi & Roever, 2017). These studies have overlooked how higher-level discourse structures, such as rhetorical devices, which collectively configure discourse, serve to achieve a communicative purpose.

Moreover, taking the position that illocutionary force is in the speaker's mind, most studies have ignored the projected effects of texts on the listener's (or reader's) experience. Understanding illocutionary force depends on the interaction between the speaker/writer and listener/reader in a local context. For example, whether the utterance "We are having a party this Friday" takes the force of an invitation or promise depends on the topic of a conversation, context of talk, and the listener's state at the time. Hence, except for institutionalized speech acts, illocution is rarely understood via isolated utterances removed from the listener/reader's experience. As such, the analysis of illocution needs to go beyond an utterance and speaker, extending to the larger unit of a text that involves interaction between the text and the listener's (or reader's) background, expectations, and preferences.

A few studies have investigated L2 pragmatic competence in written discourse or genres (e.g., Cohen & Tarone, 1994; Hyland, 1990; Ifantidou, 2011; Zhao & Kaufer, 2013). Some studies focused on the structure of the argumentative essay. For example, Hyland (1990) conducted a move analysis of 65 argumentative essays written by high school students in Papua New Guinea, as well as samples of journalistic material prepared for the British and American press. Based on the analysis, Hyland concluded that an argumentative essay has three main stages: (1) a thesis that contains an attention grabber, background information, and the writer's position; (2) an argument that contains a claim and support for that claim; (3) a conclusion that reaffirms the writer's position.

Other studies focused on the linguistic forms involved in the structure of written discourse and genres. Cohen and Tarone (1994) identified semantic strategies used by L2 English learners when writing an opinion essay and compared those to native English writers' strategies. Although the authors identified a variety of strategies, such as expressing change of opinion ("I would like to retract my previous position that…and state…"), their study still focused on isolated, utterance-level strategies and did not address rhetorical moves that characterize opinion writing. Zhao and Kaufer's (2013) study took a discourse-based approach. They used a computer program to identify genre-specific patterns to assess L2 English learners' essays in three genres: informational, descriptive, and narrative writing. The study showed that the learners were able

to produce texts that met the specifications of each genre. Ifantidou's (2011) study attempted genre-based training on metapragmatic awareness by showing L2 English learners the link between linguistic indexes (e.g., passive structures, personal pronouns) and pragmatic effects retrieved by readers (e.g., playful, manipulative). Results revealed greater training effects on description than summary or synthesis writing; however, the effects were found for higher-level but not for lower-level learners.

Focusing on the genre of persuasion, this study intends to add to the small body of literature on teaching pragmatics in discourse. Persuasive writing has a pragmatics focus because it intends to achieve the communicative purpose of convincing a reader to adopt a particular viewpoint or action by using logic and reasoning to demonstrate why one idea is more credible than another. Developing an argument requires students to analyze and evaluate content knowledge, to distinguish relevant from irrelevant information, to take a position by expressing a 'voice' and a 'stance' (Street, 2009), and to present their position in a coherent manner so that the "development of the position is reflected in a logical text structure" (Wingate, 2012, p. 146). To write an effective persuasive essay, it is necessary to present a "sequence of interlinked claims and reasons that, between them, establish content and force of the position for which a particular speaker is arguing" (Wingate, 2012, p. 146). In other words, persuasive writing requires students to embrace a particular point of view, to justify their position, to consider alternative positions, to rebut those opposing positions, to try to persuade the audience to adopt their position, and to present that position in a coherent manner. To this end, specific rhetorical moves and linguistic devices that can signal those moves are critical resources for writing an effective persuasive essay. Our study focuses on these resources.

Using the genre of persuasive writing as the instructional target, this study examines task complexity effects based on Robinson's (2011a, b) Cognition Hypothesis. We adopt the Cognition Hypothesis' concept of task complexity in this study because of the hypothesis' claim that increased task complexity leads to a greater amount of interaction among L2 learners, resulting in a greater degree of interaction-driven learning opportunities.

The Cognition Hypothesis: Task complexity and interaction-driven learning opportunities

The Cognition Hypothesis suggests that cognitively complex tasks, due to their increased communicative demands, can (a) prompt increased attention to L2 form-meaning mappings by directing attention to task-relevant linguistic elements, (b) result in greater interaction and negotiation of meaning to resolve the communicative challenge they pose, relative to simple tasks, and (c) facilitate L2 development. Robinson (2003a, 2003b) proposed a taxonomy of task characteristics known as the Triadic

Componential Framework (TCF). The TCF describes task and learner characteristics, including task complexity, task condition, and task difficulty. This study focuses on task complexity that involves cognitive demands of a task.

Task complexity can be manipulated along two dimensions: *resource-dispersing* (or *resource-depleting*) and *resource-directing* (Robinson, 2003a, 2003b). Resource-dispersing dimensions make performative and procedural demands on learners' cognition. These dimensions pose increased demands on learners' attentional and memory resources, but do not direct their attention to new linguistic forms. On the contrary, resource-directing variables of task complexity make greater demands on learners' attentional and memory capacity in a way that affects the allocation of cognitive resources to specific aspects of the L2.

The Cognition Hypothesis contends that increasing task complexity leads to increased interaction-driven learning opportunities. Those learning opportunities are often operationalized as language-related episodes (LREs), which refer to instances in which learners talk about, question, and/or self-or-other correct language use (e.g., grammar, lexis) (Swain & Lapkin, 1998). In addition to LREs, previous studies analyzed a variety of interactional features to examine the degree of learning opportunities such as clarification requests, confirmation checks, and uptake of recasts when learners complete a cognitively complex interactive task (e.g., Révész, 2011).

Several studies have revealed that increasing task complexity along resource-directing variables leads to more occurrences of interaction-driven language learning opportunities (Kim, 2012; Kim & Taguchi, 2015; Robinson, 2001; Robinson, 2007). Robinson (2001) manipulated the [± few elements] dimension, whereas Robinson (2007), Kim (2012) and Kim and Taguchi (2015) manipulated the [±reasoning demands] resource-directing variable to examine if task complexity has an impact on the amount of interaction-driven learning opportunities. These studies presented evidence that increased task complexity resulted in a greater need to interact and negotiate meaning, leading to a greater number of LREs.

Other studies, however, have produced inconsistent findings regarding the effect of task complexity on the amount of interaction (Gilabert, Barón, & Llanes, 2009; Gilabert & Barón, 2013; Kim, 2009; Nuevo, Adams, & Ross-Feldman, 2011; Révész, 2011). These studies suggest that it is not task complexity alone that affects the amount of interaction; there might be other factors that mediate the link between task complexity and interaction. Some potential factors involve task type, outcome measures, and proficiency. For example, Kim (2009) revealed that proficiency and task type mediated the effect of task complexity on the number of LREs. Low proficiency learners produced significantly more LREs in a simple picture narration task, but the opposite pattern was observed for a picture difference task. High proficiency learners produced more LREs in a complex picture narration task, but no task difference was found in picture description tasks. Révész (2011), on the other hand, suggests that the effect

of task complexity may differ depending on the aspects of interaction analyzed. In her study, ESL learners completed a simple and complex version of an argumentative task, which involved evaluating competing requests for funding city projects. The two versions (simple vs. complex) differed along the [± reasoning demands] and [± few elements] dimensions. Task complexity only had a significant effect on one aspect of LREs, namely amount of metalinguistic talk. Other interactional features (e.g., confirmation checks, recasts) did not reach significance, suggesting a mediating effect of outcome measures on the occurrence of interaction-driven learning opportunities.

A trend in the current research on task complexity and interaction is to focus on numerical data such as frequency of LREs and other communication strategies (e.g., clarification requests, confirmation checks) as an indicator of interaction (e.g., Kim, 2012; Révész, 2011; Robinson, 2001, 2007). Very few studies have taken a qualitative approach by analyzing learners' interaction turn-by-turn to reveal how increased task demands affect patterns of interaction. As a result, we do not know whether task complexity affects turn-taking patterns, conversational sequences, prosodic features (e.g., intonation, stress), and other non-verbal aspects (e.g., gaze, gestures, facial expressions). These interactional features are particularly important to analyze in a collaborative task in which participants interact with each other to achieve a shared goal, such as the activity of co-constructing a persuasive essay examined in this study. Participants' mutual orientation toward a task goal might become visible in the way they take turns, sequence negotiations, deliver their speech, and gaze at each other. Hence, it is legitimate to ask whether and how these interactional features differ depending on changing task demands. To expand the scope of analysis from LREs to quality of interaction, this study adopts a conversation-analytic perspective to examine task complexity effects on interaction. We attempt this using a collaborative writing task that requires participants to jointly produce a persuasive essay.

Collaborative writing

Collaborative writing has been defined as "an activity where there is a shared and negotiated decision-making process and a shared responsibility for the production of a single text" (Storch, 2013, p. 3). In other words, collaborative writing involves participants working together and interacting throughout all stages of the writing process, contributing to the planning, generation of ideas, and deliberating about text structure, editing and revision. Collaborative writing differs from cooperative writing, which involves the division of labor between individuals during the process of completing a writing task.

Collaborative writing has three distinguishing features: (1) substantive interaction in all stages of the writing process; (2) shared decision-making power over and responsibility for the text produced; and (3) the production of a single written

document (Ede & Lunsford, 1990). Collaborative writing involves individuals in a coordinated effort to complete a task together (Storch, 2013), and the product of a collaborative writing task is a jointly produced and shared text that cannot be reduced to the separate input of individuals (Stahl, 2006).

There are several benefits of collaborative writing. For example, L2 writers have access to the ideas and linguistic resources of other L2 writers. This enables them to draw on this larger pool of knowledge rather than on their own knowledge sources. Collaborative writing also provides learners with opportunities to deliberate over linguistic elements while completing meaning-focused tasks. When learners compose a text and become aware of gaps in their knowledge or are uncertain about how to best express an idea, their attention turns to language choice and form. Research indeed has shown that L2 writers devote much attention and time to deliberations about language choice and form. Storch and Wigglesworth (2007), for example, analyzed transcripts of collaborative writing activities and found that over 30% of the talk concerned deliberations about language choice. Loewen and Basturkmen (2005) investigated the extent to which ESL students discussed language use (grammar, discourse) during a small group-writing task. The study particularly examined the linguistic focus of LREs during the task. They found that students paid considerable attention to forms and discourse.

Collaborative writing tasks have other benefits. They provide learners with opportunities to use the target language for a range of functions. These functions include providing negative and positive feedback (e.g., recasts, correction, praise), seeking confirmation, and explaining forms. Research has shown that in such tasks learners provide suggestions and counter suggestions and extend on each other's suggestions to compose complex ideas (Storch, 2013). Learners can build on each other's suggestions, collectively scaffolding their performance to a level they could not have attained if they had worked on their own. Thus, collaborative writing may be more conducive to language learning than solitary writing. Collaborative writing also provides learners with opportunities to consolidate existing linguistic knowledge and to co-construct new knowledge that can be subsequently internalized and used later in individual writing production. Finally, the deliberation that takes place when composing a text collaboratively provides learners with a meaningful and genuine need to communicate and to use the L2 for a variety of functions that they may rarely practice (e.g., explaining, providing feedback, expressing disagreement).

Considering these benefits, this study employed a collaborative writing task with task complexity as a built-in component. We focus on task complexity in this study because increased task complexity is considered to lead to a greater number of interaction-driven learning opportunities (Robinson, 2011a, b). Using a conversation-analytic inspired approach, we explored whether and how patterns of interaction emerging from a collaborative writing task qualitatively differed between L2 English

learners who completed cognitively complex and simple tasks. The tasks used in this study had a clear pragmatics focus because the task goal was to persuade readers to take a certain point of view. Participants jointly achieved the illocutionary force of persuasion through the process of co-constructing a persuasive essay.

Purpose of the study

This study was part of a larger study that examined the effects of task complexity on the quantity and quality of task-based interaction of L2 English learners, as well as on their learning of genre-specific conventions of writing (i.e., use of rhetorical moves and linguistic forms for realizing the moves) (Gomez-Laich, 2017). In this paper, we report part of the findings from the larger study, focusing on the quality of task-based interactions affected by task complexity based on the analysis of two pairs of L2 learners.

The task involved co-constructing an essay with the goal of persuading the reader to accept the writer's point of view. A range of linguistic and discourse-level resources (e.g., linguistic forms, rhetorical devices) were provided in the task so learners could draw on those resources and effectively accomplish the pragmatic act of persuasion. However, the resources provided were different between the cognitively simple and complex task developed in this study (see the methods section). We examined how two tasks that differ in their level of task complexity affect learners' interaction patterns during a collaborative writing task involving the pragmatic act of persuasion. The following research question guided the study: Does increased task complexity manipulated along the dimension of reasoning demands influence L2 English learners' interaction patterns while constructing a persuasive essay collaboratively?

Methods

Participants

Participants in the larger study were 62 international students recruited from different sections of a freshman composition class in a university in the U.S.A. (39 females and 23 males; mean age of 18.3, ranging from 18 and 20). The class is designed for L2 English students who are highly proficient in English but still need instruction on the rhetorical and linguistic demands for academic writing. The participants' average TOEFL score was 109.88 (range: 103–116). The majority of them were from China (n = 47), 10 from South Korea, 1 from Finland, 1 from Guatemala, 1 from Nigeria, 1 from Ukraine, and 1 from Vietnam.

In order to examine students' interactional patterns in depth, we analyzed data from two pairs of participants, one high-performing pair from the simple and the complex task conditions. A high performing pair was defined as a pair: (a) who used

most target linguistic resources (rhetorical moves and linguistic forms to realize the moves) in their co-constructed persuasive essay; (b) who clearly improved their persuasive writing from pre to immediate posttest (test results not reported in this paper); and (c) who collaboratively attended to target rhetorical and linguistic resources during interaction (see below for the definitions of rhetorical moves and linguistic forms). We selected the high performing pairs because the focus of the analysis was task condition only (simple vs. complex), not on outcomes of task performance (high vs. low performing).

Instructional targets: Rhetorical moves and linguistic forms in a persuasive essay

This study used a task to develop students' ability to write persuasive essays. We conceptualized persuasive writing as consisting of two components: rhetorical moves and linguistic forms for realizing the moves. A persuasive essay involves five rhetorical moves: (1) introducing the general topic and orienting readers to a question at issue, (2) acknowledging the opponents' arguments; (3) refuting the opponents' arguments; (4) proposing a specific argument and offering evidence to support the argument; (5) providing a general summary addressing opposing viewpoints and explaining why readers should align with the writer's position (Graff & Birkenstein, 2010). To realize each of these moves, a variety of linguistic resources are required, as shown in Table 1 (Graff & Birkenstein, 2010). These linguistic resources, along with the five rhetorical moves, were the targets in this study. A task was designed to promote students' negotiation around these target linguistic forms and rhetorical moves (see the next section).

Table 1. Linguistic resources for realizing rhetorical moves in a persuasive essay

Rhetorical moves	Linguistic resources
To acknowledge and accommodate opposing viewpoints	Opponents of…maintain that *[opposing viewpoint]* Those who are against…may assert that *[opposing viewpoint]* Many people believe that *[opposing viewpoint]* It is often thought that *[opposing viewpoint]*
To refute opposing viewpoints	This view is mistaken because it overlooks… This view fails to acknowledge that… The claim that *[opposing viewpoint]* rests upon the questionable assumption that… While this position is popular, it is not supported by the facts
To propose a specific argument and offer evidence to support the argument	The main advantage/disadvantage of … An additional advantage/disadvantage of, One argument in favor of/against… Adding to my argument, I would point out that…

Conceptualization of task in this study

This study was conducted as part of the existing curriculum in a U.S. university (freshman composition). We used a collaborative writing task to assist students' development of persuasive writing abilities. We conceptualized a task as a shared goal-oriented activity where students interacted with each other and jointly constructed an essay. Hence, the illocutionary force of persuasion, the goal of the task, was jointly achieved among students. As Storch (2013) notes, mutual engagement and coordinated effort among students is expected while composing an essay together. Negotiation, feedback, and scaffolding occurring during task-based interaction are considered to direct students' attention to genre-specific rhetorical moves and linguistic resources, consequently promoting noticing and learning of these moves and resources, leading to better writing abilities. Hence, we designed a task as a vehicle to promote interaction around genre-specific rhetorical moves and linguistic resources.

Operationalization of task complexity

Students were randomly assigned to either a simple or complex task condition. In both conditions, participants formed a pair and wrote a persuasive essay collaboratively. See the task prompt below. In order to make the task realistic for students, the first author read the minutes from the University's Undergraduate Student Senate's weekly meetings. From the minutes, she identified a topic that was relevant to students' everyday campus lives.

Your task is to write a persuasive essay on the following topic:

> Some students believe that University X should provide more ethnic food options (e.g., Chinese, Italian, Greek, Mexican, Japanese, Korean, etc.). On the contrary, others argue that University X already provides enough food options and no more food varieties are needed. What is your position on the subject?

Although the task goal was the same in the two conditions (producing a persuasive essay in pairs through interaction), the degree of cognitive complexity was different between the simple and complex task. Following Robinson (2011a, b), task complexity was manipulated along the [±reasoning demands] resource-directing variable by controlling the amount of assistance given to the students. The simple version of the collaborative writing task was designed to engage students in fewer cognitive processes. The simple task provided students with (a) a T-chart with ready-made ideas to include in their essays, (b) a ready-made outline that helped them organize their essays rhetorically, and (c) a list of linguistic resources for realizing rhetorical moves of persuasive writing. This task format was considered cognitively less demanding since students did not have to think about ideas, rhetorical moves, or linguistic resources to use in their essays.

In contrast, the complex version of the collaborative writing task did not provide much assistance. We provided students with (a) a Venn diagram with ideas in a randomized order, (b) an outline in which the rhetorical moves were not in order, and (c) a list of linguistic devices for a variety of genres, not specific to persuasive writing. The complex task was expected to induce greater reasoning demands because students needed to re-arrange the ideas in order to show the connection between their viewpoint and the opposing viewpoint. They were also required to discuss how the outline should be re-arranged so it followed the Point-by-Point refutation pattern of organization. They also needed to decide which linguistic devices (out of a randomized list of devices) to use to realize the organization of persuasive writing.

The task complexity manipulation in this study was similar to that of Révész et al.'s (2016) study, which operationalized task complexity as provision or non-provision of ideas to be included in an argumentative essay. The simple task condition provided L2 English learners with content ideas, while the complex task condition withheld the content support. Based on the results of participants' survey responses and stimulated recall, the authors concluded that the complex task was perceived to be more cognitively demanding due to increased pressure on planning and translation processes.

Similarly, in this study, participant surveys, interviews, and teachers' judgments provided evidence that the complex and simple writing tasks were indeed different on the level of cognitive complexity, lending support to the validity of the researchers' task manipulations (see Gomez-Laich, 2017). Students reported that the complex task involved more mental effort and concentration than the simple task. They also judged the complex task as taking significantly more time than the time it actually took them to perform it. Teachers also judged the complex task as involving greater mental effort, concentration, difficulty, and time pressure.

Data collection procedures

The study was conducted in a laboratory format outside of class. Students first completed a persuasive essay individually as pre-test. Then, they were paired randomly[1] and completed a collaborative writing task (writing a persuasive essay in pairs) in their respective conditions (simple or complex). Students received three minutes for individual planning before they started the task. Students' interactions while completing the collaborative writing task were video-recorded. The program *ScreenFlow* was

---

1.   Due to students' time constraints and class schedules, not all students were able to work with students they were familiar with.

used to capture audio and screen movements (i.e., students' essay writing) as a video. Students took approximately 50 minutes to produce a collaborative essay. On the following day, students completed a persuasive essay individually as posttest, and two weeks later a delayed posttest.

Data analysis procedures

Students' task-based interaction data were transcribed and analyzed qualitatively to examine whether patterns of interaction differed between the two task conditions (see Appendix A for transcription conventions). Adopting a conversation-analytic inspired approach, a multimodal analysis of students' verbal interaction and para-linguistic features (e.g., facial expressions, gesture, gaze, tone, and pauses) was conducted. Data were segmented into turns and analyzed for the moment-to-moment sequential organization of turns (how one turn responds to and/or projects a relevant next action). Turns were analyzed closely to reveal whether interaction patterns differed between pairs of students who completed a complex and simple task (e.g., differences in terms of how the pairs oriented to the task, negotiated epistemic stance, planned the writing of their essay, distributed labor when co-constructing their essay).

The first author watched the video recordings of the focal pairs' task-based interaction repeatedly and closely, along with transcriptions, in order to determine what might be of interest, and went through a recursive noticing-transcription-analysis process. During this process, several differences between the participants in the simple and complex task condition emerged. The following section reports on two areas of differences: (1) pre-writing negotiation over the essay's outline, and (2) during-writing negotiation around sources of trouble and actions to take.

Results

Pre-writing negotiation over the essay's outline

Pre-writing negotiation refers to the negotiation that participants engaged in before they actually started writing their persuasive essay. To illustrate this, we will first discuss Dyad 1 from the complex task condition, which illustrates pre-writing interaction between two Chinese female students (P28 and P45). Before this segment, these two students had spent some time deciding on their position on the topic (whether or not the school should offer more ethnic food options) and opposing arguments. Once they had made these decisions, they started re-arranging the outline of the essay so that they could start writing, as illustrated in Excerpt 1.

The quality of learning in both high school and university depends on our personal choices. To learn effectively or not is a reflection of our own understanding towards learning. However, compared to high school, university provides a less constrained schedule, and it requires a stronger willingness to learn and be responsible to ourselves.

Unlike high school, university has a more flexible time schedule where students can arrange their own time table. By doing this, university provide the freedom for students to really focus on things that interest them. For example, there are interdisciplinary courses offered in university. These courses are designed for students that have not only one interest of learning but multiples. By providing a wide range of courses, students in science major can also study courses like drawing and dancing. Thus, students can really follow their dreams by gaining sufficient skills. In contrast, students in high school are always asked to have certain classes. There is barely a chance for students to choose electives. Moreover, because of the pure pressure, students will tend to choose academic classes. Therefore, a major difference between high school and university is the flexibility of schedule.

Another major difference between high school and university is students' sense of responsibility. In university there is no mandatory extracurricular activities and most of them are run by students, so it depends on students' personal willingness. If students are responsible for themselves, they would join clubs and associations that will enrich their knowledge and improve skills outside of class. In contrast, most activities in high school are arranged by teachers or parents. Students have not yet grown the sense of responsibility for their futu addition, students are responsible for their own grades, which means that if they don't ha homework or attend classes, there's no punishment expect for poor GPA. Whereas in high school, teachers will talk to students and their parents and give detentions. Although stud finish homework in the end, it's not motivated by themselves. Most importantly, we are a grown-ups in college, we have to be responsible for our behaviors by law. Once we break law, nobody can take responsibility for us. While in high school, there's greater tolerance our misbehaviors.

To summarize, there are two major differences between high school and college— flexibility of schedule and our sense of responsibility. Although we have more freedom in college, we are also expected to take responsibility for ourselves. Because it s
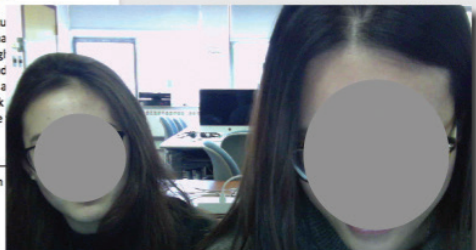
00:57:54 23

Figure 1. Sample of *ScreenFlow*'s interface

Excerpt 1.

91	P28:	I think we have to rearrange the:se °four parts° ((pointing at the task sheet))
92	P45:	mm-hm (4.0) ((looking at the task sheet))
93	P28:	these two seems like a body paragraph ((pointing at the task sheet))
94	P45:	↑OH:::↓
95	P28:	did you notice <u>that</u>, ((laughs and turns to look at P45))
96	P45:	oh I didn't notice tha:t ((smiles))
97	P28:	yeah so we have to rearrange it first before we write
98	P45:	the <u>last</u> one is the >opening paragraph<=
99	P28:	=yeah it's the introduction ((writes on the task sheet)) (5.0)
100		then the body paragraphs,
101	P45:	u::m there aren't any differences between those two <u>are</u> they, ((looking at the task sheet)) (2.0)
102		yeah >I think so<
103	P28:	where ((turns to look at P45's tasks sheet))
104	P45:	these two ((points at the task sheet))
105	P28:	oh there are no differences
106	P45:	ok
107		body one: ((smiles))
108		body two: ((smiles))
109	P28:	and the:n there i:s [conclusion]
110	P45:	[conclusion]
111	P28:	[ok]
112	P45:	[ok]

P28 started this part of the task-oriented interaction by producing a directive that is prefaced by a first person epistemic stance marker. In this turn, P28 referred to their next course of action (i.e., putting the outline in the correct order) (line 91). P45 agreed with P28's assessment, as evidenced by her affirmative response, *mm-hm*, in line 92. After a four-second pause, during which both P28 and P45 oriented to their task sheets, P28 pointed at her task sheet and stated that two of the paragraphs in the task sheet seemed to be the body paragraphs (line 93). After this sentential turn construction unit, P45 delivered the emphatic change-of-state particle ↑*OH::::↓* in line 94 (note the pitch shifting, prosodic stress, loudness and elongation), which suggests that P45 was (a) unaware that the task involved re-arranging the essay's outline, or (b) unaware that the two parts of the task sheet P28 was pointing at could be the essay's body paragraphs, or both (a) and (b). After the change-of-state particle, P28 turned to P45 and addressed the difficulty that P45 was experiencing (line 95). P28 asked P45 if she had noticed <u>*that*</u> (line 95), to which P45 replied that she hadn't (line 96). After this, in line 97, P28 explained that they needed to re-arrange the outline before

they actually started composing the essay. P28's support was successful as evidenced in line 98, where P45 was able to identify the opening paragraph. P28 immediately agreed with P45's contribution (notice the latching in line 99) and subsequently (after a five-second pause while writing on the task sheet) identified the two body paragraphs (line 100). This prompted a question on the part of P45, who asked if there were any differences in terms of the structure of the two body paragraphs (line 101). Given P45's question (note the slightly rising intonation), an answer from P28 was expectable. However, this first-pair part of the adjacency pair (i.e., question) was not followed by a second-pair part (i.e., answer) from P28. After a two-second pause, P45 answered her own question (line 102). At this point, P28 produced a first-pair part (i.e., she asked P45 to locate the paragraphs in the task sheet). After this, P45 oriented to the task sheet and produced a second-pair part (i.e., a verbal answer to P28's question) and simultaneously showed P28 which two paragraphs she was referring to (line 104). P28 confirmed that there were no differences between the two body paragraphs (line 105). P45 seemed satisfied with the answer, smiled, and pointed at the first and second body paragraphs in the task sheet (lines 106–108). P28 took the next turn and when she was about to identify the function of the remaining paragraph (line 109), P45 self-selected and P28 and P45 overlapped and uttered *conclusion* at the same time. This indicates that P28 and P45 were orienting to the same goal – identifying the function of each paragraph. This was also evidenced by another overlap in the following turn. They uttered *ok* together, which functioned as self and other confirmation.

Dyad 1 took 18 turns to re-arrange the essay's outline before they started writing the essay. This type of extended pre-writing negotiation characterizes the interaction in the complex task condition, where participants had to reason more to create an essay from an unordered outline. As illustrated above, reasoning demands induced in the cognitively complex task led students to engage in a greater amount of interaction in order to cope with the high task demands. During this process, participants mainly followed a sequence of actions to accomplish the task of re-arranging the essay's outline: (1) one participant identified the task at hand (i.e., putting the outline in the correct order), (2) both participants collaboratively identified the order of the paragraphs presented in the outline, and (3) both participants provided peer assistance to each other about task procedures.

In contrast, this type of extended interaction was almost completely absent in the simple task condition pair. Moreover, the action sequences taken by the pair in the simple task were different: (1) one participant read (verbatim) from the task sheet the information to be included in the introductory paragraph, and (2) the other participant presented candidate sentences to start the introductory paragraph. See Excerpt 2 from Dyad 2's pre-writing interaction, which involved a Chinese female student (P3) and a Ukrainian female student (P20). In this segment, P3 and P20 had already decided

on their position on the topic as well as the arguments and opposing arguments to be included in their essay. This marks a sharp contrast with Dyad 1, which took almost 11 minutes before starting to write an essay. Excerpt 2 illustrates the segment where participants were about to start writing the introductory paragraph.

Excerpt 2.

21	P3:	I think the thi:rd [one is e]asy to refute
22	P20:	[mm hm]
23	P3:	so:: are we doing the th[ird one and then,]
24	P20:	[yeah let's do the s]econd one second one
25	P3:	second one and third one
26	P20:	mm hm
27		ok we'll start with (1.0) ((looks at the task-sheet))
28	P20:	introduction to the general topic ((reads from the task sheet))
29	P3:	introduction to [the general topic]
30	P20:	[°state why the to]pic is important and thesis statement°
31		oh my go::d (2.0)
32		I have no idea what would be general topic introduction
33	P3:	some students believe that cmu should provide more ethnic
34		food options on the contrary others argue that ok ((reads the prompt
35		from the task sheet))
36		>°you should provide general information about the topic explaining the
37		situation so that the reader can make sense of the topic and the claims you
38		make°< ((reads from the task sheet)) (4.0)
39	P20:	I feel so stupid ((laughs))
40	P3:	<so cmu has like a variety of restaurants and dining locations
41		however some international students>
42	P20:	↑oh how about we start that cmu is an internationally diverse
43		university
44	P3:	ok

In this sequence, from lines 21 to 26, P3 and P20 were deciding which arguments and opposing viewpoints to include in their essay (notice that they used ordinal numbers to refer to the arguments and opposing viewpoints presented in the T-chart). Following this, the interjection *ok* and the incomplete statement produced by P3 in line 27 marked the beginning of the task proper. After this, P20 took the next turn and completed P3's statement claiming that they should begin the essay by providing an introduction to the general topic (line 28). P3 confirmed this by repeating P20's statement (line 29). Before P3 finished her turn, P20 self-selected and added that they needed to state why the topic was important and that they needed to provide a thesis statement (line 30). P20 struggled with this part of the task, as evidenced in her remark that she had no idea what a proper topic introduction could be. P3 took the next turn and read the task's prompt verbatim from the task sheet. She then turned to the third page of

the task sheet and read (verbatim) the information they were supposed to include in the introductory paragraph. P20's struggle remained, as evidenced by her comment in line 39. Following this, P3 presented a candidate sentence to start the introductory paragraph (line 40–41). P20 took the next turn and produced a high-pitched ↑*oh*, followed by an alternative introductory sentence. P3 agreed with P20's suggestion (*ok*). Upon P3's agreement, they started writing the introduction paragraph.

As we can see from this excerpt, there is not much negotiation between P3 and P20. In addition to the minimal negotiation, students in this dyad displayed a more decisive and committed speech delivery. As opposed to students in the complex task dyads, the speech of the students in this dyad in Excerpt 2 is characterized by the use of straight and final falling intonation, short pauses and by the lack of epistemic markers (both adverbial expressions, such as 'maybe', and verbal expressions, such as 'I believe' or 'I think' (see Biber et al., 1999).

In summary, the simple task condition pair (Dyad 2) simply followed the ready-made outline presented in the task sheet. Both students attended to the task sheet as their primary locus of support, and their interaction revolved around the task sheet by reading from it verbatim. This sharply contrasts with Dyad 1 from the complex condition, in which participants did not solely depend on the task sheet and instead solved problems over multiple turns. During this process, they engaged in collaborative reasoning processes in which they jointly attended to the problems at hand and resorted to each other's cognitive resources to re-organize the essay's outline prior to writing. They also complemented each other's contribution as they proceeded to achieve the task goal. Hence, although both dyads showed instances of peer support, the source of such support was different. Peer support in Dyad 2 centered around the task sheet with one participant reading directly from the task sheet for the other participant. In contrast, participants in Dyad 1 provided support from their own cognitive resources while trying to work out the information they needed to achieve the task goal. The frequent occurrence of epistemic stance markers (*I think*) during interaction also indicates that they were searching for information from their cognitive environment. Epistemic modality displays the speaker's evaluation of his/her own knowledge about a proposition. Because epistemic stance markers such as 'I think' convey the speaker's doubts and uncertainties about a proposition, they naturally trigger the listener's attention to those doubts and uncertainties. Responding to the speaker's doubts, the listener contributes his/her own thoughts and ideas to the problems at hand. As a result, the doubts and uncertainties originally expressed by the speaker become a shared state between the speaker and listener, leading to a joint reasoning process. In short, rather than merely copying information from available sources, participants in the complex task engaged in reasoning processes and offered solutions from their own perspectives. The extensive reasoning processes they went through resulted in more extended turn taking, greater negotiation sequences, and use of epistemic stance markers to display their own understanding and to respond to each other's understanding.

During-writing negotiation over sources of trouble and actions to take

Similar interaction patterns were found in the during-writing phase when participants negotiated to solve trouble sources or to clarify actions to take while completing a task. See Excerpt 3 from Dyad 1 (complex task condition) for illustration. Here, P28 and P45 were starting to write their introductory paragraph and were discussing the structure of the paragraph.

Excerpt 3.

```
113   P28:   so↑ (4.0)
114          we should start our introduction ((both students look at the task sheet))
             (9.0)
115   P45:   ok (3.0)
116          for introductio:n, ((looking at the task sheet)) (4.0)
117   P28:   wait do we:
118          oh no no no no never mind ((looking at the task sheet))
119          so do we <ackno:wledge i:t, and then refu:te i:t, <
120          or we just refu:te it, (2.0)
121   P45:   u::h
122   P28:   seems like >we have to do [both]<
123   P45:                            [you] can partially refute it
124          I think so:
125   P28:   yeah so we still have to acknowledge (1.0)
126   P45:   mmm actually here it only gives us instruction on refutation
127          ((looking and pointing at the task sheet)) (2.0)
128   P28:   probably we can do both (3.0)
129   P45:   o::r you can say it is not that convincing,
130          just don't say that it's completely wrong
131   P28:   ok ok (3.0)
```

The interjection *so*↑ produced by P28 in line 113 marks the end of the task-prefatory talk (Hellermann, 2007) and the beginning of the task proper. In line 114, P28 suggested writing the introductory paragraph. After a nine-second pause, during which both participants were looking at the task sheet (possibly looking at the information they were supposed to include in the introduction), P45 accepted P28's suggestion in line 116, but this turn is incomplete. Following this, after a four-second pause, P28 uttered an incomplete question that marks the beginning of a trouble source and negotiation sequence (line 117). The source of trouble was not made explicit until line 119 where P28 expressed her doubt. She asked whether they should acknowledge the opposing viewpoint in the introduction and then refute it, or whether they should just refute it without acknowledging it. P45 was not sure about the answer either, as evidenced in her response, *u::h*, in line 121. Following this, P28 proposed an idea in line 128 (they should do both, acknowledging and refuting). However, overlapping this turn, P45 suggested another idea (*partially* refuting it) (lines 123–124). P28 only

showed a partial agreement (*yeah*) to P45's suggestion (line 125). Responding to this, P45 oriented to the task sheet and told P28 that the task sheet only asked them to refute the opposing argument (not acknowledging it). But P28 still disagreed and said that they could probably do both (i.e., acknowledge and refute the opposing view-point). Yet, the hedging 'probably' in P28's utterance indicates her uncertainty. Then, starting the next turn with *or*, P45 presents an alternative idea (saying that the opposing viewpoint is not convincing). P28 finally accepted this idea (saying *ok* twice).

As seen in the excerpt above, participants in Dyad 1 encountered a trouble source and spent a long time figuring out what information they could include in the introduction paragraph. This was evident in the frequent within-turn and between-turn pauses appearing in the excerpt. During these pauses participants were probably processing the information in the task sheet and deciding which information to include in their introductory paragraph. Being in the complex task condition, students in Dyad 1 often encountered this kind of trouble during interaction. Intensity of those trouble sources, as evidenced in frequent pauses and lengthy negotiation sequences over uncertainty, is a reflection of the depth of processing posed by the task demands in the cognitively complex condition.

This pattern contrasts once again with Dyad 2 from the simple task group (see Excerpt 4). Here, participants (P3 and P20) were starting to write the first body paragraph, but they encountered uncertainty in terms of the information presented in their introductory paragraph. When this segment started, P3 and P20 had already written their introductory paragraph (reproduced here before Excerpt 4).

> Carnegie Mellon University is known for its international student body. In order to embrace diversity, the school tries to support the preferences of students with different backgrounds. One of the examples is the variety of on-campus dining options. While some students really enjoy the ethnic food offered by the university and wish that there are even more options, others think that the university has already offered enough meal options.

Excerpt 4.

```
271   P20:  ok so we can s[tart]
272   P3:              [oppo]nent's argument u:mmm ((reads from the task sheet))
             (5.0)
273   P20:  ↑wait should w- should we present our point of view, ((looks at the screen))
274   P3:   yeah
275   P20:  °state the position in support° u:hhh
276          we have to include our point of view over here
277   P3:   ok umm
278          we: believe that ((types))
279          >can we start with that,<
```

280   P20:   ((nods))
281   P3:   we believe that umm (2.0)
282   the university should provide more ethnic food options
283   °that cmu should° <provide more ethnic food options> ((types))

In line 271, P20 produced an incomplete turn, which is completed by P3 in line 272. This indicates that P3 possibly anticipated the next course of action: discussing how to start the first body paragraph. P3 read aloud the outline given in the task sheet, stating that they could start the first body paragraph introducing the opponent's viewpoint (line 272). After a five-second pause, P20 looked at the computer screen (i.e., the essay they were writing) and said ↑*wait* with a sharp, rising pitch. This move marked the beginning of a trouble source. P20 asked if they should present their point of view in the introductory paragraph (before moving to the body paragraph). Although they had already written the introductory paragraph (see the introductory text above), P20 realized that they had not stated their position on the topic in the introduction (i.e., whether the university should offer more ethnic food options). Following P3's acknowledgement (*yeah*), P20 read softly from the task sheet and said that they needed to add their point of view right after the last sentence in the introduction (line 276). P3 immediately agreed. She uttered and simultaneously typed *we: believe that* as a way to expressing their point of view (line 278) and then sought P20's confirmation on this linguistic expression (note the slightly rising intonation) (line 279). P20 did not produce a verbal response but nodded, so P3 kept her turn and completed her thought (and typed it) (lines 280–282).

As we can see in this excerpt, Dyad 2 needed fewer turns than Dyad 1 to resolve the trouble source. Dyad 2 took only 4 turns to resolve the trouble source, while Dyad 1 took at least 9 turns to resolve the trouble sources. In addition, the interaction between the participants in the complex task condition is characterized by the lengthening of sounds, the use of slightly rising intonation, and slower speech, all of which denote doubt and uncertainty (see Excerpt 3). However, these features were almost absent in the simple task condition. This is not surprising. The increased reasoning demands of the complex task in Dyad 1 led to more trouble sources and, consequently to greater needs to interact and negotiate, which eventually resulted in more extended negotiation sequences.

Discussion

Although teaching speech act patterns and politeness strategies has been a prevailing trend in instructed pragmatics research (Taguchi, 2015), this study adopted a task to present information about the conventions of genre (rhetorical moves and linguistic forms for realizing the moves). Hence, the study expanded the scope of instructional

targets from utterance-level lexico-grammatical forms for conveying illocutionary force, to written discourse-level systems for achieving a socially recognized communicative goal, that is, the goal of persuasion. Collaborative writing tasks designed in this study showed how learners attended to the conventions of persuasive writing during the task, thereby creating interaction-driven learning opportunities around the target pragmatic features (linguistic forms and rhetorical moves of persuasion).

In addition, our findings revealed that different cognitive demands induced by task design indeed led to different interaction demands, resulting in different patterns of interaction around the target pragmatic features. Students who performed the 'simple' and the 'complex' task differentially negotiated how the essay should be constructed and which linguistic forms should be used to realize the rhetorical moves, which probably facilitated their learning of the structure and forms. Post-task assessment of learning outcomes indeed confirmed this (see Gomez-Laich, 2017). Students in the complex task group used significantly more rhetorical moves and linguistic forms in their persuasive essays than students in the simple task group (data not reported in this paper). A peer-to-peer interaction as a social activity helped develop students' abilities to construct a persuasive essay, and the abilities were better aided by the cognitively complex task.

Although the data presented in this paper is limited, we were able to identify different types of interactional engagements between pairs in the simple and complex task condition. Analyses revealed how students co-constructed a text via interaction. The tasks used in this study afforded students opportunities for learning linguistic forms and rhetorical moves that were necessary for effective persuasion. Increased task demands in the complex task condition led to a greater amount of negotiation around an essay outline at the pre-writing stage, resulting in longer negotiation sequences and more extended turn taking than the simple task condition. The increased task demands also invoked more trouble sources as participants proceeded to achieve the task goal of collaboratively writing an essay. As we have illustrated, intensity of negotiation and repair work around the trouble sources were characterized by various interactional features, such as sound lengthening, rising intonation, slower speech, frequent pauses, and use of epistemic stance markers of uncertainty. These features indicated participants' tentative, indecisive manner of speaking, which reflected the depth of reasoning process that they were engaged in.

The different interaction patterns that emerged in the simple and complex task inform us that the quality of metalinguistic discussion around target pragmatic features can be influenced by task complexity. In other words, task complexity can be manipulated strategically in order to enhance the quality and depth of discussion about target linguistic and rhetorical features, which could eventually lead to a greater level of learning of the target features (see Gomez-Laich, 2017, which confirmed this finding). Interactional features observed in the cognitively more complex task (e.g.,

longer negotiation sequences and repair work) seem to reflect the influence of the complexity of the task.

To date, most studies on task complexity and learner-learner interaction have compared frequencies of language-related episodes (LREs) produced by learners while completing a complex or a simple task (e.g., Kim, 2012; Kim & Taguchi, 2015; Révész, 2011; Robinson, 2001b, 2007). These studies revealed that increased task complexity resulted in more interaction-driven learning opportunities as evidenced in frequent occurrences of LREs. Findings from this study complement the previous findings by revealing the nature and quality of LREs. Increased task complexity did result in different interaction patterns as evidenced in a greater number of turns and longer negotiation sequences and exchanges to clarify ambiguity and understanding. To the best of our knowledge, this is the first study that employed a conversation-analytic inspired approach to examine how students co-regulated their performance to co-construct texts and learning opportunities while performing tasks of different cognitive complexity levels. To this end, findings from this study suggest some methodological implications, notably that a close examination of collaborative dialogues is crucial to complement and validate the quantitative analyses of interaction. One of the main affordances of Conversation Analysis (CA) is its ability to describe and explain how participants achieve the organization of social action (co-construction of an essay, in this case) step-by-step in real time (Kasper & Wagner, 2014). Comparing interactions between dyads in different task conditions can help identify patterns of interactions between participants who complete cognitively complex tasks and those who complete simple tasks that are not visible by simply counting LREs.

Limitations and directions for future research

There are several limitations in this study, which need to be addressed in future research. One limitation relates to the restricted range of pragmatic acts focused on. The collaborative writing task used was limited to the communicative goal of persuading an audience. Other communicative goals such as expressing empathy, giving directions, and providing constructive criticism can be incorporated into a collaborative writing task in order to generate metalinguistic discussion around pragmatic resources used to achieve these communicative goals.

In addition, qualitative analyses of students' interaction data focused only on high-performing dyads from the simple and complex task condition. Future studies should also analyze whether patterns of collaborative interaction differ between a high and low performing dyad from each task condition. Such analyses will reveal similarities and/or differences in terms of how low and high performing pairs orient to the task, negotiate epistemic stance, plan the writing of their essay, and distribute labor when co-constructing their essay.

References

Austin, J. L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Bardovi-Harlig, K., & Mahan-Taylor, R. (2003). *Teaching pragmatics*. Washington, DC: US Department of State, Office of English Language Programs.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.

Clark, H. H. (1996) *Using language*. Cambridge: Cambridge University Press.
doi: 10.1017/CBO9780511620539

Cohen, A. & Tarone, E. (1994). The effects of training on written speech act behavior: Stating and changing an opinion. *MinneTESOL Journal*, 12, 39–62.

Crystal, D. (1997). *The Cambridge encyclopedia of language* (2nd ed.). New York, NY: Cambridge University Press.

Cutting, J. (2015). *Pragmatics* (3rd ed.). Oxford: Routledge.

Dippold, D. (2011). Argumentative discourse in L2 German: A sociocognitive perspective on the development of facework strategies. *Modern Language Journal*, 95, 171–187.
doi: 10.1111/j.1540-4781.2011.01175.x

Ede, L., & Lunsford, A. (1990). *Singular texts/plural authors*. Carbondale, IL: Southern Illinois University Press.

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. New York, NY: Routledge.

Gilabert, R., Barón, J., & Llanes, A. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, 47, 367–395. doi: 10.1515/iral.2009.016

Gilabert, R., & Barón, J. (2013). The impact of increasing task complexity on L2 pragmatic moves. In A. Mackey & K. McDonough (Eds.), *Second language interaction in diverse educational settings* (pp. 45–69). Amsterdam: John Benjamins. doi: 10.1075/lllt.34.06ch3

Gomez-Laich, M. P. (2017). *Effects of task complexity on ESL students' academic writing*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.

Graff, G., & Birkenstein, C. (2010). *They say. I say. The moves that matter in academic writing*. New York, NY: Norton.

Hellermann, J. (2007). The development of practices for action in classroom dyadic interaction: Focus on task openings. *The Modern Language Journal*, 91, 83–96.
doi: 10.1111/j.1540-4781.2007.00503.x

Huth, T. (2006). Negotiating structure and culture: L2 learners' realization of L2 compliment-response sequences in talk-in-interaction. *Journal of Pragmatics*, 38, 2025–2050.
doi: 10.1016/j.pragma.2006.04.010

Hyland, K. (1990). A genre description of the argumentative essay. *RELC Journal*, 21, 66–78.
doi: 10.1177/003368829002100105

Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, 30, 437–455. doi: 10.1016/S0378-2166(98)00009-5

Ifantidou, E. (2011). Genres and pragmatic competence. *Journal of Pragmatics*, 43, 327–346.
doi: 10.1016/j.pragma.2010.07.016

Johnstone, B. (1989). Linguistic strategies and cultural styles for persuasive discourse. In S. Ting-Toomey & F. Korzenny (Eds.), *Language, communication, and culture: Current directions* (pp. 139–156). Beverly Hills, CA: Sage.

Kasper, G., & Wagner, J. (2014). Conversation analysis in applied linguistics. *Annual Review of Applied Linguistics*, 34, 1–42. doi: 10.1017/S0267190514000014

Kim, Y. (2009). The effects of task complexity on learner-learner interaction. *System*, 37, 254–268. doi: 10.1016/j.system.2009.02.003

Kim, Y. (2012). Task complexity, learning opportunities, and Korean EFL learners' question development. *Studies in Second Language Acquisition*, 34, 627–658.
doi: 10.1017/S0272263112000368

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom contexts: The role of task complexity. *The Modern Language Journal*, 99, 656–677.
doi: 10.1111/modl.12273

Linell, P. (1998) *Approaching dialogue: Talk, interaction and contexts in dialogical perspectives*. Amsterdam: John Benjamins. doi: 10.1075/impact.3

Long, M. (2015). *Second language acquisition and task-based language teaching*. Oxford: Wiley-Blackwell.

Loewen, S., & Basturkmen, H. L. (2005). Interaction in group writing tasks in genre-based instruction in an EAP classroom. *Journal of Asian Pacific Communication*, 15, 171–189.
doi: 10.1075/japc.15.1.11loe

Martin, J. R. (1992). *English text: System and structure*. Amsterdam: John Benjamins.
doi: 10.1075/z.59

Miller, R., Mitchell, T., & Pessoa, S. (2014). Valued voices: Students' use of engagement in argumentative history writing. *Linguistics and Education*, 28, 107–120.
doi: 10.1016/j.linged.2014.10.002

Nuevo, A., Adams, R., & Ross-Feldman, L. (2011). Task complexity, modified output, and L2 development. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 175–202). Amsterdam: John Benjamins.
doi: 10.1075/tblt.2.13ch7

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, 95, 162–181.
doi: 10.1111/j.1540-4781.2011.01241.x

Révész, A., Kourtali, N., & Mazgutova, D. (2016). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, 67, 208–241.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22, 27–57.
doi: 10.1093/applin/22.1.27

Robinson, P. (2003a). Attention and memory during SLA. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell.
doi: 10.1002/9780470756492.ch19

Robinson, P. (2003b). The Cognition Hypothesis, task design, and adult task-based language learning. *Second Language Studies*, 21, 45–105.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching*, 45, 3, 193–213. doi: 10.1515/iral.2007.009

Robinson, P. (2011a). Task-based language learning: A review of issues. *Language Learning*, 61, 1–36. doi: 10.1111/j.1467-9922.2011.00641.x

Robinson, P. (2011b). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–37). Amsterdam: John Benjamins. doi: 10.1075/tblt.2.05ch1

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735. doi: 10.1353/lan.1974.0010

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139173438

Storch, N. (2013). Collaborative writing in L2 classrooms. Toronto: Multilingual Matters.

Storch, N., & Wigglesworth, G. (2007). Writing tasks: Comparing individual and collaborative writing. In M. P. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 157–177). London: Multilingual Matters.

Street, B. (2009). Hidden features of academic paper writing. *Working Papers in Educational Linguistics*, 24, 1–17.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, 82, 320–337. doi:10.1111/j.1540-4781.1998.tb01209.x

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. *Language Teaching*, 48, 1–50. doi:10.1017/S0261444814000263

Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford: Oxford University Press.

Unger, C. (2006). *Genre, relevance and global coherence: The pragmatics of discourse type*. New York, NY: Palgrave MacMillan. doi:10.1057/9780230288201

Wingate, U. (2012). Argument! Helping students understand what essay writing is about. *Journal of English for Academic Purposes*, 11, 145–154. doi:10.1016/j.jeap.2011.11.001

Zhao, H., & Kaufer, D. (2013). DocuScope: Analyzing pragmatic competence in second language writing of genres. In N. Taguchi & J. M. Sykes (Eds.), *Technology in interlanguage pragmatics research and teaching* (pp. 235–260). Amsterdam: John Benjamins. doi:10.1075/lllt.36.12zha

Appendix A. Transcription conventions

(adapted from Sacks, Schegloff, & Jefferson, 1974)

(2.0)	Timed pause (2.0 seconds or more)
.	Falling intonation
,	Slightly rising intonation
?	Raised intonation (not necessarily a question)
word	Underlined words (or parts of words) indicate stress
(word)	Single parentheses indicate uncertain hearing
( )	Empty parentheses indicate inaudible speech
::	Sound lengthening. Multiple colons indicate more prolongation
–	Abrupt cut off
((comment))	Double parentheses contain transcriber's comments or descriptions
CAPITALS	Capital letters indicate markedly loud speech
°word°	Markedly soft sound relative to surrounding context
↑	Markedly higher pitch relative to surrounding talk
↓	Markedly lower pitch relative to surrounding talk

=	Latched utterances
<word>	Slower speech relative to surrounding context
>word<	Faster speech relative to surrounding talk context
[…]	Indicates that a section of the transcript has been omitted
[	Onset of overlapping speech
]	End of overlapping speech
=	latching

Using tasks to elicit pragmatics language use

CHAPTER 5

Task modality effects on Spanish learners' interlanguage pragmatic development

Derek Reagan & Caroline Payant
Georgetown University / Université du Québec à Montréal

Within the field of second language acquisition (SLA), we have witnessed a rise in research on task-based language teaching (TBLT) and its effects on L2 development (Kim, 2015). However, few studies have examined how TBLT could facilitate the development of interlanguage pragmatics (Taguchi & Kim, 2015), an issue which this volume aims to address. Moreover, whether task modality (i.e., oral versus written tasks) mediates development has yet to be investigated. The current study with learners of Spanish focused on the effects of using pedagogical tasks and on the manipulation of task modality on learners' L2 pragmatic competence through the production of Spanish requests and speech act modifications. Two intermediate classes ($n = 25$) of Spanish completed either an oral or written story completion task. Drawing on oral and written Discourse Completion Tests, we found that tasks positively impacted learners' production of L2 requests. However, significant differences between modality groupings were not identified.

Introduction

Within the field of second language acquisition (SLA), there is substantial research on task-based language teaching (TBLT) and its effects on second language (L2) development (Bygate, 2015; Gass & Mackey, 2007; Kim, 2015; van den Branden, 2016). Placing tasks at the center of teaching and learning creates opportunities for learners to produce output, to notice gaps in their interlanguage, to obtain corrective feedback, and to modify output (Swain, 2005). A number of studies have examined the impact of task designs and task implementation on linguistic and interactional features (Ahmadian & Tavakoli, 2010; Bygate, 2001; Lynch & Maclean, 2000; Newton, 2013; Philp, Oliver, & Mackey, 2006). These empirical papers were concerned with grammar, vocabulary, and/or interactional features. Few studies have focused on the effects of specific instructional approaches, like task-based language teaching, on the development of pragmatic competence (Plonsky & Kim, 2016). Rather, within the field of instructed pragmatics, the focus has been on identifying the differences between explicit and implicit instruction (see Taguchi, 2015).

Another perceived gap within TBLT research relates to oral and written modes of production, which are hypothesized to uniquely influence the cognitive processes underlying L2 development (Gilabert, Manchón, & Vasylets, 2016). In the oral modality, language is rapid and nonpermanent, which in turn can constrain intake and feedback. When producing language in the written modality, learners can strategically distribute their cognitive resources, enabling them to plan and edit their language with greater awareness. Research in this area has shown modality effects on complexity, accuracy, and fluency of output (Kormos, 2014; Kormos & Trebits, 2012) and on interactional features (Adams & Ross-Feldman, 2008; Payant & Reagan, 2018). To date, how pedagogical tasks and task modality mediate the development of pragmatic competence is unclear. Thus, the goal of the present study was to expand our understanding of TBLT and examine the impact of modality on the development of pragmatic competence of requests with intermediate language learners of Spanish. Before turning to the study, we provide a brief review of research into interlanguage pragmatics and the impact of modality.

Literature review

Interlanguage pragmatics

Early work on pragmatic competence adopted a comparative stance where the identification of discrepancies between L2 learners' pragmatic production of target language speech acts and those of native speakers was of primary interest (Kasper, 1996; Kasper & Schmidt, 1996; Kasper & Rose, 2002). Recently, researchers have turned their focus to nonnative speakers' interlanguage pragmatics (ILP) competence. This work has generated a significant body of literature on various speech acts, including apologies, invitations, refusals, and requests (see Takahashi, 2010, for a review). For the current study, we selected the acquisition of requests as the target speech act. This is an area that has received a lot of attention within instructed pragmatics research; however, most studies have focused on methods of instruction (i.e., explicit and implicit instruction) and we have yet to explore whether task modality mediates the development of requests. In addition, from a pedagogical stance, there is value in understanding how to teach the cultural conventions of Spanish requests as this knowledge may help L2 speakers learn to minimize the imposition of face-threatening requests. In line with the field of instructed pragmatics and the goals of the present study, the following section does not discuss research into the nature of speech acts in general; instead, we emphasize empirical work that has specifically examined the acquisition of pragmatics by learners of Spanish.

Previous research on instruction of pragmatics

A salient area of instructed ILP research has been concerned with explicit and implicit instruction (Jeon & Kaya, 2006; Kasper, 2001; Taguchi, 2015; Takahashi, 2001, 2010). With English learners of Spanish, Koike and Pearson (2005) explored the effects of instruction on the acquisition of suggestions. The study included 99 low-intermediate students across five treatment conditions (i.e., the activity plus explicit pre-instruction with explicit feedback, explicit pre-instruction with implicit feedback, implicit pre-instruction with explicit feedback, implicit pre-instruction with implicit feedback, and the activity with no pre-instruction or feedback). After students read and heard dialogues focusing attention on directness and pragmatic force, students identified suggestions in television scripts and subsequently created their own script using suggestions. Participants who received explicit pre-instruction along with explicit feedback produced more target pragmatic forms during the post-tests than the other four groups. In another study, in a university-level Spanish as a foreign language context, Langer (2013) examined the benefits of explicit instruction of Spanish requests with English learners of Spanish across three proficiency levels (i.e., beginner, intermediate, advanced). The researcher provided explicit instruction of the target request forms to the treatment group over a ten-week period, while the control group did not receive any explicit pragmatic instruction. Results from the Discourse Completion Tests (DCTs) confirmed a positive impact of explicit instruction across proficiency levels.

Focusing on the acquisition of refusals, Félix-Brasdefer (2008) examined the effects of explicit instruction with fifth-semester Spanish language university-level students. Over the course of an academic semester, the experimental group received explicit instruction of refusals in the form of metapragmatic information, while the control group did not. Results from the open-ended roleplays showed that the experimental group produced fewer inappropriate direct refusal acts and more indirect refusal strategies than the control group. Félix-Brasdefer also notes that the production of refusal strategies by the experimental group held constant for the delayed posttest, although, indirect strategy use did regress slightly.

Research has also examined the acquisition of requests across different learning environments. For instance, Sykes (2009) examined the acquisition of requests by advanced learners of Spanish in a synthetic immersive environment (SIE). In this virtual world, learners completed five quests, each requiring that they made unique requests. Development of pragmatic knowledge was measured via DCTs. Although they found limited changes on the DCT scores, researchers reported an impact on the participants' awareness of Spanish pragmatics. In a subsequent study, Sykes (2013) investigated the acquisition of Spanish apologies through the use of a SIE. After completing apology scenarios within the SIE, the advanced learners of Spanish exhibited a shift from speaker-oriented to hearer-oriented apology-making strategies on the

posttest. The learners also reported a change in their self-perception of their apology-making abilities in interviews and surveys. There is substantial evidence that explicit instruction can contribute to learners' ILP development.

In sum, research on the acquisition of Spanish speech acts has primarily focused on the effects of explicit/implicit instruction. The field has started to incorporate additional cognitive models and socially-oriented theoretical paradigms as well as instructional approaches (Taguchi, 2015). While the theoretical scope of studies is expanding beyond the explicit/implicit binary, the current data regarding the effects of TBLT on the acquisition of speech acts is limited, as the next section shows.

Task-based language teaching research

In the field of instructed SLA, we have seen a significant shift toward the use of pedagogical tasks to support instruction (Kim, 2015; Loewen, 2015; Van den Branden, 2006). From a cognitive-interactionist perspective, pedagogical tasks create opportunities for learners to use language that mirrors real-world linguistic demands (Bygate, 2015). By completing pedagogical tasks, learners focus primarily on conveying meaning. In addition, learners can focus on form by interacting with authentic input, notice linguistic gaps in their interlanguage, receive corrective feedback, and produce pushed output (Gass & Mackey, 2015; Swain, 2005).

Teaching pragmatics via tasks

Empirical findings, albeit limited, appear to support the use of tasks for pragmatic instruction. Taguchi and Kim (2016) examined the influence of task-based instruction on learners' production of target pragmatic requests and pragmatic-related episodes (PREs) with 74 Korean learners of EFL. PREs were operationalized as instances where learners attend to and question their use of pragmatic forms and information. Learners completed writing tasks (i.e., drama script tasks), either individually or in pairs. ILP development was measured using written DCTs. They found that learners' use of direct head acts improved on immediate post-tests. With respect to external modifications, the use of preparators was more frequent under both conditions, and was sustained over time. Moreover, a moderate impact on internal modifications such as hedging strategies was identified in the immediate post-test only. In addition to manipulating grouping variables, Kim and Taguchi (2015) considered task complexity in relation to the production and resolution of PREs. A total of 73 Korean learners of English completed either a simple ([−reasoning]) or a complex ([+reasoning]) writing task. Results show that while task complexity did not significantly affect the quality of task performance, those who participated in collaborative tasks produced a higher frequency of target head acts and request modifications. Finally, Kim, Lee, and Kim (this volume) compared Korean heritage and foreign language students' performance

during collaborative tasks. The Korean language system requires the use of honorific particles. As such, learners of Korean must develop both pragmalinguistic knowledge (i.e., knowledge of specific relevant linguistic forms) and display sociopragmatic knowledge (i.e., understanding of how they are used according to context) to produce target-like utterances. They found that learners focused more frequently on linguistic forms (e.g., honorific verb particles, honorific noun particles) compared to discussions of sociopragmatic factors (e.g., discussing the relationships between interlocutors) that would impact the use of forms. Furthermore, heritage language learners tended to be more successful in resolving PREs, compared to foreign language learners. These three studies appear to show that task-based instruction can create authentic situations for negotiation and subsequent development of ILP. Given the limited number of studies focusing on pragmatics, however, additional studies with learners from different linguistic backgrounds and proficiency levels are needed.

Task modality

TBLT research has manipulated various task design variables, such as task complexity (Gilabert, Barón, & Llanes, 2009; Kim & Payant, 2014; Révész, Sachs, & Mackey, 2011; Robinson, 2011), as well as implementation variables, including task modality, task repetition, and task familiarity (Ahmadian & Tavakoli, 2011; Bygate, 2001; Kim, 2013; Lynch & Maclean, 2000; Payant & Reagan, 2018; Pinter, 2005, 2007a, 2007b; Sample, & Michel, 2015). There is growing interest in modality given the unique demands of speaking and writing (Bygate, 2015; Kormos, 2014). Speaking, compared to writing, limits the amount of explicit attention to the accuracy of form as it is fleeting whereas writing enables learners to more carefully plan their message and language since the output is more permanent (Williams, 2012). In addition, through the production of written output, learners may draw on declarative knowledge (i.e., explicit information relating to grammatical rules) to reflect on and edit their output over time (Ellis, 2003). The time lapses between conceptualizing a message and producing a message by activating explicit knowledge may positively impact learning. Also, oral and written forms of communication influence the interlocutor's ability to save face. During oral interaction, compared to written interaction, there may be greater immediacy for saving face in face-threatening acts. The immediate reactions experienced by the interlocutors may increase awareness of context-specific politeness strategies.

Looking into how modality affects interaction-driven learning opportunities, Adams and Ross-Feldman (2008) implemented two types of collaborative tasks with 44 English as a second language learners. The first task type was a decision-making task: Learners reached a consensus, orally, for a seating chart and described everyone's position at the table, in writing. The second task type was a story completion task. Language-related episodes (LREs), operationalized as utterances where learners question

their output, were found to be more frequent during the written components of the two tasks. In another study, Niu (2009) investigated how different modes of output impacted the production and resolution of LREs with 16 Chinese learners of English. EFL learners performed a collaborative text reconstruction task either in writing or orally on two occasions with a partner. The results showed that the written modality group produced a greater number of LREs compared to the oral group. However, the oral group discussed aspects of pronunciation more frequently than the written group, which tended to focus on spelling, perhaps not surprisingly given that that pronunciation and spelling are features of spoken and written output, respectively. Kormos and Trebits (2012) examined how English L2 performance (i.e., complexity, accuracy, and fluency) was affected by task modality, task type, and language aptitude. The authors implemented two task types (i.e., cartoon description and picture narration) in two modalities (i.e., oral and written) with 44 Hungarian-English bilinguals. After having completed the four narrative tasks, the participants displayed greater accuracy in their written output and tended to use more varied vocabulary compared to their oral output. No differences in terms of syntactic complexity were identified between the two modalities.

Finally, Payant and Kim (2015) assessed how task modality influenced the specific mediating functions of language produced during learner-learner interaction and further examined how learners mediated their target language output through use of their entire linguistic repertoire (i.e., first, second, and third languages). The Spanish-English learners of French as a third and foreign language completed three types of tasks, each having an oral and written modality. The findings indicated a greater amount of turns focusing on grammar during the written modality compared to greater discussions of vocabulary and task management during the oral modality. In sum, task modality appears to be a variable that mediates interactional features and language output. To better understand the role of modality, it is useful that we expand the current research to include additional language features, such as pragmatic competence.

Purpose of the study

The purpose of this study is to explore to what extent task-based instruction promotes language learners' development of pragmatics while engaged in either an oral or a written task. In particular the study examines whether low-intermediate learners of Spanish develop pragmatic competence of requests through the completion of story completion tasks in either an oral or a written modality. The specific research questions are:

1. How does task-based language instruction impact the development of L2 pragmatic competence in terms of differences in the production of requests, external modifications, and internal modifications?

2. How does task modality (oral or written) affect the development of L2 pragmatic competence in the form of requests with learners of Spanish? Are there differences in the production of requests, external modifications, and internal modifications?

Methods

Instructional context

The study was conducted at a land grant institution in the Pacific Northwest. Data was collected in two sections of a low-intermediate Spanish college-level course. The Spanish course met four times per week for 50-minute lessons and instruction was supplemented by an additional one-hour laboratory that focused on listening and speaking skills. In this setting, the Spanish program relies on a structure-based syllabus and the teacher reported never having experimented with TBLT. As preparation, the two authors met with her to discuss the rationale behind TBLT and provided illustrations of various task types. Four meetings were held prior to the data collection. In the first meeting, the authors discussed the major tenets of TBLT (~1 hour). During the second meeting, the course syllabus and target linguistic structures that would need to be emphasized during classroom instruction were shared with the researchers by the teacher (~30 minutes). This information was used to guide the creation of the target pedagogical materials. In the third meeting, the task-based materials were shared and discussed with the teacher to ensure that they were in line with the unit's learning outcomes (~1 hour). During the final meeting before collecting the data, the teacher rehearsed the PowerPoint (PPT) Presentation, which provided explicit instruction on the use of pragmatics in Spanish, the instructions for the DCTs, and the instructions of the tasks (~ 1 hour).

Participants

In total, 33 students participated in the study; however, only data from those who completed all stages of the study were included in the analysis. In the end, the data from 25 learners of Spanish was included. The learners' average age was 19.8 (range: 18–24). In order to register for this course, the learners needed to have taken the equivalent to two semesters of college level Spanish, or have had one to two years of high school Spanish study. Their average months of Spanish study was 13.6 (range: 8 months – 20 months) with a standard deviation of 3.9 months. Students in the humanities and social sciences are required to take two years of a foreign language at this institution; however, it should be noted that only two learners had declared Spanish as their major and three as their minor.

The participants were registered in two different sections of the same Spanish course. Each section was randomly assigned to one task condition: Learners from the

Oral Group (OG) ($n = 13$) completed a collaborative story completion task in the oral modality and those from the Written Group (WG) ($n = 12$) completed a collaborative story completion task in the written modality.

Instructional pragmatic targets

The target pragmatic form was the speech act of request. While considered to be a commonly researched pragmatic target, few studies have examined whether tasks can mediate its development. In order to identify the internal organization of requests performed in Spanish, a pilot study was conducted with six native speakers of Spanish from Spain ($n = 3$), Guatemala ($n = 1$), and Mexico ($n = 2$). Based on the results of the pilot study, corroborated with previous research on the acquisition of Spanish pragmatics (see, e.g., Blum-Kulka, et al., 1989; Félix-Brasdefer, 2005), we included three aspects: request head acts (direct and indirect), internal modification (syntactic downgraders), and external modifications (supporting reason). Table 1 defines each aspect and provides illustrative examples of each target element.

Table 1. Target pragmatic elements

Head acts	Indirect strategy Reference to requests that are hearer-oriented thereby placing the hearer in a position of control. Example: Can you let me use your notes? *¿Me puedes dejar usar tus notas?*
Internal modifiers	Syntactic downgraders Reference to syntactic choices that will reduce the impositive form of a target request, for example, conditional tense Example: Could you lend me five dollars? ¿Me *podrías* prestar cinco dólares?
External modifiers	Supporting reasons Reference to information to reduce the impositive force before or after the request head act. Example: *I was not in class yesterday*, could you let me borrow your notes? *No estuve en clases ayer*, ¿Me dejarías usar tus notas?

Instructional task

Within the TBLT framework, tasks have been operationalized in different ways (see, e.g. Ellis, 2003; Samuda & Bygate, 2008). For the current study, we followed Ellis (2003) who 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). To generate opportunities for the participants to communicatively use their linguistic and pragmatic resources, we devised a two-way story completion gap task. The language

produced by the participants to create the story line and to develop dialogues between the characters would mirror authentic language output.

The images used for the story completion task focused on environmental initiatives. Specifically, for this two-way information gap task, learners were required to sequence their images and create a story that showcased two characters, one of whom was focused on helping the environment by seeking the help and participation of her fellow classmates. While the requests and responses were prompted based on task input, pragmatic meaning was a central aspect and learners were required to use language that reflected real-world demands (Ellis, 2003). To encourage authentic output and activate the target language for making requests and refusals between the characters, each learner described the characters and their actions of their three respective photos to their peer while keeping their pictures hidden. They were instructed to carefully listen to the descriptions as they were tasked with identifying a story sequence for the six photos. Once the dyads determined the sequence of the story, they created a story based on the images. On the back of each image, more specific information that needed to be considered was provided. For instance, Figure 1 shows the two main characters who were bringing garbage to the dump. There, the young boy is thinking about playing videogames while his classmate is imagining starting a recycling program. On the back of this image, three requests were listed along with a recommended response, namely, (1) stop producing waste, a request initiated by the girl (refused by the boy), (2) play videogames, a request initiated by the boy (refused by the girl), and (3) reduce waste, a request suggested by the girl (accepted). As a part of the story, learners had to mitigate their responses for the requests and the refusals. The decision for diversifying the response (i.e., accepting and refusing) was to ensure that both the speaker and the listener were involved in the interactions for each card. For a complete set of images with requests and refusals, see Appendix A.

Request	Response
Stop producing waste (girl)	Negative
Play videogames (boy)	Negative
Reduce waste (girl)	Positive

Figure 1. Sample materials

In the OG, each learner took the role of one of the two characters on each card and had 15 minutes to role-play the information for the six cards. After their 15-minute rehearsal timeframe, they were given 10 minutes to record a final version of their role-play. In the WG, the two learners had 25 minutes to collaboratively write a scripted version of the story that included the characters and the recommended requests.

Assessment materials

ILP development was measured via oral and written DCTs. While it is important to note that DCTs are artificial in nature (Golato, 2003), as learners respond based on how they believe they would respond in a particular context which may not correspond to how they would respond in the actual context (Mackey & Gass, 2005), DCTs allow for the evaluation of learners' knowledge of normative conventions of pragmatic language use and are commonly used in research. Each DCT comprised ten items, written in English: five situations involving request speech acts and five distractor items involving non-target speech acts. The contents for the DCTs were developed in conjunction with the course materials, ensuring familiarity with target vocabulary, and did not repeat any from the target task (to minimize practice effects). Sample requests included borrowing class notes, asking for help to move, and inviting a friend to the movies. DCTs were administered to both treatment groups in the oral modality followed by the written modality, for each testing phase (i.e., pre-test, post-test, and delayed post-tests). Three versions of the oral DCTs and the written DCTs were created and were counterbalanced across the testing phases. Participants were given 5 minutes to audio record their responses to the DCT items, which were provided in writing, on their individual recorders for the oral DCTs and 8 minutes to write their responses for the written DCT.

Procedure

Given the learners' unfamiliarity with tasks, the participants completed a practice information-gap task (i.e., a decision-making task) prior to the data collection. One week later, the target task was implemented. Each Spanish section was randomly assigned to either the OG or the WG and completed the study within regularly scheduled class time. Learners self-selected their peer for the story completion task.

On Day 1, learners from both treatment groups were allotted 5 minutes to complete the pre-test in the oral modality and 10 minutes in the written modality (see Table 2). Immediately following, the teacher provided an explicit, 10-minute lesson on the request head act and target speech act modifications (Rose, 2005; Taguchi, 2015). The decision to include a brief explicit lesson was motivated by the fact that Spanish pragmatics was not part of the curriculum. The lesson, created by the authors, was delivered by the Spanish teacher via a PPT and briefly explained the importance of being familiar with Spanish pragmatics. Then, learners saw and practiced a brief dialogue that included a request and a refusal. The teacher drew the learners' attention to the three core components of Spanish requests, namely, addressing the interlocutor, setting-up the stage by providing support/justification,

and the request. Learners were provided with some helpful vocabulary and structures for each of these stages (formal and informal ways of address and the use of direct and indirect ways of making a request). On Day 2, learners completed the story completion task according to their groups' modality. On Day 4, learners repeated the same story completion task with a new, self-selected partner. Immediately following task completion, learners completed the immediate post-tests in the oral modality, followed by the written modality. Fourteen days later, learners completed the delayed post-tests, in the oral and written modalities, respectively. The rationale for the sequencing of the DCTs (oral to written) was that learners were less likely to draw on explicit knowledge and strategies during oral performance given the immediate requirements of oral output, limiting the amount of transfer to the written DCT.

Table 2. Data collection procedure

	Oral modality	Written modality
Day 1	Pre-Test	
	Oral DCT and written DCT	
	Explicit instruction via PPT	
Day 2	Story completion task (Time 1)	
Day 4	Story completion task with a new partner (Time 2)	
	Immediate post-test	
	Oral DCT and written DCT	
Day 14	Delayed post-test	
	Oral DCT and written DCT	

Data coding

Each participant was equipped with an audio-recorder allowing them to individually audio-record their oral DCT responses which were transcribed verbatim. The handwritten DCT responses were typed. Both written and oral DCT responses were coded for (1) request strategy (e.g., direct, indirect), (2) external modifications (e.g., supporting reasons) and (3) internal modifications (modals in present and conditional). We allocated one point for each instance of the head acts and modifications and subsequently calculated the means for each response. Examples (1) to (5) show learners' responses to the request scenario, 'You ask your professor for help in order to study for the final', in the oral modality. In Example (1), the imperative form is used to request help. This request is preceded by a single supporting reason. This example was coded as having 1 direct request and 1 supporting reason. In Example (2), the request for help was mitigated by a modal. This request was preceded by 1 supporting reason.

In Example (3), the modal is expressed in the present tense and in Example (4), the modal is expressed in the conditional. Finally, in Example (5), the learner provided a reason before making the request.

Example 1. Direct head act

    S: Hola profesora, yo necesito *ayudar por la final. Ayudame por favor.
                [Hello teacher, I need help for the final. *Help me* please.]

Example 2. Indirect head acts

    S: Hola Ani, necesito ayuda con mi tarea. ¿Puedes ayudarme por favor?
                [Hello Ani, I need help with my homework. *Can you* help me please?]

Example 3. Modalization – present

    S: Perdón profesora, yo no entiendo la tarea. ¿Puedes ayudarme con mi tarea?
                [Sorry professor, I don't understand the homework. *Can you* help me with the homework?]

Example 4. Modalization – conditional

    S: Podría ayudarme?
                [*Could you* help me?]

Example 5. Supporting reasons

    S: Hola, tengo muchos problemas con mi tarea. ¿Quizas puedes ayudarme con *el?
                [Hello, *I having trouble with my homework*. Maybe you can help me with it?]

Reliability and analysis

Both authors discussed the coding criteria after reviewing students' responses. This stage was followed by individual coding where each author coded 17% of the DCT responses. The percentage agreement was 93.5% for the DCTs. We subsequently met and resolved any differences before the first author coded the remaining responses. Alpha was set at .05 for all statistical tests.

Results

The first research question examined the impact of tasks on the development of L2 pragmatic competence in the form of requests with learners of Spanish. Table 3

provides the raw numbers of the direct and indirect head acts from the responses to the 10 request DCT items (5 oral DCT items and 5 written DCT items), for the two groups combined ($N = 25$), the mean scores, and standard deviation. Table 3 also provides the percentages of direct and indirect head acts at each testing phase.

Table 3. Descriptive results for direct and indirect head acts for OG and WG combined

	Pre-test			Immediate post-test			Delayed post-test		
	N	M	SD	N	M	SD	N	M	SD
Direct	101 (75.4%)	4.04	2.37	72 (38.5%)	2.88	2.21	102 (50.0%)	4.08	2.45
Indirect	33 (24.6%)	1.32	1.95	115 (61.5%)	4.60	2.56	102 (50.0%)	4.08	2.58
Total	134 (100%)			187 (100%)			204 (100%)		

Results of a two-way ANOVA showed a significant effect of instruction on the use of direct head acts, $F(2, 24) = 4.609$, $p < .05$; $\eta^2 = 0.167$, for the two groups combined. The mean of direct head acts dropped from 4.04 at the pre-test stage to 2.88 at the immediate post-test. However, this difference was not maintained on the delayed post-test ($M = 4.08$). A significant effect of instruction was also identified on indirect head acts. Specifically, $F(2, 24) = 34.425$, $p < .05$; $\eta^2 = 599$. The higher use of indirect head acts at the immediate post-test stage was maintained on the delayed post-test.

The first research question also examined whether the use of tasks as a vehicle for pragmatic instruction would impact the use of external modifications (i.e., supporting reasons) as well as the use of internal modifications (i.e., modals) (see Table 4). For external modifications, in the form of supporting reasons, results of a two-way ANOVA showed a significant main effect for treatment, $F(2, 24) = 57.928$, $p < .05$; $\eta^2 = 716$. The use of supporting reasons was maintained on the delayed post-test.

Table 4. Descriptive results for external and internal modifiers for all learners

	Pre-test			Immediate post-test			Delayed post-test		
	N	M	SD	N	M	SD	N	M	SD
Supporting reasons	96	3.8	2.66	235	9.24	2.96	239	9.48	2.88
Modals: Present	22	.84	1.74	117	3.2	2.95	127	4.24	3.39
Modal: Conditional	4	0.16	0.80	38	1.64	2.09	21	1.48	2.41

With respect to syntactic downgraders in the form of present tense modal, we also found a significant main effect for treatment, $F(2, 24) = 20.557$, $p < .05$; $\eta^2 = 472$. Moreover, their use continued to increase between the immediate post-test ($M = 3.2$) and the delayed post-test ($M = 4.24$). Finally, results of the two-way ANOVA also showed a significant main effect for treatment on conditional use of *poder*, $F(2, 24) = .7.504$, $p < .05$; $\eta^2 = 246$. A slight decrease between the immediate and delayed post-tests was identified. In sum, there is some support for the use of tasks with low-inter-mediate learners of Spanish for the development of ILP.

For the second research question, we investigated the potential impact of task modality on the development of L2 pragmatic competence. In our analysis, we compared the learners' performance from the OG and the WG for head acts (direct and indirect), external modifications (supporting reasons), and internal modifications (modals). Table 5 shows the descriptive statistics for each group for the three testing phases.

Table 5. Oral and written modalities on direct and indirect head acts

	OG (N = 13)						WG (N = 12)					
	Pre-test		Immediate post-test		Delayed post-test		Pre-test		Immediate post-test		Delayed post-test	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Direct head act	3.62	2.14	2.54	1.94	3.54	2.29	4.05	2.16	3.25	2.49	4.67	2.57
Indirect head act	1.54	2.29	4.92	1.93	4.23	2.48	1.08	1.56	4.25	3.16	3.92	2.77

Results of a two-way ANOVA did not show a significant difference for direct head acts across the two modalities between the testing phases: Pre-test: $F(2, 1) = .0.863$; $p > .05$; Immediate post-test: $F(2, 1) = .0.432$; $p > .05$; Delayed post-test: $F(2, 1) = .0.258$; $p > .05$. Similar results were identified for indirect head acts such that no significant differences were identified across the modalities: Pre-test: $F(2, 1) = .0.571$; $p > .05$; Immediate post-test: $F(2, 1) = .0.524$; $p > .05$; Delayed post-test: $F(2, 1) = .0.768$; $p > 0.05$.

We further compared the learners' performance from the OG and WG on the use of supporting reasons and syntactic downgraders. Table 6 shows the descriptive statistics for each group on the three testing phases. The raw numbers for each request features, with the exception of conditionals, followed a very similar distribution across the two modalities.

As illustrated in Figure 2, the OG produced 26 conditionals on the immediate post-test where the WG only produced 12. Although the OG experienced a larger drop on the delayed posttest, they continued to outperform the WG.

Table 6. Oral and written modalities on external and internal modifiers

| | OG | | | | | | | | |
| | Pre-test | | | Immediate post-test | | | Delayed post-test | | |
	N	M	SD	N	M	SD	N	M	SD
Supporting reasons	57	4.38	3.02	131	10.07	2.75	129	9.76	3.21
Present	0	0.61	1.19	40	3.15	2.37	53	4.07	3.30
Conditional	1	0.07	0.27	26	2	1.91	14	1.07	1.97
	WG								
	Pre-test			Immediate Post-test			Delayed Post-test		
	N	M	SD	N	M	SD	N	M	SD
Supporting reasons	39	3.16	1.85	104	8.33	3.02	110	9.16	2.58
Present	13	1.08	2.23	39	3.25	3.59	53	4.41	3.63
Conditional	0	0	0	12	1	1.71	7	0.58	1.73
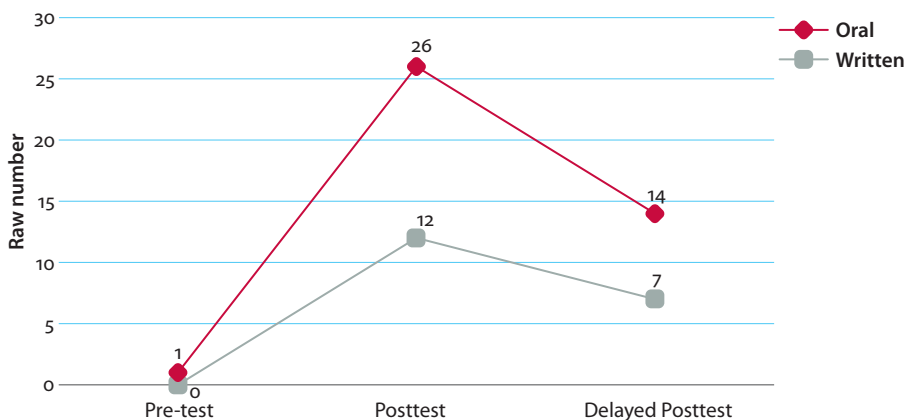
Figure 2. Conditionals across modalities

We conducted a two-way ANOVA for each measure. For supporting reasons, significant differences were not identified across the modalities: Pre-test: $F_{(2, 1)} = .0.262$; $p > .05$; Immediate post-test: $F_{(2, 1)} = .0.145$; $p > .05$; Delayed post-test: $F_{(2, 1)} = .0.613$; $p > .05$. Moreover, we did not find any significant differences across modalities for modals in the present tense: Pre-test: $F_{(2, 1)} = .0.515$; $p > .05$; Immediate post-test: $F_{(2, 1)} = .0.937$; $p > .05$; Delayed post-test: $F_{(2, 1)} = .0.809$; $p > .05$. And finally, despite some differences in raw numbers, no significant differences across the modalities were

found: Pre-test: $F(2, 1) = .0.347$; $p > .05$; Immediate post-test: $F(2, 1) = .0.183$; $p > 0.05$; Delayed post-test: $F(2, 1) = .0.515$; $p > .05$.

In summary, the main findings of the study indicate that TBLT had some impact on the ILP development of indirect request head acts, internal modifications and external modifications. On the other hand, modality did not appear to mediate the development of ILP competence. However, it is important to acknowledge that the statistical analyses were performed on a relatively small sample size. In the following section, we will discuss our findings in relation to previous studies.

Discussion

Building on previous research in ILP and TBLT, the present study examined whether TBLT and task modality affects learning of request-making L2 Spanish. Results indicated that learners' use of request forms improved after receiving a brief explicit demonstration and completing two story completion tasks. Specifically, the number of indirect head acts significantly increased and were sustained on the delayed post-tests. These findings are valuable considering that indirect requests, particularly hearer-oriented requests, are preferred to direct requests in Spanish (Flores Salgado, 2011; Márquez, 2000). In addition, learners increased their production of supporting reasons and were able to continue with this practice on the delayed post-tests. Not only are supporting reasons important for minimizing the imposition of the face-threatening requests (Félix-Brasdefer, 2005), research shows that Spanish speakers tend to provide multiple supporting reasons to mitigate potential threats (Flores Salgado, 2011). Finally, while the use of present tense modals was evident, the findings did not show sustained development of conditional modals.

TBLT researchers provide strong evidence that tasks serve as a vehicle for the development of lexis and grammar (Bygate, 2015; van den Branden, 2016). With respect to the effect of TBLT on ILP development, research suggests that participation in tasks appears to benefit some aspects of learners' development of pragmatic knowledge. However, findings are somewhat inconclusive. For instance, unlike Taguchi and Kim's (2016) findings, in the present study the participants' use of indirect head act strategies was sustained over time. One possible explanation to account for this difference lies in task implementation. In the present study, the participants repeated an identical task on two occasions; however, in Taguchi and Kim (2016), learners repeated the same procedure but worked on new contents. Previous studies on task repetition suggest that repeating identical tasks may push learners to notice the target language (Payant & Reagan, 2018). The idea that rehearsal may be important in the learning of pragmalinguistic forms warrants further attention.

The second research question focused on the role of task modality in ILP development. In this context, modality did not play a role in promoting ILP development as

both groups performed similarly in terms of both head acts and modifications on the DCTs. These findings did not support our initial hypothesis and previous findings that writing would have stronger effects than oral production on output (Gilabert et al., 2016). However, we should point out that existing studies on task modality have generated mixed results. For example, as noted above, Payant and Kim (2015) identified modality effects such that language-related discussions were more frequently observed during the written modality than during the oral modality whereas discussions about the task at hand, namely task management functions, were frequently observed during the oral components of the tasks. Kormos (2014) found impacts on learners' task performance with learners using more varied vocabulary and producing a higher proportion of error-free clauses in the written modality than in the oral modality. However, Kuiken and Vedder (2011) did not identify striking differences between the written and spoken mode in the output of Dutch learners of Italian with regards to complexity, accuracy, and fluency. Specifically, after having completed an oral or a written version of a decision-making task, the only difference between the two groups was the use of dependent clauses, identified in the written group's data. Although the current literature of modality effects is expanding, it would appear as though learner variables and task variables mediate the impact of modality on L2 development.

Seeking to account for the lack of differences between the two groups, we examined the learners' interaction data. Overall, learners from the OG tended to repeat the target structures more frequently than learners from the WG. See Example (6) below. Learner A and Learner B from the OG produced seven conditionals while brainstorming the content of their story. In Turn 6 alone, Learner B repeated the conditional modal three times.

Example 6. OG dyad brainstorming content

1	A:	Podrías… could you take… [*Could you*… could you take]
2	B:	Podrías or we could say could you go to the recycling center so podrías *vas a… [*Could you* or we could say could you go to the recycling center so *could you go to*]
3	A:	Luisa habla Luisa says [*Luisa talks* Luisa says]
4	B:	What did you say podrías? [What did you say *could you*?]
5	A:	Uh hmm
6	B:	Podrías podrías or maybe we could say could we… so podrías *vamos al centro de reciclaje? [*Could you could you* or maybe we could say could we… so *could you* we go to the recycling center]?

The interactions between learners in the WG were qualitatively different. In Example (7), Learners C and D produced only five conditionals in thirteen conversational turns while attempting to formulate a request. However, they are engaged in discussion about how to make the request (PRE) (Turns 2–5) and the optimal language to express it (i.e., LRE) (Turns 6–13).

Example 7. PRE and LRE in WG dyad

1	C:	Podrías could you uh [*Could you* could you uh]
2	D:	Podrías like go with me [*Could you* like go with me]
3	C:	Sí, conmigo [*Yes, with me*]
4	D:	Podrías what's… would we just say podrías [*Could you* what's… would we just say *could you*]…
5	C:	No, that's with me.
6	D:	We just need the 'could you go with me.' So it'd just be…
7	C:	Is it just *va*? [Is it just *go*]?
8	D:	ir? [*to go*]?
9	C:	Oh
10	D:	Cause is podrías a verb? What is that? [Cause is *could you* a verb? What is that?]
11	C:	It's could you.
12	D:	So would it be like unconjugated to go or would it be…?
13	C:	ir conmigo or va conmigo [*to go with me* or *goes with me*]?

While these are selected examples, they tended to mirror the learners' performance across the two groups. In the OG (Example (6)), learners appeared to be focusing on the meaning but in processing what they wish to say, they are repeating the target structure without analyzing it. This rehearsal component may enable learners to internalize the target structures. In the WG (Example (7)), Learners C and D appeared to be focusing more on their grammatical accuracy. For instance, in Turn 5, rather than using the Spanish conditional form, they turn to English to translate the conditional and the target is only expressed once in Spanish, in Turn 10. In the OG, when learners are focusing on the contents, they may in fact be rehearsing target features. In the WG, learners may be processing the structures more deeply and may use declarative knowledge and their first language (L1) to discuss the target structures. Ultimately, learners appear to be in a position to produce the target in similar ways as a result of repetition (oral) or analysis (written). Based on the current findings and post hoc analysis of task performance, it would seem worthwhile to examine the quantity and quality of LREs and PREs produced by each group to examine whether modality impacts task performance as well as the role of the L1.

Since both the OG and WG showed similar gains for the target Spanish requests, we propose the following explanations as to the similar pragmatic gains between the groups. First, both groups received the same brief explicit instruction on the target request speech acts, an important aspect to consider as part of the teaching cycle. Second, the type of pragmatic feature (Spanish request), while different, is not drastically different than the request form in English (participants' familiar language). This cross-linguistic similarity can support the interlanguage development of the target pragmatic form (Rose, 2005). Third, in this study, learners from each modality group repeated the same task twice. Since repetition has been shown to benefit L2 development (Kim

& Tracy-Ventura, 2013), doing the same task twice may have mitigated the gains in pragmatic knowledge. To measure the effect of task repetition on the modality group-ings' gains, it would be important for future research to measure pragmatic production between repetitions.

While the present study offers new insights, some limitations should be taken into account. This classroom-based research, implemented with two groups of intermediate learners of Spanish, drew on a single task type. As such, it provides but a small glimpse into the relationship between task modality and the development of pragmatic com-petence and may not be generalizable to other contexts. In addition, the current study implemented a task design in which the necessary requests were explicitly prompted via the information given on the back of the story completion images. While the pragmatic requests were central to the completion of the overall task and learners were instructed to describe the characters' actions during the picture description stage, the actual dia-logues between characters that prompted authentic requests/responses were provided on the cards and part of the story completion component. This direct approach may not have encouraged authentic language production therefore future research should explore various task designs that prompt authentic pragmatic responses. Also, the pres-ent study examined the development of pragmatic competence via DCTs. It might be beneficial to examine task performance more systematically and more authentically in order to better understand how the participants engaged with the learning material in both modalities. It also remains unclear how learners with more advanced proficiency would have performed on a similar task since pragmatic competence is often acquired at later stages of proficiency. In sum, we believe that the present study offers a potential research avenue in terms of task type, learner proficiency, and interactional features.

Conclusion

The intersection of TBLT and interlanguage pragmatics is becoming an important area of research, as evidenced by this edited volume. While our findings have sug-gested that modality did not mediate acquisition, it seems imperative that research-ers continue to investigate how different variables concerning task modality such as sequencing of modalities, attitudes towards modalities, and different modes within an overarching modality (i.e., collaborative writing task vs text-based writing task within writing as a modality) may affect overall language development. This study contrib-utes to the existing research in that there have been few papers focusing on Spanish L2 pragmatic instruction (Rose, 2005; Takahashi, 2010; Taguchi, 2015). As pragmatic research has been limited in Spanish language instruction classrooms, future research may consider collecting additional data about how task-based pragmatic instruction influences learners' gains as well as learners' perceptions towards the task-based prag-matic instruction.
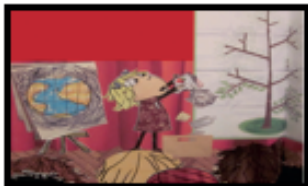
References

Adams, R., & Ross-Feldman, L. (2008). Does writing influence learner attention to form? In D. D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 243–266). Ann Arbor, IN: University of Michigan Press.

Ahmadian, M.J., & Tavakoli, M. (2010). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, 15, 35–59.

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, 15(1), 35–59. doi:10.1177/1362168810383329

Blum-Kulka, S., House, J., & Kasper, G. (1989). *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex.

Bygate, M. (2001). Effects of task repetition on the structure and control of language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Task-based learning: Language teaching, learning, and assessment* (pp. 23–48). London: Longman.

Bygate, M. (Ed.). (2015). *Domains and directions in the development of TBLT: A decade of plenaries from the international conference*. Amsterdam: John Benjamins. doi:10.1075/tblt.8

Child, L., & Hurst, B. (2011). *Look after your planet*. London: Puffin Books.

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

Félix-Brasdefer, J. C. (2005). Indirectness and politeness in Mexican requests. In D. Eddington (Ed.), *Selected proceedings of the 7th Hispanic linguistics symposium* (pp. 66–78). Somereville, MA: Cascadilla Press.

Félix-Brasdefer, J. C. (2008). Pedagogical intervention and the development of pragmatic competence in learning Spanish as a foreign language. *Issues in Applied Linguistics*, 16, 49–84.

Flores Salgado, E. (2011). *The pragmatics of requests and apologies developmental patterns of Mexican students*. Amsterdam: John Benjamins. doi:10.1075/pbns.212

Gass, S., & Mackey, A. (2007). Input, interaction and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 175–200). Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S., & Mackey, A. (2015). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition. An introduction*, (pp. 180–206). London: Routledge.

Gilabert, R., Barón, J., & Llanes, À. (2009). Manipulating cognitive complexity across task types and its impact on learners' interaction during oral performance. *International Review of Applied Linguistics in Language Teaching*, 47(3–4), 367–395. doi:10.1515/iral.2009.016

Gilabert, R., Manchón, R., & Vasylets, O. (2016). Mode in theoretical and empirical TBLT research: Advancing research agendas. *Annual Review of Applied Linguistics* 36 117–135. doi:10.1017/S0267190515000112.

Golato, A. (2003). Studying Compliment Responses: A Comparison of DCTs and Recordings of Naturally Occurring Talk. *Applied Linguistics*, 24(1), 90–121. doi:10.1093/applin/24.1.90

Jeon, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development: A meta-analysis. In J. M. Norris & L. Ortega (Eds), *Synthesizing research on language learning and teaching* (pp. 165–211). Amsterdam: John Benjamins. doi:10.1075/lllt.13.10jeo

Kasper, G. (1996). Introduction: Interlanguage pragmatics in SLA. *Studies in Second Language Acquisition*, 18(02), 145–148. doi:10.1017/S0272263100014856

Kasper, G. (2001). Classroom research on interlanguage pragmatics. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 33–62). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524797.006

Kasper, G., & Rose, K. R. (2002). Pragmatic development in a second language. *Language Learning: A Journal of Research in Language Studies*, 52(1), 1–352.

Kasper, G., & Schmidt, R. (1996). Developmental issues in interlanguage pragmatics. *Studies in Second Language Acquisition*, 18(02), 149–169. doi:10.1017/S0272263100014868

Kim, Y. (2013). Promoting attention to form through task repetition in a Korean EFL Context. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse educational Contexts* (pp. 3–24). Amsterdam: John Benjamins. doi:10.1075/lllt.34.04ch1

Kim, Y., & Payant, C. (2014). A pedagogical proposal for task sequencing: An exploration of task repetition and task complexity on learning opportunities. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 151–177). London: Bloomsbury Academic.

Kim, Y. (2015). The role of tasks as vehicles for learning in classroom interaction. In N. Markee (Ed.), *Handbook of classroom discourse and interaction* (pp. 163–181). Malden, MA: Wiley-Blackwell.

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom contexts: The role of task complexity. *The Modern Language Journal*, 99(4), 656–677. doi:10.1111/modl.12273

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, 41(3), 829–840. doi:10.1016/j.system.2013.08.005

Koike, D. A., & Pearson, L. (2005). The effect of instruction and feedback in the development of pragmatic competence. *System*, 33(3), 481–501. doi:10.1016/j.system.2005.06.008

Kormos, J. (2014). Differences across modality of performance: An investigation of linguistic and discourse complexity in narrative tasks. In H. Byrnes & R. Manchón (Eds.), *Task-based L2 language learning: Insights from and for L2 writing* (pp. 193–216). Amsterdam: John Benjamins.

Kormos, J., & Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning*, 62(2), 439–472. doi:10.1111/j.1467-9922.2012.00695.x

Kuiken, F., & Vedder, I. (2011). Task performance and linguistic performance in L2 writing and speaking: The effect of mode. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 91–104). Amsterdam: John Benjamins. doi:10.1075/tblt.2.09ch4

Langer, B. D. (2013). Teaching requests to L2 learners of Spanish. *Journal of Language Teaching and Research*, 4(6), 1147–1159. doi:10.4304/jltr.4.6.1147-1159

Loewen, S. (2015). *Introduction to instructed second language acquisition*. New York, NY: Routledge.

Lynch, T. & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4(3), 221–250. doi:10.1191/136216800125078

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Márquez, R. R. (2000). *Linguistic politeness in Britain and Uruguay: A contrastive study of requests and apologies*. Amsterdam: John Benjamins. doi:10.1075/pbns.83

Newton, J. (2013) Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research*, 17(2), 164–187. doi:10.1177/1362168812460814

Niu, R. (2009). Effect of task-inherent production modes on EFL learners' focus on form. *Language Awareness*, 18(3–4), 384–402. doi:10.1080/09658410903197256

Payant, C., & Kim, Y. (2015). Language mediation in an L3 classroom: The role of task modalities and task types. *Foreign Language Annals*, 48, 706–729. doi:10.1111/flan.12161

Payant, C., & Reagan, D. (2018). Manipulating task implementation variables with incipient Spanish language learners: A classroom-based study. *Language Teaching Research*, 22(2), 169–188.

Pinter, A. (2005). Task repetition with 10-year-old children. In C. Edwards & J. Willis (Eds.), *Teachers exploring tasks in English language teaching* (pp. 113–126). Houndmills: Palgrave Macmillan.

Pinter, A. (2007a). Some benefits of peer-peer interaction: 10-year-old children practicing with a communication task. *Language Teaching Research*, 11(2), 189–207.
  doi:10.1177/1362168807074604

Pinter, A. (2007b). What children say: Benefits of task repetition. In K. Van den Branden, K. Van Gorp, & M. Verhelst (Eds.), *Tasks in action: Task-based language education from a classroom-based perspective* (pp. 131–158). Newcastle upon Tyne: Cambridge Scholars.

Plonksy, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics* 36 73–97. doi:10.1017/S0267190516000015.

Philp, J., Oliver, R., & Mackey, A. (2006). The impact of planning time on children's task-based interactions. *System* 34 547–565. doi:10.1016/j.system.2006.08.004

Révész, A., Sachs, R., & Mackey, A. (2011). Task complexity, uptake of recasts, and second language development. In P. Robinson (Ed.), *Researching task complexity: Task demands, task-based language learning and performance* (pp. 203–238). Amsterdam: John Benjamins.

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–38). Amsterdam: John Benjamins. doi:10.1075/tblt.2.05ch1

Rose, K. R. (2005). On the effects of instruction in second language pragmatics. *System*, 33(3), 385–399. doi:10.1016/j.system.2005.06.003

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Houndmills: Palgrave Macmillan.

Sample, E., & Michel, M. (2015). An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners' oral task repetition. *TESL Canada Journal*, 31, 23–46.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.) *Handbook of research in second language teaching and learning* (pp. 471–483). Mahwah, NJ: Lawerence Erlbaum Associates.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal*, 82, 320–337.
  doi:10.1111/j.1540-4781.1998.tb01209.x

Sykes, J. (2009). Learner request in Spanish: Examining the potential of multiuser virtual environments for L2 pragmatics acquisition. In L. Lomika & G. Lord (eds.), *The Second Generation: Online collaboration and social networking in CALL*, 2009 (CALICO Monograph). San Marcos: TX, 199–234.

Sykes, J. (2013). Multiuser virtual environments: Learner apologies in Spanish. In N. Taguchi & J. Sykes (eds.), *Technology in interlanguage pragmatics research and teaching*. Amsterdam: John Benjamins, 71–100. doi:10.1075/lllt.36.05syk

Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. State-of-the-art article. *Language Teaching* 48 1–50.
  doi:10.1017/S0261444814000263

Taguchi, N., & Kim, Y. (2016). Collaborative dialogue in learning pragmatics: Pragmatics-related episodes as an opportunity for learning request-making. *Applied Linguistics*, 37, 416–437.

Takahashi, S. (2001). Explicit and implicit instruction of L2 complex request forms. In P. Robinson, M. Sawyer, & S. Ross (Eds.), *Second language acquisition research in Japan* (pp. 73–100). Japan: The Japan Association for Language Teaching.

Takahashi, S. (2010). Assessing learnability in second language pragmatics. In A. Trosborg (Ed.), *Pragmatics across languages and cultures* (pp. 391–421). Berlin: Mouton de Gruyter.

Van den Branden, K. (Ed.) (2006). *Task-based language education: From theory to practice*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667282

Van den Branden, K. (2016). Task-based language teaching. In G. Hall (Ed.), *The Routledge handbook of English language teaching* (pp.238–251). London: Routledge.

Williams, J. (2012). The potential role(s) of writing in second language development. *Journal of Second Language Writing*, 21(4), 321–331. doi:10.1016/j.jslw.2012.09.007

Appendix A.

Images	Target Language
	Request / **Response** Stop producing waste — Negative Play videogames — Negative Reduce waste — Positive
	Request / **Response** Put the paper in — Negative Avoid the use of plastic — Negative Recycle plastic — Positive
	Request / **Response** Pick up the paper — Negative Clean the classroom — Negative Take out the recycling — Positive
	Request / **Response** Develop a program — Negative Protect the environment — Negative Help at school — Positive
	Request / **Response** Separate paper and aluminum — Negative Be outdoors — Positive Plant a tree — Positive
	Request / **Response** Resolve problems — Negative Change habits — Positive Celebrate — Positive

Developing pragmatic competence through tasks in EFL contexts

Does proficiency play a role?

Mayya Levkina
University of Barcelona

Although task-based language teaching, on the one hand, and teaching pragmatics on the other, are not considered to be new trends anymore, there is still a gap in the research regarding how pragmatic features can be taught using a task-based approach. To fill the gap, the present study examines whether speech acts of apology, justification, and thanking are teachable through tasks in English as a Foreign Language (EFL) contexts. The study further investigates the role of L2 proficiency in the development of pragmatic awareness. The participants are 38 third-year university students in Spain who are enrolled in English for the Media course with two different levels of proficiency in English (B1 and C1 on the CEFR). The experiment consists of a pretest involving e-mail writing tasks and a posttest. Pretest/posttest tasks are analysed by rating the appropriateness of speech acts of apology, justification, and thanking on a six-point appropriateness scale and on grammar scales. Results display some positive effects of the task-based instruction on the learning of pragmatics in both proficiency groups. However, the results of the posttests for the grammar scales and the overall gains did not correlate with the proficiency scores, which suggests that L2 development of pragmatics cannot be fully attributed to L2 learners' proficiency.

Introduction

Over the last two decades, various topics on task-based language teaching (TBLT) (e.g., task types, modes, and the role of task complexity and difficulty) have been investigated to advance our knowledge on improving the quality of TBLT. Multiple studies were carried out to investigate second language (L2) oral and written production (Foster & Skehan, 2008; Gilabert, 2005; Gilabert, Manchón, & Vasylets, 2016; Kuiken, Mos, & Vedder, 2005; Kuiken & Vedder, 2007a, 2007b; Michel, Kuiken, & Vedder, 2007; Robinson, 2001). However, there is still a gap in the literature, specifically in how tasks can be used in L2 classroom contexts. A few studies have investigated L2 acquisition based on TBLT (Baralt, Gilabert, & Robinson, 2014), and these studies focused mainly

on such linguistic features as lexis, grammar, and syntax with very little attention to pronunciation or pragmatics. On the other hand, L2 pragmatics research has mainly examined instructional methods and approaches in teaching pragmatics (explicit versus implicit and deductive versus inductive) with very little attention to how pragmatics can be developed using tasks (Taguchi, 2015). So, this study intends to address this gap by examining the role of task-based instruction in teaching pragmatic features with learners from different proficiency levels.

Background

TBLT and pragmatics

In the past several decades, scholars have advocated the importance of task-based instruction that corresponds to the real-life needs of second language learners (Nunan, 2004; Long, 1985), including a few TBLT studies that have been carried out recently with a particular focus on L2 development in different educational settings (Baralt et al., 2014). However, these studies were mainly interested in examining the effects of manipulating task design and implementation factors in L2 learning in situ. So far, most TBLT studies have opted for a focus on grammatical features (Lambert & Robinson, 2014) or vocabulary (Levkina & Gilabert, 2014), and little research has been conducted to examine the learning of pragmatics through task performance. Additionally, TBLT studies have drawn more attention to how tasks might be designed and sequenced. To put tasks into a sequence, most studies were guided by Robinson's SSARC model (2010) where tasks are organised according to their level of cognitive complexity from cognitively simple to cognitively complex. However, less attention has been devoted to input itself (but see Révész, Sachs, & Hama, 2014).

By comparison, in the field of L2 pragmatics, studies that focused on teaching pragmatics were mostly concerned with the relative effects of explicit and implicit instruction, on the one hand (Martínez-Flor & Fukuya, 2005; Takahashi, 2013; Trosborg & Shaw, 2008), and deductive and inductive teaching, on the other (Rose & Ng, 2001; Takimoto, 2008) (for a review, see Takahashi, 2010; Taguchi, 2015). These studies have revealed mixed findings regarding the advantage of one method over the other, where some demonstrated a positive effect of explicit over implicit instruction and deductive over inductive methods across different individual teaching and learner conditions, although other studies showed that implicit and inductive methods are equally efficient (Taguchi, 2015). Inconclusive findings may be due largely to differences in learner characteristics, target pragmatic features, and measures of learning outcomes across studies (Alcón, 2005; Martínez-Flor & Fukuya, 2005; Safont, 2003; Trosborg & Shaw, 2008). Additionally, the mixed findings may be explained in light of the Noticing Hypothesis (Schmidt, 2001) where the learner's attention must be drawn to a linguistic feature so as to notice and intake it.

Regarding teaching pragmatics via TBLT, there is still a gap in both fields, with just a few recent publications addressing the possible interface (Kim & Taguchi, 2015; Kim & Taguchi, 2016). Kim and Taguchi (2015) analysed the effect of task complexity manipulated through the amount of reasoning demands on the development of request-making expressions. 73 Korean high school students took part in the experiment, for which they were assigned to one of the three groups (simple, complex, or control). The experimental design consisted of a pretest, posttest, and delayed posttest. Students' oral performance was measured in terms of LREs, and in addition a discourse completion test was administered to them. The results showed that both experimental groups outperformed the control group in the posttest; however, no differences were observed between the simple and the complex group in the immediate posttest. Cognitive task complexity was relevant for the retention of acquired knowledge, with the students of the complex group maintaining the gains reached after the treatment.

In a more recent publication, Kim and Taguchi (2016) revised the analysis of the data collected with the Korean high-school students with a main focus on learner-learner interaction. The writing tasks containing request-making expressions were analysed for the frequency of pragmatically related episodes (PREs). The results revealed that task complexity had an effect on the PREs related to sociopragmatic aspects rather than pragmalinguistic aspects, which is an important finding for the promotion of sociopragmatic awareness. All in all, task design and task implementation have received little attention in teaching pragmatics, and pragmatics has received little attention in TBLT instruction (Kasper & Rose, 1999, 2002; Rose, 2005; Taguchi, 2015).

Speech acts as a measure of pragmatic knowledge

In pragmatics, researchers very often use speech acts as a target item of analysis; it represents "a minimal unit of linguistic communication" (Searle, 1969, p. 16). In the present research, the main focus is on the speech acts of apology, justification, and thanking. So, in what follows a general overview of those speech acts is given.

Multiple studies in pragmatics have dealt with apology (Bergman & Kasper, 1993; Olshtain & Cohen, 1990; Rose, 1992), justification, and also thanking (Coulmas, 1981; Watts, 2003) among other speech acts. Bergman and Kasper (1993) aimed to see how strategy choices were affected by Thai and American speakers' perceptions of the conversational situation. In their research, assessment questionnaires of contextual factors such as *severity of offence, likelihood of apology acceptance*, and *obligation to apologise* were used. The results showed that in all of the 20 apology situations, the cultural perception of the context differed between Thai and American speakers, where the highest discrepancy was observed in the perception of the *obligation to apologise*.

Olshtain and Cohen (1990) explored teaching the speech act of apology by using a pretest and a posttest questionnaire and three-session materials. A total of 18 advanced EFL learners of English and 11 native speakers of American English took

part in the study. A set of three 20-minute dialogues was used to teach apology strategies. Posttest results were promising as the EFL participants significantly increased their awareness of apology strategies in the contexts where American counterparts would use them.

Several recent studies showed that some additional factors such as gender (Holmes, 1993), socio-cultural context (i.e., social position and roles attributed to it in different cultures; see Bergman & Kasper, 1993; Blum-Kulka, House, & Kasper, 1989; Cohen & Olshtain, 1994; Olshtain & Cohen, 1983, 1990; Rose, 1992; Trosborg, 1995), or variations between different types of English (Blum-Kulka & Olshtein, 1984) should be taken into account as well.

Moving on to thanking, this speech act can be considered a way of showing respect and of being polite (Watts, 2003). According to Watts (2003), thanking in English can be classified as a semi-formulaic form of politeness and respect, whereas Hickey (1991) showed that in Spanish, thanking tends to be used literally. Coulmas (1981) was a pioneer in establishing a link between the speech act of apology and thanking which consists in indebtedness. It was observed that in some languages, like Japanese or Korean, expressions of apology may be used to thank someone (Coulmas, 1981). However, the use of thanking and apology scenarios can also be conditioned by some cross-cultural realisations where an apology may be followed by an explanation and thanking. For instance, English and Italians have been found to use thanking as a way to close a conversation (Aston, 1995; Colston, 2002). However, apart from socio-cultural constraints affecting the use of a speech act, some other factors like L2 proficiency of the students may well affect L2 pragmatic awareness (Taguchi, 2015; Takahashi, 2010; Trosborg, 1995; Xiao, 2015).

Proficiency in TBLT and L2 pragmatics studies

L2 proficiency has been investigated in both TBLT and pragmatics instruction (Rose, 2000; Taguchi, 2007; Malicka & Levkina, 2012). In TBLT, proficiency has been taken into consideration when analysing L2 oral and written production, as well as effects of different types of input and feedback (e.g. recasts) (Housen, Kuiken, & Vedder, 2012; Ortega, 2003; Spada, 1986). The main objective of investigating proficiency within TBLT studies has been to identify the impact of different proficiency levels (beginner to advanced) on task performance and instructional outcomes. In pragmatics, in turn, one of the key questions is whether learners with different levels of proficiency benefit equally from instruction in pragmatics and whether any additional factors come into play.

Takahashi (2005), for example, investigated the role of proficiency and motivation in development of L2 pragmalinguistic awareness. More specifically, she examined whether L2 proficiency or motivation played a role in the noticing of six different types of L2 pragmalinguistic situations under implicit instruction. Eighty participants

completed a motivation questionnaire and a proficiency test first and then took part in a noticing-the-gap activity as the treatment. They were assessed through a retrospective awareness questionnaire. The results showed that learners' motivation was significantly related to the development of the pragmalinguistic awareness but not L2 proficiency.

Taguchi (2011) noted that, although a number of cross-sectional studies have compared pragmatic competence across different proficiency groups, only a few studies have assessed the effect of pragmatic instruction on learners of different proficiency levels. Furthermore, as Roever (2009) suggested, more studies are needed on the learnability of pragmatic features (e.g., speech acts, formulae) at different levels of proficiency. As a result, there is no consensus on whether proficiency plays a role in the learnability of pragmatics. Additionally, Takahashi (2010) in her metalinguistic analysis concluded that teachability and learnability of pragmatic features are closely related to the nature of the intervention and individual characteristics of the learners, such as proficiency and students' motivation. Moreover, Trosborg (1995) in his study on the interlanguage development of pragmatic competence in requesting, complaining, and apologizing showed that students with limited grammar knowledge had a disadvantage in the development of a given speech act.

More recently, Xiao (2015) conducted a synthesis of several cross-sectional studies into the effect of proficiency on the development of pragmatic competence. This study suggests that proficiency does correlate with higher pragmatic competence, although having a high proficiency level does not always guarantee a native-like level of pragmatics, which also depends on social status or power relationship, among others.

Purpose of the study and research questions

To begin to address the gaps in the literature on TBLT and pragmatics, this study investigates whether pragmatics can be taught through instructional tasks with a focus on meaning. In addition, the present experiment addresses the role of proficiency in the development of pragmatic awareness through task performance, as several previous studies have revealed proficiency as a potentially important factor in both pragmatics learning and TBLT (Bardovi-Harlig, 1999; Bardovi-Harlig & Bastos, 2011; Malicka & Levkina, 2012). The following research questions guided the present study:

1. Is there any effect of task-based pragmatic instruction on the use of the speech acts of apology, justification, and thanking in L2 English learners' email writing?
2. How does task-based pragmatic instruction impact the e-mail writing (in terms of pragmatic appropriateness) by students of two different proficiency levels?
3. How does task-based pragmatic instruction impact the e-mail writing (in terms of linguistic appropriateness) by students of two different proficiency levels?

Methodology

Participants

Participants of the study were 66 third-year university students enrolled in a compulsory specific English course (i.e., English for the Media) at the University of Barcelona. The study was conducted during their regular English class. A total of 12 males and 54 females (mean age = 22.54; range: 20–28) were enrolled in the class, but 38 learners' data remained for the final analysis based on the following criteria:

–   students who completed all parts of the experiment (i.e., pretest, a treatment session, and a posttest);
–   students who had B1 or C1 level of English proficiency based on the Common European Framework of Reference (Council of Europe, 2001).

Students' proficiency was measured with the Oxford Placement Test,[1] which was administered by the classroom instructor. Students were divided into two proficiency groups: 18 higher proficiency students (C1 level) (4 males and 14 females with a mean age of 23.42) and 20 lower proficiency students (B1 level) (2 males and 18 females with a mean age of 21.25).

Materials and procedure

Proficiency test
The Oxford Placement Test was administered at the beginning of the course to determine the overall English level of the students whose primary degree was Journalism and Audiovisual Communication.

Test tasks
The data collection consisted of three parts: pretest, treatment, and posttest. The pre and posttests were used to assess students' learning outcomes after the task-based instruction on email writing. There were two versions of the tests that were equivalent in terms of format and task requirements, but were slightly different in terms of content. This decision was made to reduce task repetition effects and to counterbalance the tasks used for the pretest or the posttest. Half of the students were given Version A as pretest and Version B as posttest, while the other students were given Version B as pretest and Version A as posttest (see Appendix A). The test task involved the reply

---

1.   The Oxford Placement Test is a multiple-choice test with 60 items, and it is targeted mainly at lexis and syntax. It distinguishes the following proficiency levels: 0–17 beginner, 18–29 elementary, 30–39 lower intermediate, 40–47 upper intermediate, 48–54 advanced, and 55–60 very advanced.

to a formal e-mail written by someone who was recently interviewed by a newspaper journalist and felt unhappy by the published interview because of a series of mistakes made by the editing board (e.g., place of the article within the issue, omission of some details, no additional copies of the issue sent to the interviewee). The students were not explicitly given any additional instructions regarding the style, content, or format of the email; however, the content of the e-mail was supposed to elicit the target speech acts of apology, justification, and thanking.

Students completed the pretest and posttest tasks in a computer lab on two different days. They received the initial email to their personal email accounts. They read the email and then wrote a reply. They received 15 minutes to complete the test tasks.

Treatment tasks

The treatment consisted of two parts and was entirely delivered in one class session (i.e., 1 hour and 30 minutes). In the first part (see Appendix B) students were given three e-mails – replies to the pre-test e-mail which were written by native speakers of American English and contained different expressions for thanking, apologising, and justifying. First, students were asked just to read them, and afterwards a class discussion on their impression of the style and content took place (implicit part of the treatment). Later, they were asked to fill in a grid with the expressions for thanking, apologizing, and justifying. After completing the task, the teacher provided overall class feedback on the grid completion for both the simple and complex condition. The feedback was controlled for the amount of information given to students in both classes (i.e., the guidelines for the feedback were written beforehand and then given to both teachers to make it as similar as possible).

The second section of the treatment consisted of a series of writing tasks to be completed in pairs and individually. Four tasks were provided in sequence according to the SSARC model of task sequencing, which is based on task complexity levels (Robinson, 2010). In this experiment, reasoning demands were used to manipulate task complexity levels. For example, in the simple version of the task, the students were asked to write an e-email of apology for some presumably insignificant act, so no much reasoning was needed to apologize (see Appendix C). In the most complex version of the task, the students were given an e-mail to answer with compulsory justification, apology, and thanking in a series of very serious mistakes made, which required much more reasoning to come up with some plausible justification. Students were not given individual feedback at this stage.

Data coding and analysis

Two types of measures were used to analyse the learners' pretest and posttest e-mails: appropriateness ratings and a grammar scale. Appropriateness was defined as the

ability to react adequately and appropriately to an oral or written situation on the basis of language knowledge and strategic competence (Bachman & Palmer, 1996). Following several previous studies (Cohen, 1994; North, 2000; Taguchi, 2006, 2007), a six-point rating scale ranging from zero to five was used (see Table 1).

Four native speakers of American English (EFL teachers and / or researchers) evaluated the speech acts of the e-mails based on a six-point appropriateness scale (see Table 1). All raters participated in a norming session, where the author provided detailed information about the rating scale and descriptors, along with practice rating. After the norming session, pretest and posttest emails (38 from each) were randomly distributed to the raters (half of the pre-test and post-test emails). Every e-mail was evaluated by two raters. The interrater reliability correlation was 0.88 for the entire data set. In two cases of a major discrepancy (i.e., a difference of more than 1 point), an average score was used as the final score after discussion (see Appendix D).

Table 1. Appropriateness rating scale for the pragmatic writing task (taken from Taguchi, 2007)

Ratings	Descriptors
5 Excellent	Expressions are fully appropriate. No or almost no grammatical or discourse errors.
4 Good	Expressions are almost appropriate. Very few grammatical and discourse errors.
3 Fair	Expressions are only somewhat appropriate. Grammatical and discourse errors are noticeable, but they do not interfere with appropriateness.
2 Poor	Due to the interference from grammatical and discourse errors, appropriateness is difficult to determine.
1 Very poor	Expressions are very difficult or too little to understand. There is no evidence that the intended speech act is performed.
0	No performance

In addition to appropriateness, grammaticality of emails was assessed on a five-point scale. The scale was developed based on Celaya and Barón (2015). It consisted of five bands, each of which described the grammatical accuracy of speech acts of apology, justification, and thanking (see Table 2). As in the case of the appropriateness rating scale, the grammaticality scale was previously piloted with the same raters and a norming session was conducted. The interrater reliability correlation reached 0.92.

Table 2. Grammar scale (adapted from Celaya & Barón, 2015)

Bands	Descriptors
Band 1	Simple sentences, occasional coordination. Basic modals are used (must, can). Ungrammatical structures are present which impede full understanding of the message.
Band 2	Simple sentences, some attempts to use more complex sentences with coordination. Modals are more frequently used. Some ungrammatical structures are still present.
Band 3	Complex and compound sentences occasionally appear, but not common yet. Formulaic language starts to appear. More modality is used. Ungrammatical structures are less frequent.
Band 4	Grammatical and syntactic complexity is more common, coordination and subordination. Tense, aspect and modality are usually correct. Formulaic language is more frequently used. Infrequent ungrammatical structures.
Band 5	Overall grammatical and syntactic complexity. Tense, aspect and modality correctly used. Formulaic language is widely used. No or almost no ungrammatical structures.

Baseline data were collected from 10 native speakers of American English (3 males and 7 females) which were also analysed using the same classification framework.

To answer the first research question on the possible positive outcomes of teaching pragmatics in a TBLT context, paired-samples t-tests were calculated on the appropriateness scores and the grammar scale scores. Regarding the second and the third research questions on the role of students' proficiency in the development of pragmatic and linguistic awareness, in addition to paired-samples t-tests, a Pearson correlation was calculated to analyse the relationship between proficiency (an independent variable in the present analysis) and gains in pragmatic knowledge after treatment. The Kolmogorov-Smirnov test for normality, as well as tests for skewness and kurtosis, confirmed normal distribution of the data. The significance level was initially set at 0.5. However, because the study used four statistical comparisons, the alpha level was adjusted to 0.01 using the Bonferroni correction.

Results

Research Question 1

The first research question concerned the effectiveness of the use of tasks in teaching pragmatics in the context of a classroom. When comparing the means of the pretest and the posttest on appropriateness scores (apology, thanking, and justification),

certain improvements were observed (see Table 3). The paired-samples t-test revealed a significant difference between the pretest and posttest scores for most comparisons (apology, $t(36) = -9.08$, $p < .001$; thanking, $t(34) = -6.47$, $p < .001$; justification $t(36) = -7.85$, $p < .001$). The results indicate the overall development of pragmatic knowledge in a TBLT lesson context.

Table 3. Appropriateness scores for apology, justification and thanking ($n = 36^2$)

	Speech act	Mean	SD	Min.	Max.
Pre-test	Apology	2.28	.94	1	4
	Thanking	1.03	.62	0	4
	Justification	2.33	1.21	0	4
Post-test	Apology	3.78	.93	2	5
	Thanking	1.89	1.97	0	5
	Justification	3.83	.76	2	5

Research Question 2

To answer the second question, proficiency groups were analysed separately to see if the results would remain at the same level of significance for pragmatic appropriateness. The descriptive statistics (see Table 4) revealed some differences in favour of the posttest scores which were confirmed by the paired-samples t-test, where the higher proficiency group obtained significant differences for apology ($t(14) = -6.14$, $p < .001$); thanking ($t(13) = -3.42$, $p = .003$); and justification ($t(14) = -5.02$, $p < .001$), and the lower proficiency group obtained significant differences for apology ($t(20) = -6.75$, $p < .001$); thanking ($t(18) = -4.32$, $p = .002$), and justification ($t(20) = -6.06$, $p < .001$). Thus, differences between the students' pragmatic knowledge before and after treatment were significant for all participants when analysed jointly, and the same significant differences in the L2 development of pragmatic competence were displayed when splitting the sample into two proficiency groups.

To further analyse the role of L2 proficiency in the development of pragmatic skills a series of Pearson correlations was run (see Table 5). First, the results of the Oxford Placement Test and the pre and posttest scores were compared. The results of the Oxford Placement Test showed an overall significant correlation with pretest apology ($r(38) = .0.65$, $p < .001$); posttest apology ($r(38) = .0.74$, $p < .001$); posttest thanking ($r(38) = .0.45$, $p = .002$); and posttest justification ($r(38) = .0.54$, $p < .001$).

---

2.   Students who obtained a score of 5 in the pre-test were excluded from the analysis.

Table 4. Appropriateness scores for apology, thanking and justification

	Speech act	Mean		SD		Min.		Max.	
		high[a]	low[b]	high	low	high	low	high	Low
Pre – Test	Apology	3.02	1.91	1.12	.68	1	1	4	3
	Thanking	1.22	.53	0.84	.44	0	0	4	1
	Justification	2.25	2.27	1.31	.88	0	1	4	4
Post-test	Apology	4.30	3.36	.90	.79	3	2	5	5
	Thanking	2.23	1.36	2.31	1.59	0	0	5	4
	Justification	4.13	3.55	.73	.80	0	2	5	5

Note.
[a] Group 1 ($n = 16$): higher proficiency students;
[b] Group 2 ($n = 20$): lower proficiency students

Table 5. Correlations between opt and pretest scores, posttest scores and gains

	Pre-test	Post-test	Gains
Apology	.65**	.74**	.34
Thanking	.14	.45**	.05
Justification	.30	.54**	.16

Next, the gains between the pre and the posttest results were computed and the scores were correlated. This analysis was performed to see whether proficiency was also related to the significant intake of new pragmatic information. Although the descriptive statistics revealed some gains in the appropriate use of the speech acts of apology, thanking, and justification (see Table 6), none of the correlations was significant (see Table 5).

Table 6. Gains in speech act scores

	Mean		SD		Min.		Max.	
	high[a]	low[b]	high	low	High	Low	high	low
Apology	1.52	1.32	.86	.95	0	0	3	3
Thanking	1.32	1.41	2.12	2.13	−1	−2	5	5
Justification	1.20	1.50	1.01	1.19	−1	−1	3	4

Note.
[a] Group 1 ($n = 16$): higher proficiency students;
[b] Group 2 ($n = 20$): lower proficiency students

Research Question 3

To answer the third research question on the L2 development of linguistic awareness at different proficiency levels, the results of the grammar scale were analysed. Descriptive statistics displayed a certain improvement in the scores on the posttest (see Table 7). The paired-samples t-test showed a significant difference between the pretest and posttest scores for the grammar scales ($t(35) = -9.03, p < .001$).

Table 7. Gains in grammar scores ($n = 35$)[3]

Grammar score	Mean	SD	Min.	Max.
Pretest	3.35	.75	2	4
Posttest	4.15	.80	3	5

When looking at two proficiency groups separately (see Table 8), the paired-samples t-tests showed that both higher and lower proficiency groups obtained significant gains for the grammar score ($t(15) = -3.22\ p < .001$ for Group 1, and $t(20) = -4,28\ p < .001$ for Group 2, respectively).

Table 8. Gains in grammar scores

Grammar score	Mean		SD		Min.		Max.	
	high[a]	low[b]	high	low	high	low	high	low
Pretest	3.81	3.77	1.17	1.11	3	2	5	5
Posttest	4.25	4.27	.87	.77	3	3	5	5

Note.
[a] Group 1 ($n = 15$): higher proficiency students;
[b] Group 2 ($n = 20$): lower proficiency students

Pearson correlation analyses were run to further explore the relation between L2 proficiency and the development of linguistic awareness. Interestingly, no significant correlation was found between the OPT score and the grammar scale scores for the pretest and posttest (see Table 10), which suggests that the overall level of proficiency as measured by the OPT was not significantly related to the participants' overall performance on the pretest and posttest tasks. Nor was there a significant relation between gains in the posttest and the participants' proficiency level (see Table 9).

---

3.   Students who obtained a score of 5 in the pre-test were excluded from the analysis.

Table 9. Grammar score gains per groups

	Mean		SD		Min.		Max.	
	high[a]	low[b]	high	low	high	low	high	low
Grammar Score	.40	.50	.96	.51	0	0	1	1

Note.
[a] Group 1 (*n* = 15): higher proficiency students;
[b] Group 2 (*n* = 20): lower proficiency students

Table 10. Correlations between OPT and pretest, posttest grammar scale scores and gains

	Pre-test	Post-test	Gains
Grammar Score	.00	.01	.17

Discussion

The results of the present study showed that task-based pragmatic instruction has several positive effects on L2 development as measured by the appropriateness scales and the grammar scale. The findings suggest that the speech acts of apology, thanking, and justification can be taught using e-mail writing tasks. In this context, on the one hand, a combination of explicit and implicit input was employed, which according to Takahashi (2010) is the most efficient way of delivering pragmatic information to ensure intake. On the other hand, a series of real-life tasks were used in the practice part of the treatment, which were organized according to the cognitive criteria from simple to complex based on Robinson's SSARC model.

The successful intake of pragmatic features when comparing the results of the pretest and the posttest may be due to the fact that the participants were provided with both a combination of implicit and explicit input and task-based practice in the same 1.5 hour session. This allowed students to become implicitly aware of the pragmatic structure of the e-mails, to receive explicit feedback, and to put their knowledge into practice. The main point of successful pragmatic development did not lie in the production of more sophisticated linguistic formulae, but rather in the way formal e-mails should be structured and the language itself mitigated.

Similar to what Economidou-Kogetsidis (2011) observed when analysing data gathered from Cypriot students, Spanish participants showed the same tendency of being much more direct than English native speakers (e.g., "I make an apology for your dissatisfaction, but we must cut the interview because there was not enough space to publish all the details"), so it may well be the reason for less frequent use of the speech act of thanking in the data. Moreover, when looking at the L1 Spanish data which was

also gathered from the students, Spanish native speakers, rather than being direct, were more likely to exaggerate when apologizing without providing a justification of what happened (e.g., "We perfectly understand your disappointment, it's a shame that we couldn't place the interview in the first pages so please, receive my apologies"). On the other hand, they would look for a way to compensate a hearer much more eagerly than their English counterparts (e.g., "Due to the issues, we will send you 30 copies tomorrow. Please, accept it"). These cultural differences were reflected in the way Spanish students apologised in English and even after a treatment session, the L1 influence was not mitigated.

A closer look at the descriptive statistics regarding speech acts showed that among NNS students there was very little tendency to use a thanking technique to apologise at the pretest. Even though they received several examples of e-mails with thanking as the first speech act that was used in this type of e-mail, students kept ignoring this particular speech act. This can be explained by the lack of an apology function of thanking described by Coulmas (1981). Spanish NSs use the speech act of thanking to directly thank someone for something, not as a mitigating technique, although it must be admitted that some students after the treatment started using thanking in their e-mails for mitigating purposes (e.g., "I would like to take a moment to thank you for finding time for the interview. It was a real pleasure to work with you."). On the contrary, in order to make their apology more visible, students in the posttest intensified the lexical expression of their apology or even repeated the act of apology several times. That had not been observed in the English NS e-mails (e.g., "I'm sorry to hear about your dissatisfaction… once again accept our apologies").

With regards to proficiency, it has been seen that it correlated with the overall L2 development of pragmatic skills in performing speech acts as measured by the scale of appropriateness. The correlations showed that in the posttest students scored according to their level of proficiency in English. However, this finding is not applicable to the correlational analyses performed on grammar scales and proficiency, where no significant correlation was found. This can be explained by the fact that the main focus of the treatment was on the pragmatic side of e-mail writing, not on the grammar issue of appropriate formulaic language. So, it was to be expected that the results of the Oxford Placement test would not correlate with the grammar score because there was not substantial improvement in grammar scores from the pretest to the posttest as captured by a grammar assessment scale.

Another interesting finding had to do with the relation between proficiency and overall gains. Even though proficiency seemed to correlate with the overall test results for the L2 development of speech acts, there was no significant impact on overall gains for any of the two proficiency groups in any of the tests.

On the other hand, as noted by Bardovi-Harlig (1999), with regards to grammatical and pragmatic competence, if students have an extensive knowledge of advanced grammar rules, that does not imply that they are pragmatically competent.

On the positive side, the mean of the gains for both proficiency groups is promising, which suggests that development of pragmatic knowledge did take place as a result of the treatment. However, the amount of the intake did not depend on the students' level of proficiency. This finding suggests that some other factors such as motivation (as in the study of Takahashi, 2005) may influence the intake by students. It also suggests that pragmatic knowledge is not necessarily related to a high level of proficiency and, therefore, can be taught at lower levels as well. So, pragmatics should be considered as another linguistic skill to teach along with grammar, syntax, lexis, and phonetics.

Gender and age may also have influenced the final results, as they have been shown to play a significant role in pragmatic development and overall L2 acquisition of other language features (King & Holmes, 2014; Lakoff, 1975). However, in the current study, this could not be explored, because the participants were mainly university students with no wide age range; most of them were 21–22 years old at the time of data collection. Moreover, female participants were predominant in both institutional and proficiency groups, which added to the homogeneity of the group.

Limitations and future research

One limitation is the fact that the number of students who took part in the experiment was not large enough to run more complex statistical analyses. The same criticism applies to the number of raters (i.e., four).

Regarding the materials in the study, only formal e-mail writing was used, and the focus lay on only one speech act (apology) which was connected with thanking and justification in this particular context. Future research should compare the teachability of other speech acts across different registers (formal versus informal e-mails) and different addressees (e.g., friends, teachers, co-workers, unknown people). Additionally, in the current study, an American version of English was adopted for the whole study (i.e., input materials, raters). Further research should use other varieties of English (e.g., British English, Australian English). So, a comparative study of the teachability of English pragmatics across cultures would also be of interest.

As for different types of input and practice based on tasks, in this study a combination of implicit and explicit input was selected, followed by one type of task sequencing (from cognitively simple to cognitively complex). For future research, I would suggest testing different combinations of input (explicit versus implicit, deductive versus inductive) and task sequences (e.g., from cognitively complex to cognitively simple, randomized, with the same level of complexity maintained) to explore which combinations are efficient for the development of specific types of pragmatic knowledge. Additionally, only e-mail writing was a focus in the present study, while other tasks such as direction-giving or asking a favor are worth being explored in the future.

In the present study, students with various levels of proficiency were compared. However, only two levels (B1 and C1) could be drawn from the available classroom groups. In future research, students with more different levels of proficiency doing the similar kind of tasks would merit special attention. Age and gender were not the focus of the present study; thus, much more attention should be paid to both of them, as it has been shown that gender affects the way people produce speech acts. Likewise, people of different ages are very likely to have different ways of expressing pragmatic meaning within the same culture.

Finally, this study did not explore individual differences between learners, although these have been shown to be of importance in L2 acquisition (Dörnyei & Skehan, 2003; Skehan, 2008; Winke, 2007). Factors such as attention, motivation, ambiguity tolerance, and working memory should also be taken into consideration in future research on L2 development of pragmatic competence.

Conclusion

The present study aimed to investigate whether pragmatic awareness can be raised through TBLT instruction. The study focused specifically on the speech acts of apology, justification, and thanking as part of e-mail writing. An additional factor analysed in the present research was L2 proficiency of the students. The results showed that students' pragmatic awareness increased and L2 proficiency did not seem to play a significant role. It must be acknowledged, though, that no results of a control group with an alternative teaching approach (or no teaching) were reported in the present study. This should be a focus of future research.

Acknowledgements

References

Alcón, E. (2005). Does instruction work for learning pragmatics in the EFL context? *System* 33(3), 417–435. doi:10.1016/j.system.2005.06.005

Aston, G. (1995). Say 'thank you': some pragmatic constraints in conversational closings. *Applied Linguistics*, 16, 57–86. doi:10.1093/applin/16.1.57

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Baralt, M., Gilabert, R., & Robinson, P. (Eds.) (2014). *Task sequencing and instructed second language learning*. London: Bloomsbury.

Bardovi-Harlig, K. (1999). Exploring the interlanguage of interlanguage pragmatics: a research agenda for acquisitional pragmatics. *Language Learning*, 49, 677–713.
doi:10.1111/0023-8333.00105

Bardovi-Harlig, K., & Bastos, T. (2011). Proficiency, length of stay, and intensity of interaction and the acquisition of conventional expressions in L2 pragmatics. *Intercultural Pragmatics*, 8(3), 347–384. doi:10.1515/iprg.2011.017

Bergman, M. L., & Kasper, G. (1993). Perception and performance in native and nonnative apology. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 82–107). Oxford: Oxford University Press.

Blum-Kulka, S., & Olshtain, E. (1984). Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics*, 5(3), 196–212.
doi:10.1093/applin/5.3.196

Blum-Kulka, S., House, J., & Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex.

Celaya, M. L., & Barón, J. (2015). The interface between grammar and pragmatics in EFL measurement and development. *European Journal of Applied Linguistics*, 3(2), 181–203.

Cohen, A. (1994). *Assessing language ability in the classroom*. Rowley, MS: Newbury House.

Cohen, A., & Olshtain, E. (1994). Researching the production of second-language speech acts. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 143–156). Hillsdale, NJ: Lawrence Erlbaum Associates.

Colston, H. L. (2002). Pragmatic justifications for nonliteral gratitude acknowledgements: Oh sure, any time. *Metaphor and Symbol*, 17, 205–226. doi:10.1207/S15327868MS1703_3

Coulmas, F. (1981). "Poison to your soul": Thanks and apologies contrastively viewed. In F. Coulmas (Ed.), *Conversational routine: Explorations in standardized communication situations and pre-patterned speech* (pp. 69–91). The Hague: Mouton.

Council of Europe E.C. (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Oxford: Blackwell. doi:10.1002/9780470756492.ch18

Economidou-Kogetsidis, M. (2011). "Please answer me as soon as possible": Pragmatic failure in non-native speakers' e-mail requests to faculty. *Journal of Pragmatics*, 43(13), 3193–3215.
doi:10.1016/j.pragma.2011.06.006

Foster, P., & Skehan, P. (2008). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*, 3(3), 215–247.
doi:10.1177/136216889900300303

Gilabert, R. (2005). *Task complexity and L2 oral narrative production*. Unpublished PhD dissertation, University of Barcelona.

Gilabert, R., Manchón, R., & Vasylets, O. (2016). Mode in theoretical and empirical TBLT research: advancing research agendas. *Annual Review of Applied Linguistics*, 36, 117–135.

Hickey, L. (1991). Comparatively polite people in Spain and Britain. *Association for Contemporary Iberian Studies*, 4, 2–6.

Holmes, J. (1993). New Zealand women are good to talk to: An analysis of politeness strategies in interaction. *Journal of Pragmatics*, 20(2), 91–116. doi:10.1016/0378-2166(93)90078-4

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency*. Amsterdam: John Benjamins. doi:10.1075/lllt.32

Kasper, G., & Rose, K. R. (1999). Pragmatics and SLA. *Annual Review of Applied Linguistics*, 19, 81–104. doi:10.1017/S0267190599190056

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

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. doi:10.1111/modl.12273

Kim, Y., & Taguchi, N. (2016). Learner-learner interaction during collaborative pragmatic tasks: The role of cognitive and pragmatic task demands. *Foreign Language Annals*, 49, 42–57.

King, B. W., & Holmes, J. (2014). Gender and pragmatics. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (1–6). Hoboken, NJ: Wiley & Sons. doi:10.1002/9781405198431.wbeal1453

Kuiken, F., Mos, M. & Vedder, I. (2005). Cognitive task complexity and second language written performance. In S. Foster-Cohen, M. P. García-Mayo, & J. Cenoz (Eds.), *EUROSLA Yearbook* (Vol. 5) (pp. 195–222). Amsterdam: John Benjamins.

Kuiken, F., & Vedder, I. (2007a). Cognitive task complexity and linguistics performance in French L2 writing. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp.117–135). Clevedon: Multilingual Matters.

Kuiken, F., & Vedder, I. (2007b). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, 17(1), 48–60. doi:10.1016/j.jslw.2007.08.003

Lakoff, R. (1975). *Language and woman's place*. New York, NY: Harper and Row.

Levkina, M., & Gilabert, R. (2014). Task sequencing in the L2 development of spatial expressions. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 37–71). London: Bloomsbury Academic.

Lambert, C., & Robinson, P. (2014). Learning to perform narrative tasks: A semester-long classroom study of L2 task sequencing effects. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 207–230). London: Bloomsbury Academic.

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language training. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 77–99). Clevedon: Multilingual Matters.

Malicka, A., & Levkina, M. (2012). Measuring task complexity. Does EFL proficiency matter? In A. Shehadeh & C. Coombe (Eds.), *Task-based language teaching in foreign language contexts. Research and implementation* (pp. 43–66). Amsterdam: John Benjamins. doi:10.1075/tblt.4.06mal

Martínez-Flor, A., & Fukuya, Y. J. (2005). The effects of instruction on learner's production of appropriate and accurate suggestions. *System*, 33(3), 463–480. doi:10.1016/j.system.2005.06.007

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics*, 45(3), 241–259. doi:10.1515/iral.2007.011

North, B. (2000). *The development of a common framework scale of language proficiency*. New York, NY: Peter Lang.

Nunan, D. (2004). *Task-based language teaching*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667336

Olshtain, E., & Cohen, A. D. (1983). Apology: A speech-act set. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 18–35). Rowley, MA: Newbury House.

Olshtain, E., & Cohen, A. D. (1990). The learning of complex speech act behavior. *TESL Canada Journal*, 7, 45–65.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518. doi:10.1093/applin/24.4.492

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. doi:10.1111/lang.12061

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. doi:10.1093/applin/22.1.27

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 243–268). Amsterdam: John Benjamins. doi:10.1075/celcr.13.17rob

Roever, C. (2009). Teaching and testing pragmatics. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 560–577). Malden, MA: Wiley-Blackwell. doi:10.1002/9781444315783.ch29

Rose, K. R. (1992). Speech acts and questionnaires: The effect of hearer response. *Journal of Pragmatics*, 17(1), 49–62. doi:10.1016/0378-2166(92)90028-A

Rose, K. R. (2000). An exploratory cross-sectional study of interlanguage pragmatic development. *Studies in Second Language Acquisition*, 22, 27–67. doi:10.1017/S0272263100001029

Rose, K. R. (2005). On the effects of instruction in second language pragmatics. *System*, 33(3), 385–399. doi:10.1016/j.system.2005.06.003

Rose, K. R., & Ng, C. (2001). Inductive and deductive teaching of compliments and compliments responses. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 145–170). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524797.013

Safont, M. P. (2003). Instructional effects on the use of request acts modification devices by EFL learners. In A. Martínerz-Flor, E. Usó Juan, & A. Fernández Guerra (Eds.), *Pragmatic competence and foreign language teaching* (pp. 211–232). Castellón de la Plana, Spain: Publicacions de la Universitat Jaume I.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139173438

Skehan, P. (2008). Individual differences in second language learning. *Studies in Second Language Acquisition*, 13(2), 275–298. doi:10.1017/S0272263100009979

Spada, N. (1986). The interaction between types of contact and types on instruction: Some effects on the second language proficiency of adult learners. *Studies in Second Language Acquisition*, 8(2), 181–199. doi:10.1017/S0272263100006070

Taguchi, N. (2006). Analysis of appropriateness in a speech act of request in L2 English. *Pragmatics*, 16, 513–535. doi:10.1075/prag.16.4.05tag

Taguchi, N. (2007). Task difficulty in oral speech act production. *Applied Linguistics*, 28, 113–135. doi:10.1093/applin/aml051

Taguchi, N. (2011). Teaching pragmatics: Trends and issues. *Annual Review of Applied Linguistics*, 31, 289–310. doi:10.1017/S0267190511000018

Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. State-of-the-art article. *Language Teaching*, 48, 1–50.
doi:10.1017/S0261444814000263

Takahashi, S. (2005). Pragmalinguistic awareness: Is it related to motivation and proficiency? *Applied Linguistics*, 26, 90–120. doi:10.1093/applin/amh040

Takahashi, S. (2010). Assessing learnability in second language pragmatics. In A. Trosborg (Ed.), *Pragmatics across languages and cultures* (Handbooks of pragmatics, Vol. 7, pp. 391–421). Berlin: De Gruyter Mouton.

Takahashi, S. (2013). Awareness and learning in second language pragmatics. *Language, Culture and Communication*, 5, 53–76.

Takimoto, M. (2008). The effect of deductive and inductive instruction on the development of language learners' pragmatic Competence. *The Modern Language Journal*, 92(3), 369–386.
doi:10.1111/j.1540-4781.2008.00752.x

Trosborg, A. (1995). *Interlanguage pragmatics: Requests, complaints and apologies*. Berlin: Mouton de Gruyter. doi:10.1515/9783110885286

Trosborg, A., & Shaw, P. (2008). Deductive and inductive methods in the teaching of business pragmatics: Not an 'either/or'! In R. Geluykens & B. Kraft (Eds.), *Institutional discourse in cross-cultural contexts* (pp. 193–220). Munich: Lincom.

Watts, R. (2003). *Politeness*. Cambridge: Cambridge University Press.
doi:10.1017/CBO9780511615184

Winke, P. M. (2007). The psychology of the language learner: Individual differences in second language acquisition. *Studies in Second Language Acquisition*, 29(1), 143–144.
doi:10.1017/S0272263107310061

Xiao, F. (2015). Proficiency effect on L2 pragmatic competence. *Studies in Second Language Learning and Teaching*, 5(4), 557–581. doi:10.14746/sllt.2015.5.4.3

Appendix A

Dear editor,

I am writing to express my strong dissatisfaction with the interview you published in the last issue of your newspaper. I was very disappointed to find that you omitted most of the details regarding my dog! Furthermore, I should have been informed that you would place the interview at the end of the issue and not at the beginning as promised.

On the other hand, I wonder where I could get 15 copies of this issue as they are already sold out.

I look forward to hearing from you at your earliest convenience.

Yours sincerely,

Simon Trustworthy

Appendix B

Dear Mr. Trustworthy,

Thank you so much for your message and for taking the time to allow our reporter to interview last week. It is a shame that we didn't have enough space to include more details about Rover,

particularly since our reporter said he was indeed a very sweet dog. Unfortunately, as I'm sure you can appreciate, we need to continually make decisions about the length and placement of every article since there is always fresh news to report. Due to the last minute news about the upcoming presidential election, this particular issue needed to be rearranged quite a bit in order to accommodate all of the new information.

Best regards,
Editor

Appendix C

Simple version
You work as a correspondent of an Australian newspaper in Europe. You travel quite a lot around Europe. This time you were asked to go to Preston (England) to interview a local judge. Instead of Preston you went to Bristol. You still have some time to get to Bristol on time. Write an e-mail of apology to the chef editor of your newspaper agency.

Complex version.
You approved the final version of a newspaper issue to be printed. But now you have realized that you got confused with the placement of several headlines. Meanwhile, half of the print run is already out. Write an email of apology justifying your actions to the executive director of the newspaper agency you are working at.

Appendix D

Ratings	Descriptors	Example – act of apology
5 Excellent	Expressions are fully appropriate. No or almost no grammatical or discourse errors.	It is a shame we could not include all the information that was at our disposal. We do apologise for that.(ALGU_07)
4 Good	Expressions are almost appropriate. Very few grammatical and discourse errors.	We should apologise for our mistake. But it was out of our hands. (SAMU_25)
3 Fair	Expressions are only somewhat appropriate. Grammatical and discourse errors are noticeable, but they do not interfere with appropriateness.	I say sorry for publish just something, not everything. In the magazine is not enough space. (GAGA_43)
2 Poor	Due to the interference from grammatical and discourse errors, appropriateness is difficult to determine.	I would apologized for the error, but it wasn't my guilty. (RECA_50)
1 Very poor	Expressions are very difficult or too little to understand. There is no evidence that the intended speech act is performed.	Editor have the decision, so he must sorry for this. (LAMA_13)
0	No performance	

CHAPTER 7

Independently measuring cognitive complexity in task design for interlanguage pragmatics development

Roger Gilabert[1] & Júlia Barón[1,2]
[1]University of Barcelona / [2]International University of Catalonia

This study examines the effects of task design on interlanguage pragmatics, bringing together task complexity and L2 pragmatic outcomes measured holistically. Fifteen expert judges were asked independently to assess task complexity and to evaluate the pragmatic performance of 60 EFL learners who had written a response to four email messages at 4 different levels of complexity. On the basis of needs analysis, complexity had been manipulated by along the following parameters: +/−frequency of input, +/−familiarity with interlocutor, +/−intentional, +/−causal reasoning, +/−dependency of steps, +/−number of elements, and +/−dual task. In order to validate complexity independently, two techniques were applied. On the one hand, a subjective perception questionnaire evaluating mental effort and difficulty was answered by 15 experienced teachers. On the other hand, subjective judgments on a pragmatic scale by the same teachers followed by retrospective protocol analysis were used in the classification of tasks from simple to complex. Pragmatic outcomes were assessed on a holistic rating scale of pragmatic performance. While descriptive statistical results pointed in the direction of predictions, inferential statistics only confirm a difference between the first level of complexity and the other three. As for the effect of sequence, expert judges did not find the outcomes of the two sequences to differ. Results are discussed in light of task performance, task sequencing, and interlanguage pragmatics development.

Introduction

The field of pragmatics is interested in how language is used in context, and one of its many lines of research is interested in how language is used to achieve communicative goals through speech acts. This interest, which was investigated mainly in speakers' and writers' first languages initially, has also become the focus of interlanguage pragmatics (ILP) as well as *acquisitional* pragmatics (henceforth L2 pragmatics). Beyond the initial wave of studies in the 80s and 90s (Faerch & Kasper, 1989; House, 1996;

Kasper & Schmidt, 1996), in recent years we have seen a renewed concern with issues related to L2 language in context (Barron, 2012; Barron, Gu & Steen, 2016). As a consequence of this renewed attention to pragmatics in second language acquisition (SLA), it is not surprising that other research agendas within SLA such as task-based language teaching (TBLT) research have started to delve into in this rich and fruitful area. As Long (2005) has suggested, most if not all of our daily activity can be described in the form of tasks. A lot of those tasks happen interactively and are therefore affected by the context of communication in which they happen. Yet, a considerable proportion of TBLT research has either focused on the performance and development of language ability as measured by general indices like complexity, accuracy and fluency (CAF) or on the development of second language (L2) lexical or morphosyntactic knowledge. Only recently, task-based studies have started to focus on the effects of task design manipulation on pragmatic performance and/or development.

Two important issues that are relevant to the area of pragmatics and have not been addressed within task-based research agendas in relation to pragmatics are, firstly, how to measure task complexity independently (Norris, 2010; Norris & Ortega, 2003; Révész, 2014) and, secondly, how to measure the impact of task sequencing on L2 performance and development (Kim & Taguchi, 2015; Baralt et al., 2014). Regarding the former issue, increases in the cognitive demands that tasks impose on learners through design are expected to affect communicative demands (Robinson, 2011) and as such they are particularly relevant to L2 pragmatics. As experienced by L2 teachers worldwide, more demanding communicative situations (e.g. a face-to-face job interview as opposed to a simple information request over e-mail), which are typically achieved in an instructional context through task design, call for a higher number and also more sophisticated use of pragmatic devices. Consequently, having some degree of control over the cognitive demands of tasks and their perception by task performers becomes crucial. As suggested by a number of scholars (Norris, 2010; Norris & Ortega, 2003; Révész, 2014), however, it is essential that we should operationalize and measure differences in task complexity in an autonomous and independent manner. To our knowledge, this issue has been largely unexplored in task-based studies in general (see however Révész et al. 2016 or Sasayama 2016 for examples of a systematic exploration of the issue) and in task-based interlanguage pragmatics in particular. As far as sequencing is concerned, after three decades of study into task and syllabus design little is known about how tasks should be best sequenced (e.g., from simple to complex, simple-complex-simple). Recent specific proposals (Baralt et al., 2014) have suggested that task complexity could be used as the basis for sequencing decisions with simple tasks performed first, setting the grounds for and facilitating more complex performance. However, the empirical research into the effects of such sequencing is still scarce and scarcer when it comes to effects on L2 pragmatics.

The gap that the study described in this chapter attempts to fill is that of the impact of task design, and more specifically task complexity and task sequencing, on overall L2 pragmatic performance. The study is part of a larger study also covering the effects of task complexity and task sequencing on specific pragmatics moves (Gilabert & Levkina, 2018) which is only mentioned but not fully described here. The goal of this quasi-experimental classroom study is two-fold: (1) to independently assess task difficulty and mental effort involved in the performance of email writing tasks, as perceived by experienced teachers, in order to tap into different levels of task complexity; and (2) to examine whether there are differences between experts' perceptions of pragmatic task performance, as measured by a holistic pragmatic scale, depending on the sequence of presentations of tasks.

In order to meet those two goals, in the following sections we first review the intersection between TBLT and L2 pragmatics in general, and then zoom in on the specific issue of task design for the promotion of interlanguage pragmatics use and development. In the second part of our review, we look at important issues in connection to the independent measurement of task complexity that have only slowly started to make their way into the area of L2 pragmatics. This leads us to a section on task grading and sequencing which provides the theoretical base to our second research goal. Our review ends with a revision of our knowledge on the measurement of interlanguage pragmatics that is the basis for the holistic measurement employed in this study.

Background

TBLT and pragmatics

As mentioned above, among topics related to TBLT, including syllabus design, task development and teacher cognition, researchers have increasingly examined students' task performance and L2 development in terms of fluency, accuracy, and complexity (see Jackson & Suethanapornkul, 2013 for a meta-analysis; Sasayama et al. 2015), and little attention has been paid to pragmatics, pronunciation, or quality of task performance (Plonsky & Kim, 2016), especially in the domain of task complexity research. In fact, TBLT aims at improving students' skills through meaningful and real-life related tasks where social interaction is commonly an important aspect to be considered. As Kasper (2010) claimed, pragmatics is reflected in social interaction, not only in an individual's acts, but also in interactional exchanges. From this view, language development should not only be seen as a cognitive process but also as a social phenomenon (Firth & Wagner 2007). Therefore, pragmatic exchanges could take place and be easily elicited through collaborative tasks, since the tasks involve students' naturalistic interactions.

In line with this idea of examining how social interaction is built, one of the areas on which ILP studies have focused is e-mail communication, since pragmatic aspects such as degree of imposition, social distance, and formality are commonly reflected in such tasks. Apart from this, e-mail has become a common means of communication among people, so it might reflect what users of a language would 'say' when writing e-mails, even if it is an elicited task. Two sets of research studies are particularly useful here to understand the use of speech acts in e-mail communication. On the one hand, in studies looking at speech acts in general, requests have been one of the most often-examined pragmatic moves, due to their face-threatening nature. The studies in ILP have commonly found that the degree of imposition of the speech act, as well as the social distance, have an effect on the way L2 learners perform the speech act. This has been a common finding in studies that have focused on requests, where L2 learners seem to be more direct when the degree of imposition and social distance are low. The use of more indirect requests, on the contrary, tends to be closely related to high degree of imposition and high social distance (Hartfort & Bardovi-Harlig, 1996; Chen, 2001; Biesenbach-Lucas, 2006, 2007; Economidou-Kogetsidis, 2011; Felix-Brasdefer, 2012). Similar findings have been shown in terms of modification of requests. In general, studies have found that non-native speakers (henceforth NNSs) tend to mitigate the request less when both degree of imposition and social distance are low (Woodfield & Economidou-Kogetsidis, 2010; Econimidou-Kogetsidis, 2011; Alcón, 2013, 2015). On the other hand, recently a number of studies have begun to explore how e-mail communication can be taught in the language class (see Econimidou-Kogetsidis, 2015). The studies looking at e-mail communication have mainly analyzed the openings and closings of e-mails, and the speech acts employed by e-mail users.

In classroom contexts, ILP researchers have examined the effects of instruction on pragmatic development (Alcón & Martínez-Flor, 2008; Economidou-Kogetsidis, 2015; Rose & Kasper, 2001; Martínez-Flor & Usó-Juan, 2010; Rose, 2005; Safont, 2005; for a review see Taguchi, 2015). Most of these studies have looked at the effects of both explicit and implicit instruction, and their findings seem to be inconclusive. Although the effectiveness of explicit instruction has been generally confirmed (Takahashi, 2001), a few studies have revealed that implicit instruction can be equally effective (Alcón, 2012; Koike & Pearson, 2005; Takimoto, 2008). Therefore, even though ILP research has revealed the positive effects of instruction, it has not examined which methodological approach is most effective. More importantly, existing instructional studies in ILP are limited in the scope of the instructional methods examined. Most studies focused on explicit and implicit approaches, while other approaches and theories have been rarely explored. Hence, the potential contribution of TBLT to L2 pragmatics is a worthwhile investigation.

As far as instructional task design is concerned, only a few studies have examined the effects of task design along different degrees of task demands on pragmatic

development. One of the first studies was by Taguchi (2007), who examined how tasks designed for the promotion of pragmatic development affected the performance of requests and refusals among L2 English learners. The main variables explored in this study, and later on in Kim and Taguchi (2015, 2016), were social distance, power relationships, and degree of imposition. The findings of these studies suggested that different degrees in social distance, power relationships, and degree of imposition manipulated during task design have an effect on how social variables are projected in interaction. As mentioned above, the same variables were used in Kim and Taguchi (2015) involving L2 English learners. The learners were enrolled in different conditions: complex, simple, and control. The task treatment groups performed drama script tasks focusing on request expressions at two different complexity levels (simple versus complex) in pairs. Task complexity was operationalized along reasoning demands (Robinson, 2011). In the simple version learners had detailed scenario descriptions and matching pictures, whereas in the complex section they did not have any detailed descriptions of the scenario which was meant to have learners in further reasoning. Their findings showed that those in the complex condition produced more pragmatic-related episodes (PREs) than those from the simple groups. Kim and Taguchi (2016) analyzed the interactional features, including PREs and the number of turns produced by students who performed simple and complex tasks. They further included both high imposition (a request for an extra lesson to a busy professor) and low imposition scenarios (asking a friend to show the person some photos during a break), and examined the interaction effects on the amount of interactional features. They found that complexity played a role by generating a higher number of PREs related to sociopragmatics rather than pragmalinguistics. As expected, the high imposition scenarios elicited more interactional features.

In line with these studies, Gilabert and Barón (2013) examined the effect of task complexity in the use of suggestions in two different types of tasks: a problem-solving task and a decision-making task. Task complexity was operationalized along the number of elements and reasoning demands. Results showed that a greater number of pragmatic moves (e.g., requests or suggestions) were triggered in the complex versions of the decision-making task compared with the simple versions, but that was not the case for the problem-solving task. In the latter, the amount of pragmatic moves in the simple and complex versions was not significantly different. This study explores the effects of manipulating task design along the cognitive complexity of tasks on the perception by expert judges. Tasks were designed with the intent to promote the development of interlanguage pragmatics with a focus on pragmatic moves such as requests, apologies, expressions of gratitude and ways of advising others.

The independent measurement of task complexity

Two of the key challenges of the task complexity agenda within TBLT research have been how to operationalize both independent and dependent variables in experimental

studies. This follows calls (Norris, 2010; Norris & Ortega, 2003; Révész, 2014) to obtain validity for independent (e.g., task complexity) and dependent variables (e.g., CAF or pragmatic development measures). More specifically in the case of task complexity, Norris (2010), Norris and Ortega (2003), and Révész (2014) have suggested that independent measurement of cognitive processes is needed in order to avoid circularity. Often task complexity researchers have made predictions about how the manipulation of independent variables (e.g., increasing reasoning demands or number of elements) would affect task performance (e.g., complexity, accuracy, and fluency). If dependent variable measures confirmed the prediction, this was taken as confirmatory evidence of the designed differences, rather than independently measuring task complexity effects on cognitive processes. As Révész puts it, "it needs to be shown rather than assumed that the task version designed to be more complex is indeed more cognitively demanding. Likewise, independent evidence needs to be gathered for the causal processes that are predicted to take place instead of inferring based on linguistic performance data whether they have occurred" (2014, p. 8).

Since the advent of those calls, a number of task complexity studies have strived to autonomously measure the effects of manipulating cognitive demands on the cognitive causal processes. The techniques employed have ranged from more subjective ones such as student ratings of task difficulty (see for example some recent studies by Baralt, 2013; Gilabert et al., 2014; Kim & Tracy-Ventura, 2011; Malicka & Levkina, 2012; Michel, 2011; Révész, 2011; all based on the original first use of subjective ratings by Robinson 2007) or expert judgments of task difficulty and mental effort (examples of recent studies include Révész, 2014; Révész et al., 2016;), to more objective ones such as time estimation (see for example Baralt, 2013; Malicka & Levkina, 2012; Sasayama, 2013, 2016), dual task methodology (Révész et al., 2014; Révész, et al. 2016; Sasayama 2016), or eye-tracking (Michel et al., 2018; Révész et al., 2014; Révész & Gurzynski-Weiss, 2016). In this section, we review the studies that have used expert judgments, which is the technique used in this study.

Expert judgment is a technique by which, for example, testers or teachers are asked to assess the difficulty of an item in a test or the mental effort that task design is meant to exert on learners. In the area of task design, Révész et al. (2014) employed expert judgments to measure the cognitive load generated by tasks varying in complexity by asking two doctoral students to rate tasks using a 5-point Likert scale. They found that the two raters assessed the most complex versions of the task as having a higher cognitive load, which they then additionally triangulated with eye-tracking and dual-task methodology. They used their results to confirm and validate their manipulation of task design and their intended effects, and suggested that expert judgments offer a useful and valid technique for gauging differences in cognitive load. Recently, Révész, Michel, and Gilabert (2016) triangulated subjective ratings with expert judgments and dual-task methodology. As for expert judgments, they used 61 expert teachers

who were asked to evaluate, among other variables, the difficulty and cognitive load of three different task types (a narrative task, a map task, and an instruction-giving task) with two levels of complexity. Their results showed that teachers' perceptions were aligned with their intended design of simple and complex tasks.

In this research project expert judgments area utilized as an independent measurement of task complexity. It is an issue whether such a technique will help to discriminate among the four levels of complexity intended for the tasks in this study, which have been designed to promote the development of interlanguage pragmatics. Task complexity is also the basis for task grading and sequencing, an issue we now turn to.

Task grading and sequencing

The issue of task grading and task sequencing is still an unresolved one in TBLT, and there exists an array of proposals to grade and sequence tasks (Ellis, 2003; Prahbu, 1987; Robinson, 2001, 2011; Robinson & Gilabert, 2007; Skehan, 1998). While the goal of this chapter is not to provide a critical review of the models or hypotheses for task grading and sequencing (see Baralt et al., 2014 for a critical review), we will briefly discuss the criteria they advance for manipulating task design along a continuum of task difficulty/complexity. In the Bangalore project, Prahbu (1987) advanced five criteria for grading tasks, which included the amount of information (from a few elements to many elements), the amount of reasoning (from few steps to many steps), the degree of precision (from no need of precise terms to need of precise terms), the degree of familiarity (from familiar to unfamiliar), and the degree of abstractness (from objects and action to concepts). Skehan (1998) suggested that tasks can be sequenced from easy to difficult along 4 types of variables: (a) code complexity (in terms of linguistic complexity and variety, and vocabulary load and variety); (b) cognitive complexity (in terms of familiarity with the topic, discourse type, and task type) and cognitive processing (organization, amount, clarity, and sufficiency of information); (c) communicative stress (in terms of time pressure, participants, modality, stakes, and opportunity for control); and (d) learners' variables (such as intelligence, breadth of imagination, or experience). Robinson (2001; Robinson & Gilabert, 2007) proposed three sets of factors: cognitive factors (e.g., along resource-directing variables such as amount of reasoning or number of elements and resource-dispersing variables such as planning time or familiarity), interactive factors (i.e., task design factors and interactant factors), and learner factors (i.e., cognitive and affective individual differences factors). Finally, Ellis (2003) brought together ideas from previous models to grade tasks around four criteria: input (including medium, code complexity, and cognitive complexity), conditions (interactant, task demands, discourse mode), processes (cognitive operations), and outcomes (medium, scope, discourse mode).

To the best of our knowledge, from all the models for task grading presented above, the only explicit model for task sequencing available in the literature is the SSARC (stabilize, simplify, automatize, restructure, and complexify) model by Robinson (2005). While all models share the idea that some criteria are necessary for determining, grading, and sequencing tasks, only Robinson's model makes specific predictions about the effects of sequencing on L2 performance and development. Also to our knowledge, there are no studies bringing together task sequencing and L2 pragmatics. In general terms, in reference to the SSARC model, Baralt et al. (2014) state that task *sequencing* should be carried out by designing, and asking learners to perform, tasks that initially are simple on all relevant parameters of task demands, and then gradually augmenting their cognitive complexity on later versions of the tasks. They claim that such sequences may foster *cumulative* learning, since each task version is only slightly different from the previous one, but also includes a slight *increase in the conceptual and communicative challenge.* This has the potential to prompt learners to make adjustments and expand their interlanguage resources to meet those task demands. Such adjustments and expansions, in turn, create the conditions for L2 development. As will be argued below, however, our approach in this study was not a top-down approach by which we took a model and applied it to our task sequencing. Most previous studies have selected task variables (e.g., reasoning demands or number of elements) from existing models to generate task design. One of the risks of such an approach is that task design may not correspond to what actually happens with target tasks (Long, 2005; 2015a). Instead, a bottom-up approach was followed here on the basis of task-based needs analysis (NA). We drew information about the variables that were relevant to each task from domain experts in the field of communication, and the variables were then related to existing models. This decision contributed to enhancing the ecological validity of our task design and rendered the tasks as similar to what communication experts would do in real life.

Measuring pragmatic competence

Testing L2 pragmatic competence is one of the main issues in ILP (for a review, see Taguchi & Roever, 2017). On the one hand, the current practice of using researcher-created tests is a problem in the field, since learners' pragmatic language use is not examined in natural interactions but by using artificial tests such as oral/written discourse completion tests or role-plays. On the other hand, the analysis of such elicited productions is another issue to consider: how should we evaluate the learners' productions? Should we follow native speaker (NSs) norms as the target? In fact, this has been the general trend in ILP, since the NSs' language becomes the goal in second language acquisition. However, this might be problematic when dealing with pragmatics, since aspects such as sociolinguistic variability or identity might interfere in conceptualizing

the native speakers' pragmatic norms (Kasper, 1998). Despite these problems, in ILP, comparison with NS data is what has been mainly used to evaluate L2 learners' pragmatic competence.

However, what has been even more controversial is the use of rubrics in ILP testing. Studies such as those by Liu (2006) and Walters (2007) have developed rubrics in which NSs have rated learners' productions. In such studies, both analytic (e.g., Hudson el al., 1995; Walters, 2007) and holistic pragmatic rating scales have been used (Liu, 2006). As Liu (2011) claims, existing studies have typically followed Hudson et al.'s criteria to evaluate pragmatic competence (rated from 'very unsatisfactory' to 'completely appropriate'). Those criteria include appropriateness, directness, politeness, and formality speech act performance, use of formulaic language, and amount of information provided. However, as Liu (2011) highlights in her review of L2 pragmatics testing, such rubrics need further validation because only few studies have addressed this issue and findings have been inconclusive. Liu (2001) also points out some questions which are left unanswered, such as the type of rubrics that should be used (holistic or analytic) or the background of the raters, since raters' different social backgrounds, like social status or age, might affect the raters' judgments. Furthermore, an extra complication may arise when raters are not NSs of the language, especially in foreign language contexts, in which raters are commonly teachers with different L1 linguistic, social, and cultural backgrounds. In such a context, the question is which norms to use when assessing L2 pragmatic competence and how to articulate those norms clearly in rating rubrics.

This study is a response to the need for further research on the effects of task design on pragmatic knowledge, performance, and development. Firstly, this study moves away from previous research that has often focused on performance in order to focus on overall pragmatic development. As will be further detailed below, the study uses rubrics to assess holistically pragmatic performance. Secondly, the study also addresses calls in the literature for the independent measurement of task complexity, now seen as a fundamental prerequisite for any type of research into task complexity. This is particularly important in an area such as L2 pragmatics, in which researchers have only started to address this issue. Thirdly, task variables for grading and sequencing, which typically have come from existing taxonomies in theoretical models, are drawn from needs analysis in this study. These are used as the basis for instructional task design that is manipulated in terms of cognitive complexity and independently assessed in terms of pragmatic criteria by expert judges.

Research questions

With our research goals in mind, the present study set out to answer the following research questions:

1. Do the different levels of task complexity generated during task design match expert teachers' perceptions of task difficulty and cognitive load?
2. What is the impact of the sequencing of tasks (i.e., simple-to-complex condition versus randomized condition) on the overall pragmatic performance of L2 learners as holistically assessed by teachers?

Because of the scarcity of empirical studies in this area, we could not formulate any specific directional hypotheses for any of the two questions. We predicted, however, that for research question 1, expert judgments would be a good way to discriminate among different levels of task complexity. The prediction was that experts would perceive the simplest task as easier and requiring less mental effort than the other more complex versions. For research question 2, we predicted that task sequencing from simple-to-complex, as predicted by the SSARC model (2005; Baralt et al. 2014), would facilitate all aspects of performance and thus have positive effects for overall pragmatic performance.

Methods

Participants

All 15 teachers who participated in the current study had extensive experience in the teaching of English as a foreign language, and they were all familiar with task-based language teaching. They came from three different institutions in the Barcelona region. Except for one, they were all non-native speakers of English, but they all reported a high command of English. Their ages ranged from 23 to 40 and the average age was 31.5. In order to answer research question 1, all teachers rated the level of difficulty they perceived and the mental effort they invested to perform the task for all tasks. For research question 2, the 15 teachers were randomly assigned to three groups. All teachers in all groups were given 20 e-mail samples written by students and rated them using a holistic scale. They rated the performances of a total of 60 upper-intermediate students of L2 English from two universities in Spain who participated in the current study. These students were selected from four intact groups, all of which were classified by their institutions as B2. Their ages ranged from 19 to 22 and they had all taken 12 years of obligatory courses in English as a foreign language (EFL). Typically, EFL courses in their teaching context are heavily focused on grammar aspects and they neglect conversational aspects of the language. Furthermore, there is little or no focus on pragmatics. At the time of data collection, the students were taking specialized courses in their different areas, and one of the commonly used tasks was an e-mail writing task, which is further described below.

Tasks

Four e-mail writing tasks that varied in complexity levels were used in the entire project: three training tasks integrated into full units, and one final and most complex outcome task measuring participants' overall pragmatic performance (see Appendix A for a sample simple and complex task). In this paper, we report on the perception of difficulty and mental effort (research question 1) for each of the four versions of the tasks by expert teachers. We also report on the most complex outcome task (Task 4) for the assessment of holistic pragmatic performance (research question 2), since this is the task that was performed last and should show the impact (if any) of the sequencing treatment (see Figure 1 below). The design of the tasks was based on the samples of e-mail exchanges collected during needs analysis (Gilabert, 2005). The first three tasks used in the whole experiment were three e-mail messages students were prompted to write, and they were integrated into full units including pre-task, task, and post-task. These served as training for the performance of a final task (which in the overall experiment served as the post-test, i.e., outcome task).

In pedagogic terms, the three training task units followed the same structure (Figure 1) and the same procedure was followed. The e-mail tasks were contextualized in a fictional press office. The students were asked to assume the role of a press officer throughout the experiment. In the pre-task, students were presented with sample e-mail messages containing requests, apologies, thanking, and advice-giving expressions. They were first asked general questions about each email, including the purpose of the message, the relationship between the writer and the recipient, the appropriateness of the opening and closing expressions, and the overall tone of the message. Then they were asked to identify target speech acts in the sample e-mail messages and classify them into four categories (requesting, apologizing, thanking, and advice-giving). During the task phase, they were asked to respond to one e-mail message in pairs, which they presented to the class later. Then for each unit (see Figure 1), students were asked to prepare another email at home individually (which was the focus of this study). All tasks included a post-task with a focus on language aspects (e.g., expressions, prepositions, among others) related to the pre-task and task-cycle.

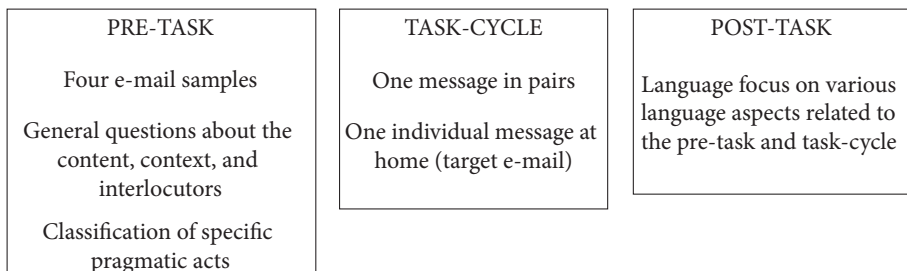| PRE-TASK	TASK-CYCLE	POST-TASK
Four e-mail samples General questions about the content, context, and interlocutors Classification of specific pragmatic acts	One message in pairs One individual message at home (target e-mail)	Language focus on various language aspects related to the pre-task and task-cycle

Figure 1. Unit design for each of the three training tasks

Based on Bardovi-Harlig's (2013) criteria, the tasks used in this experiment were neither authentic nor consequential. They were non-interactive, single-user written tasks, which allowed for comparison across writers. As described in the following section, each task was designed to achieve a distinct level of complexity.

Expert judgments via online questionnaires

In order to answer research question 1, the 15 judges (non-native English speaking instructors) were presented with different versions of the e-mail writing task (see next section for descriptions of different tasks). The judges were asked to rate each task design for the perceived degree of difficulty and the mental effort on a 9 point Likert-scale. They saw an image of each task, which included an e-mail message to respond to and associated instructions, and they were asked to carefully assess its difficulty and the mental effort required to complete the task. They were also asked to provide comments regarding their decisions by responding to an open-ended question. This final open-ended question was meant to examine how different teachers perceived task difficulty and mental effort differently or similarly.

For research question 2, the 15 judges were organized into three groups of 5 raters each (groups were labelled as A, B, C). Each rater received 20 randomized email messages written by students and were asked to evaluate them for overall pragmatic appropriateness on a 7-point Likert scale, ranging from '0' (pragmatically inappropriate) to '6' (fully pragmatically appropriate) (see Appendix A). Each scale category was described in detail, taking into account the following aspects: the adequacy of the e-mail opening, the acknowledgment of the addressee, the structure of the e-mail, formality, appropriateness of speech act, and the use of mitigation. These scale descriptions were based on the Common European Framework of Reference (CEFR, see Council of Europe, 2001) that addresses pragmatics as part of communicative competence. In fact, the CEFR considers discourse factors as an important aspect to consider when dealing with pragmatics. Hence, we considered context, situation, speech acts, appropriateness, and politeness as the key aspects when assessing L2 learners' pragmatic competence.

Operationalization of task complexity and task sequencing

As argued by Gilabert (2005), needs analysis (NA) provides not only an opportunity to define in detail all tasks, processes, and the language associated with them, but it also allows researchers to obtain information about the task components that contribute to their relative complexity. In order to determine the variables affecting task complexity, we followed a bottom-up approach to task design rather than a top-down approach. This means that the criteria used to establish the different levels of complexity were drawn from the information obtained from real e-mail writing samples during NA, rather than deciding on the variables on the basis of pre-existing models of task design,

complexity, grading, and sequencing (Ellis, 2003; Prahbu, 1987; Robinson, 2001, 2011; Robinson & Gilabert, 2007; Skehan, 1998). The criteria that were used to configure the different levels of task complexity are listed below, and we relate them to task design models that have provided suggestions for the variables that affect task performance:

Input frequency

This refers to the relative frequency of the target pragmatic expressions that were used in our tasks. Expressions were selected on the basis of their frequency in the British National Corpus (BNC) but also on the corpus of e-mail messages that we collected during NA. The four levels of complexity ranged from highly frequent expressions (e.g., 'thanks for') to low frequency ones (e.g., 'how could we possibly thank you for'). This variable relates to the concept of 'code complexity' in Skehan's (1998) model 'input' in Ellis (2003).

Familiarity with the interlocutor: this is connected to the idea of 'status' in pragmatics which suggests that part of our pragmatic competence is to adapt our discourse to the different interlocutors we speak to, who may have equal, lower or higher status than us. The four complexity levels in our task design went from high familiarity (e.g., an employee writing to an employee whom he/she knows well) to low familiarity (e.g., writing to an unfamiliar interlocutor with a higher status). This operationalization relates to the +/−familiar one in Robinsons' (2001; Robinson & Gilabert, 2007) Triadic Componential Framework. Increases in complexity along this variable raises attentional and memory demands on the learner.

Causal reasoning

According to the expert sources participating in our needs analysis, e-mails can range from very simple ones that entail the transmission of simple information to highly complex ones involving causes, consequences, and effects. In our task design we used increasing reasoning which ranged from a simple request of information to the complex reasoning of solving a problem with many edges. Simple tasks also used simple independent instructions (e.g., to apologize for a late response, to request information, to thank their interlocutor for their collaboration) or highly complex and interconnected instructions (e.g., where some actions depend on other actions and are interrelated). In Robinson's Triadic Componential Framework (2011; Robinson & Gilabert, 2007), this is related to +/−causal reasoning, while it is also related to 'amount of reasoning' in Prahbu (1987), to 'amount of computation' in Skehan's (1998) model, and 'processes' in Ellis (2003).

Intentional reasoning

Expert sources involved in the NA reported that intentional reasoning may contribute to the complexity of the task. Working out what one person is trying to achieve during communication may be completely straightforward or it may entail a more

complex interpretation of intentions. In this experimental design, the simple versions of the task involved straightforward messages while the most complex tasks pushed the learner to interpret more complex intentions. Notwithstanding the specific operationalization used here, this may be related to Robinson's (Robinson & Gilabert, 2007) 'intentional reasoning' task variable, or to 'clarity of information' in Skehan's (1998), or 'degree of abstractness' in Prahbu (1987).

Number of elements

The number of elements has received considerable research attention in task complexity studies. In our design, the number of elements was operationalized by manipulating the number of steps, so that in the simple versions of task learners would have a few instructions to cope with while complex versions of tasks included more instructions. This can be related to Skehan's (1998) model that suggests that task difficulty may increase along 'the amount of information', as Ellis (2003) also suggests.

As can be seen, the analysis of e-mails and interviews with experts during NA pointed out the multiplicity of factors affecting the complexity of an e-mail message. Most task complexity studies have traditionally focused on one or two variables, but here we took a bottom approach in order to achieve and preserve ecological validity. By considering the factors described above, we were able to create four different tasks that differed in the levels of task complexity (see Figure 1 – see also Appendices A and B for the simplest and most complex version of tasks).

Operationalization of task sequencing

In order to present the tasks with varying degrees of cognitive complexity, we followed Robinson's (2015) SSARC model (see also Baralt, Gilabert, & Robinson, 2014). This model suggests that tasks in a language program should be presented to learners in increasing complexity in terms of the demands that they impose on attention and memory resources, rather than in terms of linguistic difficulty. This means that the basic criteria for task organization should involve intrinsic task complexity variables, such as the cognitive variables that were introduced in the Triadic Componential Framework (Robinson, 2001, 2011; Robinson & Gilabert, 2007). We claim that intrinsic cognitive variables can distinguish one task from another regardless of interactional conditions or individual learner characteristics. This is important because in a typical language course instructors do not have information about individual students' characteristics or their preferred interactional conditions readily available to them. Hence, it is reasonable to use cognitive variables (as identified during NA) to determine the sequencing of tasks in a syllabus, and we followed this rationale in this study. The SSARC model hypothesizes that by providing simple versions of the task earlier, instructors can promote learners' better performance on more complex tasks.

So far, findings on regarding this hypothesis are mixed, with some studies showing positive effects for simple-to-complex sequencing (Levkina & Gilabert, 2014), while other studies showed no effects (Baralt, 2014). Our study intended to test this hypothesis in the area of pragmatics.

Two conditions were created for this experiment: the simple-to-complex condition and the randomized condition. In the former condition, learners completed four tasks in a sequence. They performed the simple version of the task first and then proceeded to the 'less simple' version, to the complex version, and finally to the highly complex version. In the randomized condition, learners performed the complex version first, followed by the simplest version, then by the 'less simple version' and finally by the most complex version.

Condition 1: simple to complex sequence

SIMPLE TASK 1	LESS SIMPLE TASK 2	COMPLEX TASK 3	HIGHLY COMPLEX TASK 4
High frequency input High familiarity Low causal reasoning Low intentional reasoning Very few instructions	Less frequent input familiarity Moderate causal reasoning Less explicit message Few instructions	Infrequent input Partial lack of familiarity High causal reasoning Less explicit message Many instructions	Highly infrequent input Lack of familiarity Very high causal reasoning Implicit message A lot of instructions

Condition 2: randomized sequence

COMPLEX TASK 1	SIMPLE TASK 2	LESS SIMPLE TASK 3	HIGHLY COMPLEX TASK 4
Infrequent input Partial lack of familiarity High causal reasoning Less explicit message Many instructions	High frequency input High familiarity Low causal reasoning Low intentional reasoning Very few instructions	Less frequent input Less familiarity Moderate causal reasoning Less explicit message Few instructions	Highly infrequent input Lack of familiarity Very high causal reasoning Implicit message A lot of instructions

Figure 2. Task sequence conditions

Specific data collection procedures

This quasi-experimental study was fully integrated in the existing language programs of both institutions over one semester. Students were randomly assigned to one of the following task sequencing conditions, resulting in 30 students being in the simple-to-complex condition and 30 in the randomized condition. In the first phase, data were collected about the participants' receptive and productive pragmatic knowledge of target pragmatic acts (not reported in this study; see Gilabert & Levkina, 2018). Then the three training tasks (see Figure 1) were distributed throughout the term. Learners wrote two e-mail messages as part of each task (two messages, one in class in pairs and one individually at home for Task 1; two messages for Task 2; two messages for Task 3) and a final message in the outcome task (Task 4) at the end of the term.

Statistical tools and procedures

IBM SPSS statistics 24 was used for the calculation of the obtained descriptive statistics, correlations, and inferential statistics. In order to calculate interrater reliability among the different raters, intraclass correlation coefficient (ICC) tests were calculated for the 15 raters involved in research question 1 and the three rater groups involved in research question 2.

Regarding the first research question that looked at the different levels of task complexity, Shapiro-Wilk normality tests indicated that the data were not normally distributed for some of the variables and suggested the need to use non-parametric tests. Overall differences were therefore tested by means Friedman's 2-way ANOVA for related samples, which was followed by Wilcoxon Signed Rank comparisons to detect any significant differences between task complexity conditions.

As for the second research question, Shapiro-Wilk normality tests showed that only some variables were normally distributed, so non-parametric tests were chosen for the analysis. To explore the impact of the sequence (single-to-complex versus randomized) on overall pragmatic performance, the Kruskal-Wallis test was used.

Results

Research Question 1

The first research question in this study asked whether the perception of difficulty and mental effort by expert teachers would actually discriminate among the different levels of task complexity designed by the researchers. Intraclass correlation coefficients used for interrater agreement reached ICC was .896, which is considered high, and so it was decided to keep all answers from all 15 judges. Descriptive statistics in Table 1 showed that the tasks were perceived by judges (non-native speaker teachers) in accordance with the difficulty levels as intended by task design. The simple task was perceived as less difficult than the less simple, and the less simple task in turn was perceived as less difficult than the complex task, which was perceived as less difficult than the highly complex task.

Table 1. Descriptive statistics of difficulty ratings

	N	Minimum	Maximum	Mean	Std. deviation
Simple	15	2	5	2.73	1.10
Less simple	15	2	8	5.73	1.71
Complex	15	4	9	6.07	1.44
Highly complex	15	5	8	6.33	1.11

Line graphs show how the perception of tasks was aligned with the intended levels of task complexity.
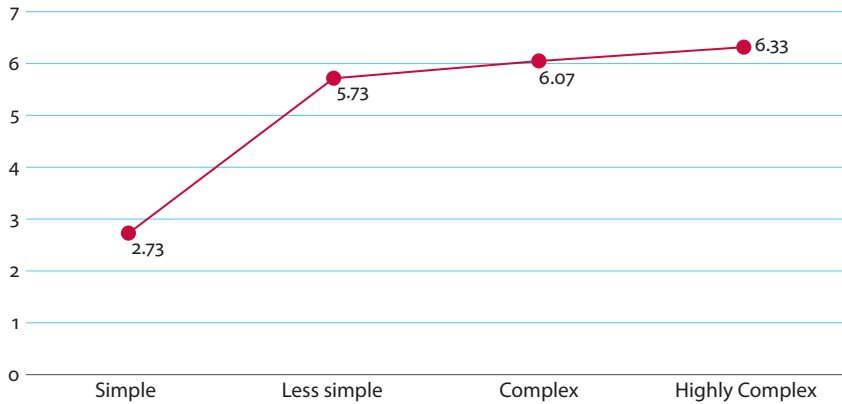
Figure 3. Line graph of perceived difficulty

For the items about perceived degree of mental effort, interrater agreement was .696, and it is considered acceptable. Descriptive statistics showed again that the tasks were perceived by judges as requiring mental effort in line with task design. The simple task was perceived as requiring little mental effort, and the perception of mental effort increased steadily to the highly complex task, which was perceived to require the highest level of mental effort.

Table 2. Descriptive statistics of mental effort

	N	Minimum	Maximum	Mean	Std. deviation
Simple	15	1	7	3.13	1.69
Less simple	15	2	9	6.07	1.94
Complex	15	4	9	6.27	1.62
Highly complex	15	4	8	6.40	1.40

Line graphs again show that the intended increases in task complexity were also perceived by the judges.

In order to check if the differences between the levels were significant, Friedman's tests were run followed by pairwise comparisons between different tasks. For the perceived level of difficulty, Friedman's 2-way ANOVA for related samples revealed significant differences between the levels of task complexity ($p < .000$). Pairwise comparisons showed that the difference was only significant ($p < .000$) between the simple

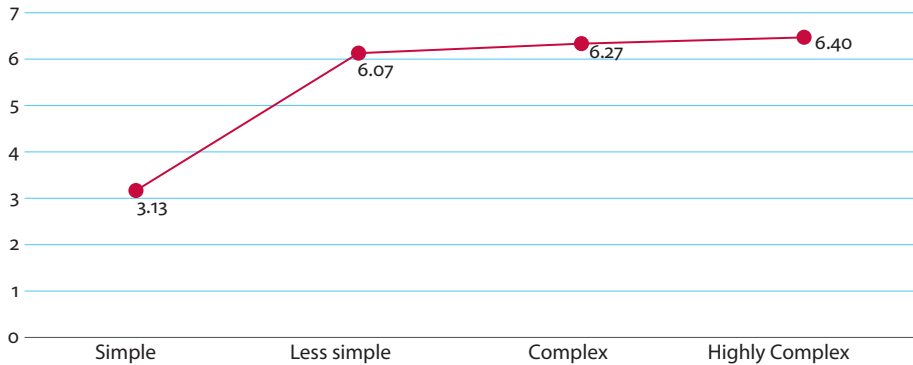Figure 4. Line graph of mental effort

task and other three tasks, but there were no significant differences among the less simple, complex, and highly complex tasks. These findings therefore suggest that the intended levels of task complexity were not sufficiently discriminatory across tasks to yield significant differences in judges' perceived difficulty. Exactly the same patterns were found for the perceived degrees of mental effort. In sum, both difficulty and mental effort ratings by judges revealed the existence of two clearly distinct levels of task complexity (i.e., between the simplest task and the other three), rather than four.

In addition, correlational analyses were conducted between difficulty and mental effort ratings. The correlations between these two constructs ranged from moderate to high: in the simple task, the correlation was high, $r = .731$, $p = .002$; in the less simple task the correlation was also high, $r = .861$, $p = .000$; in the complex task the correlation was again high, $r = .777$, $p = .001$; in the highly complex task the correlation was moderate, $r = .547$, $p = .035$. These findings suggest that raters may have interpreted difficulty and mental effort across task versions in a similar manner. Hence, in future research, only one measurement (difficulty or mental effort) may be necessary.

Research Question 2

In order to answer research question 2, we first needed to check the interrater reliability of expert judges in the rating of overall pragmatic performance of all participants. Firstly, intraclass correlations coefficients were employed for each of the three groups of raters, each one formed by 5 raters (Group A: .697, Group B: .890, and Group C: .031). As a consequence it was decided to leave out the data rated by Group C. This allowed us to calculate the average of ratings of all raters and compare their assessment of the written outcomes in the simple-to-complex sequence to the ones in the randomized one.

In order to measure differences in pragmatic performance, Kruskal-Wallis tests with sequencing as a grouping variable revealed that there were no differences between the performance of learners in the simple-to-complex sequence and the randomized sequence. All judges but one (rater 3 who rated the outcome task in the simple-to-complex sequence as higher) holistically assessed overall pragmatic performance in the outcome tasks (Task 4) to be similar regardless of sequence. Performing tasks with increasing cognitive complexity (i.e., simple, less simple, complex, highly complex) as predicted by the SSARC model did not result in an advantage for the performance on the most complex version of the task. Learners that went through the randomized sequence performed similarly in terms of overall pragmatics than their simple-to-complex counterparts.

Discussion

This study set out to investigate two questions: first, it asked whether expert judgments by experienced teachers would serve as a good technique to discriminate among different levels of task complexity; secondly, it asked whether a simple-to-complex sequence would facilitate overall pragmatic performance to a higher degree than a randomized sequence. In what follows we discuss our findings and relate them to the available literature.

Regarding research question 1, descriptive statistics showed that our operationalization of task complexity in four tasks was actually matched by the perception of teachers in terms of the raw rank order of difficulty and mental effort ratings. The simple task was clearly perceived as the easiest of all tasks, and the most complex task was viewed as the most difficult one, though with smaller differences among the less simple, the complex, and the highly complex task. This is in line with previous studies (Révész et al., 2014; Révész et al., 2016) that have shown teachers to be sensitive to interpreting design decisions. However, design decisions were not radical enough to guarantee a larger difference in perception between the less simple task, the complex task, and the highly complex task. Open comments by teachers reveal interpretations which are in line with our intended design. This, we believe, is a desirable and important goal to achieve prior to task implementation in order to then make any claims about how task design may affect performance. Several comments referred to the *number of elements* in the tasks as contributing to its difficulty:

– T1: "The amount of information"
– T4: "The points to include in the emails are quite a few for intermediate students"
– T6: "Some were difficult due to the amount of requests/refusals they were expected to make. Others were difficult due to the amount of information having to be conveyed in pragmatic ways";

– T8: "More mentally challenging: the task is very long and the instructions can-
not be followed point by point, but require comprehensive reworking keeping all
information in mind."

The *amount of intentional reasoning* imposed by the tasks was also captured by teach-
ers' comments:

– T7: "More difficult: when explicit instructions are given regarding pragmatics
(e.g., be polite, be formal).

Familiarity of the interlocutor was also mentioned as a source of difficulty by
expert teachers:

– T8: "Task 2 is the easiest one, because the message is quite simple and he knows
the recipient, so he can use informal language and doesn't need to worry about
formalities
– T12: "mentally effortful changes depending on who you are writing too you have
to think about how it would be received so that makes it more difficult if you know
or don't know the person").

No comments were made about *causal reasoning* (i.e., the fact that in the more com-
plex task we built in more interrelated causes and consequences) or the *frequency
of the input* in the prompts (i.e., more frequent pragmatic input in simple tasks like
"thanks" versus less frequent input "we are grateful to you") contributing to perceived
difficulty. Because of the bottom-up approach we adopted in our design, in which we
manipulated several variables simultaneously instead of isolating one single variable
for the study, it is impossible to know which one contributed more significantly to the
higher cognitive complexity of tasks. Teachers' comments, however, seem to suggest
that the *number of elements, intentional reasoning*, and *familiarity with the interlocutor*
mainly contributed to their perception of difficulty.

In addition to the small differences between levels two to four suggested by
descriptive statistics, inferential statistics, did not confirm the 4-level sequencing
intended in our design. While there were significant differences between the simple
task and the three other tasks, there were no significant differences among the three
other levels. This would suggest that teachers perceived two main levels of complexity,
rather than four. Several studies (Sasayama, 2013; Révész et al., 2016) have reported
similar difficulties in discriminating levels: only extreme levels seem to yield signifi-
cant differences, with intermediate levels of complexity being a lot more similar and
therefore a lot more difficult to distinguish during task design. Our design may have
suffered from the same problems of lack of radical distinction. Task designers have
no reference in the literature as to what contributes to a real difference in the level of
complexity for any of the task variables that may be manipulated during task design.

At which point do differences in frequency of the input start discriminating between two levels of complexity? How much design must be added (in terms of conditions or elements to deal with) so that the reasoning caused by a task increases significantly from one version of a task to the next? No answers to these questions are available yet. Another question that may be raised whether with a larger number of judges the differences between levels would have turned out to be significant. Our data also brought up the issue of overlap between the constructs of difficulty and mental effort. As suggested by Révész et al. (2016), mental effort more clearly connects with the construct of cognitive load in cognitive psychology, which seems to make it more appropriate in the context of task design and the study of task complexity. Indeed, some experts made attempts to distinguish between the two concepts:

–   T1: "In my opinion, the mental effort depends on the number of different subtasks to do, such as answering many questions or writing them in order, and the difficulty depends on the type of message you have to write, to whom you are writing to, the subject you are dealing with, etc."
–   T7: "Mental effort, for me, depended on the stakes involved."
–   T14: "To me, both are interrelated. The tasks require more mental effort as the more difficult they get."

As for research question 2, there were no significant differences between learners in terms of their pragmatic performance in the outcome tasks regardless of the sequence they were assigned to (see Table 3). Our prediction was based on the general idea advanced by the SSARC model that by providing simple versions of tasks first which are followed by increasingly more complex versions, L2 task performance can lead learners to better schedule their memory and attentional resources to solve task-related problems.

Firstly, it would be reasonable to believe that by dealing with more manageable input at the same time as dealing with simple mental operations (few instructions, low in causal and attentional reasoning, and with high familiarity) would let learners better integrate information and devote memory and attentional resources to solving the task. By slowly increasing task complexity, students would progressively make more efficient use of their resources to meet task demands in the most complex version of the task. It is a well-known fact in SLA that learners prioritize meaning over form. When learners are faced with a new task, a lot of their attentional and memory resources are employed in making sense of the task, in focusing on content and task structure, and in working out what the learner thinks how he or she is expected to perform the task. By progressively building schema in the process of performing simple to more complex tasks (Sweller, Van Merriënboer, & Paas, 1998), we expected their pragmatic knowledge to be slowly but effectively stretched for the students to better meet the high complexity version of the task. By comparison, we expected that students in the

Table 3. Descriptive statistics of raters assessing outcome tasks in Sequences 1 (simple-to-complex) and Sequence 2 (randomized)

		N	Mean	Std. deviation	Minimum	Maximum
Rater 1	1	10	3.70	1.25	2.0	6.0
	2	10	3.80	1.32	2.0	6.0
	Total	**20**	**3.75**	**1.25**	**2.0**	**6.0**
Rater 2	1	10	3.80	1.32	2.0	6.0
	2	10	4.00	1.33	2.0	6.0
	Total	**20**	**3.90**	**1.29**	**2.0**	**6.0**
Rater 3	1	10	4.50	.85	3.0	6.0
	2	10	2.00	.67	1.0	3.0
	Total	**20**	**3.25**	**1.48**	**1.0**	**6.0**
Rater 4	1	10	4.20	.92	3.0	6.0
	2	10	4.80	.63	4.0	6.0
	Total	**20**	**4.50**	**.82**	**3.0**	**6.0**
Rater 5	1	10	3.90	1.29	2.0	6.0
	2	10	4.30	1.42	2.0	6.0
	Total	**20**	**4.10**	**1.33**	**2.0**	**6.0**
Rater 6	1	10	3.80	1.32	2.0	6.0
	2	10	3.60	1.26	2.0	6.0
	Total	**20**	**3.70**	**1.26**	**2.0**	**6.0**
Rater 7	1	10	3.20	1.39	1.0	5.0
	2	10	2.50	.84	1.0	4.0
	Total	**20**	**2.85**	**1.18**	**1.0**	**5.0**
Rater 8	1	10	5.00	1.05	3.0	6.0
	2	10	4.70	1.15	3.0	6.0
	Total	**20**	**4.85**	**1.09**	**3.0**	**6.0**
Rater 9	1	10	3.80	1.13	2.0	5.0
	2	10	3.80	1.03	3.0	6.0
	Total	**20**	**3.80**	**1.06**	**2.0**	**6.0**
Rater 10	1	10	3.50	1.08	2.0	5.0
	2	10	3.20	.92	2.0	5.0
	Total	**20**	**3.35**	**.99**	**2.0**	**5.0**

randomized sequence would disperse their resources over the different components of the task and, as a consequence, not strategically allocate their resources in subsequent versions of the task. We did not expect the creation, construction, and automation to be so efficient when tasks of higher task complexity were presented first. However, this was not confirmed by our data. L2 learners performed equally in the final outcome task regardless of the sequence in which they had been prepared. A sequence of four tasks may not be enough to capture any substantial changes to emerge. This is in line with previous studies (Baralt, 2014) which found a minor effect of sequencing along task cognitive complexity on L2 performance. Baralt stated (2014) that: "as opposed to a specific task sequence, those sequences that contained more complex tasks" (p. 114) had a higher impact on performance and development (in her case, it was a higher number of language-related episodes). At least as it has been operationalized in this study, and in general in the sequencing literature so far, sequencing from simple-to-complex may not be the only way to impact performance and development. Finally, and as rightly pointed out by one of the reviewers, time distribution may have played a role. A relatively long lag between the sessions may have mitigated or even eliminated any effects of task sequencing. Future research should address whether potential sequences effects differ depending on whether we present tasks in an intensive, concomitant fashion or in a more widely distributed fashion.

Conclusion

There are a number of limitations to this study. Firstly, other techniques for the independent measurement of task complexity should have been used. Subjective ratings by the learners themselves could be collected right after task performance. This was not possible in this study but future studies should definitely use triangulation of sources. Secondly, in retrospect, fewer levels of task complexity could have been created. We could have worked, for instance, with only one intermediate level between the simple and the highly complex version of the task. In our view, the measurement of pragmatic knowledge will need further research attention in the future, and it will be need to be combined with the measurement of other linguistic dimensions, such as CAF measures. In line with this, and as pointed out by Liu (2011) more studies dealing with how to test pragmatic competence through holistic scales need to be carried out. More methodological guidelines on how to create such scales are required so replication studies can take place without each new study having to create scales *ad hoc*. This would not only benefit the research agenda but would also have important pedagogical implications in the teaching and testing of L2 pragmatics.

Beyond limitations, we believe a number of conclusions can also be drawn from our study. If we believe that real target tasks, which are often complex and demanding,

need to be broken down into smaller, more manageable pedagogic versions than can be practiced in class (Long, 2015a), then getting task design right is of utmost importance. When it comes to task design as associated with pragmatics, the expert judgements technique seemed to work satisfactorily as an independent measurement of task complexity. The fact that their impressions matched our design intentions, even when our design did not radically distinguish among the three higher levels of complexity, is an indication that the technique has some potential. Additionally, asking teachers to rate task complexity is a ready available technique that does not require expensive equipment (e.g. eye-trackers or reaction time software). It therefore seems reasonable to ask expert instructors to report on how they perceive instructional materials. They provide interesting and valuable insights that can help task and syllabus designers with their design decisions. If teachers happen to be the ones who design the tasks, it is an accessible, uncostly and relatively quick option to show the tasks to other teachers to confirm whether design is actually matched by their colleagues' subjective perception.

A second conclusion that can be drawn from this study is that the different versions of the task need to be more radical in their design. The fact that teachers perceived two levels instead of four levels of complexity in this study is proof of that. It seems imperative than in order to distinguish different levels of task complexity more of a radical increase between tasks should be applied during design. In TBLT we do not have specific guidelines as to how this may be done since task sequencing is largely an unresolved issue in SLA. Despite some recent proposals (Baralt et al., 2014) exact guidelines for sequencing still remain elusive and certainly a lot more in-depth theoretical and empirical work is needed in this area. In the third place, we were hesitant about the use of pragmatic rating rubrics and we still feel they need specific validation work. We need to be able to assess tasks performance holistically from a pragmatics point of view, and this study has been a step towards that goal.

References

Alcón, E. (2012). Teachability and bilingualism in effects on third language learners' pragmatic knowledge. *Intercultural Pragmatics*, 9(4), 511–541.

Alcón, E. (2013). Mitigating e-mail requests in teenagers' first and second language academic cyber-consultation. *Multilingua*, 32, 779–799.

Alcón-Soler, E. (2015). Pragmatic learning and study abroad: Effects of instruction and length of stay. *System*, 48, 62–74.

Alcón, E., & Martínez-Flor, A. (Eds.). (2008). *Investigating pragmatics in foreign language learning, teaching and testing*. Clevedon: Multilingual Matters.

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(4), 689–725. doi:10.1017/S0272263113000429

Baralt, M. (2014). Task complexity and task sequencing in traditional versus online language classes. In M. Baralt, R. Gilabert, & P. Robinson (Eds), *Task sequencing and second language learning* (pp. 52–122) . New York: Bloomsbury.

Baralt, M., Gilabert, R., & Robinson, P. (Eds.). (2014). *Task sequencing and instructed second language learning*. London: Bloomsbury.

Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning*, 63(1), 68–86. doi:10.1111/j.1467-9922.2012.00738.x

Bardovi-Harlig, K., & Hartford, B. S. (2005). Institutional discourse and interlanguage pragmatics. In K. Bardovi-Harlig & B. S. Hartford (Eds.), *Interlanguage pragmatics: Exploring institutional talk* (pp. 7–36). Mahwah, NJ: Erlbaum.

Barron, A. (2012). Interlanguage pragmatics: From use to acquisition to second language pedagogy. *Language Teaching*, 45(1), 44–63. doi:10.1017/S0261444811000462

Barron, A., Gu, Y., & Steen, G. (Eds.). (2016). *Routledge handbook of pragmatics*. London: Routledge.

Biesenbach-Lucas, S. (2006). Making requests in email: Do cyber-consultations entail directness? Toward conventions in a new medium. *Pragmatics and Language Learning*, 11, 82–107.

Biesenbach-Lucas, S. (2007). Students writing emails to faculty: An examination of e-politeness among native and non-native speakers of English. *Language Learning and Technology*, 11(2), 59–81.

Chen, C-F. E. (2001). *Making e-mail requests to professors: Taiwanese vs. American students*. Paper presented at the Annual Meeting of the American Association for Applied Linguistics in St. Louis, February 2001. (ERIC Document Reproduction Service No. ED 461 299).

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Press Syndicate of the University of Cambridge.

Economidou-Kogetsidis, M. (2011). "Please answer me as soon as possible": Pragmatic failure in non-native speakers'e-mail requests to faculty. *Journal of Pragmatics*, 43(13), 3193–3215.

Economidou-kogetsidis, M. (2015). Teaching E-mail Politeness in the EFL/ESL Classroom. *ELT journal*, 69(4), 415–424. doi:10.1093/elt/ccv031.

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

Færch, C., & Kasper, G. (1989). Internal and external modification in interlanguage request realization. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics* (pp. 221–247). Norwood, NJ: Ablex.

Félix-Brasdefer, J. C. (2012). E-mail requests to faculty. In M. Economidou-Kogetsidis, & H. Woodfield (Eds), *Interlanguage request modification* (pp. 217–287). Amsterdam: John Benjamins.

Firth, A., & Wagner, J. (2007). Second/Foreign language learning as a social accomplishment: elaborations on a reconceptualized SLA. *The Modern Language Journal*, 91(1), 800–819. doi:10.1111/j.1540-4781.2007.00670.x

Gilabert, R. (2005). Evaluating the use of multiple sources and methods in needs analysis: A case study of journalists in the autonomous community of Catalonia (Spain). In M. H. Long (Ed.), *Second language needs analysis* (pp. 182–99). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667299.007

Gilabert, R., & Barón, J. (2013). The impact of increasing task complexity on L2 pragmatic moves. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse educational contexts* (pp. 45–69). Amsterdam: John Benjamins. doi:10.1075/lllt.34.06ch3

Gilabert, R., Barón, J., & Levkina, M. (2011). Manipulating task complexity across task types and modes. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 105–140). Amsterdam: John Benjamins.

Gilabert, R., & Levkina, M. (2018). The impact of task complexity and task sequencing on the development of pragmatic moves. Manuscript in preparation.

House, J. (1996). Developing pragmatic fluency in English as a foreign language: routines and metapragmatic awareness. *Studies in Second Language Acquisition*, 18(2), 225–252. doi:10.1017/S0272263100014893

Hudson, T., Detmer, E., & Brown, J. D. (1995). *Developing prototypic measures of cross-cultural pragmatics*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawaii at Manoa.

Jackson, D., & Suethanapornkul, S. (2013). The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning*, 63(2), 330–367.

Kasper, G. (1998). Interlanguage pragmatics. In H. Byrnes (Ed.), *Learning Foreign and Second Languages: Perspectives in Research and Scholarship* (pp. 183–208). New York, NY: The Modern Language Association of America.

Kasper, G. (2010). Interlanguage pragmatics. *Variation and Change: Pragmatic Perspectives*, 6, 141–154.

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

Kasper, G., & Schmidt, R. (1996). Developmental issues in interlanguage pragmatics. *Studies in Second Language Acquisition*, 18, 149–69. doi:10.1017/S0272263100014868

Kasper, G., & Blum-Kulka, S. (1993). Interlanguage pragmatics: An introduction. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 3–17). Oxford: Oxford University Press.

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom contexts: The role of task complexity. *The Modern Language Journal*, 99(4), 656–677. doi:10.1111/modl.12273

Kim, Y., & Taguchi, N. (2016). Learner–learner interaction during collaborative pragmatic tasks: The role of cognitive and pragmatic task demands. *Foreign Language Annals*, 49(1), 42–57.

Kim, Y., & Tracy-Ventura, N. (2011). Task complexity, language anxiety, and the development of the simple past. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 287–306). Amsterdam: John Benjamins. doi:10.1075/tblt.2.18ch11

Koike, D. A., & Pearson, L. (2005). The effect of instruction and feedback in the development of pragmatic competence. *System*, 33(3), 481–501. doi:10.1016/j.system.2005.06.008

Levkina, M., & Gilabert, R. (2014). Task sequencing in the L2 development of spatial expressions. In M. Baralt, R. Gilabert, & P. Robinson (Eds), *Task sequencing and instructed second language learning* (pp. 37–71). New York: Bloomsbury.

Liu, J. (2011). Testing interlanguage pragmatics knowledge. In Trosborg, A. (Ed.), *Pragmatics across language and cultures* (pp. 467–488). Berlin: Walter de Gruyter.

Liu, J. (2006). Assessing EFL learners' interlanguage pragmatic knowledge: Implications for testers and teachers. *Reflections on English Language Teaching*, 5(1), 1–22.

Long, M. H. (2005). *Second language needs analysis*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667299

Long, M. (2015). *Second language acquisition and task-based language teaching*. Malden, MA: Wiley-Blackwell.

Malika, A., & Levkina, M. (2012). Measuring task complexity: Does EFL proficiency matter? In A. Shebadeh & C. Combee (Eds.), *Task-based language teaching in foreign language contexts* (pp.43–66). Amsterdam: John Benjamins. doi:10.1075/tblt.4.06mal

Martínez-Flor, A., & Usó-Juan, E. (2010). Pragmatics and speech act performance. In A. Marínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical and methodological issues* (pp. 3–20). Amsterdam: John Benjamins. doi:10.1075/lllt.26.01mar

Michel, M.C. (2011). Effects of task complexity and interaction on L2-performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 141–174). Amsterdam/Philadelphia: John Benjamins.

Michel, M., Gilabert, R., & Révész, A. (2018). Eye-tracking and stimulated recall as means of investigating cognitive load and processes during L2 tasks. Manuscript in preparation.

Norris, J. M. (2010). *Understanding instructed SLA: Constructs, contexts, and consequences.* Plenary address delivered at the annual conference of the European Second Language Association (EUROSLA), Reggio Emilia, Italy.

Norris, J. M., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 716–761). Malden, MA: Blackwell. doi:10.1002/9780470756492.ch21

Plonsky, L., & Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics*, 36, 73–97.

Prabhu, N. S. (1987). *Second language pedagogy.* Oxford: Oxford University Press.

Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics*, 35(1), 87–92. doi:10.1093/applin/amt039

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, 95(1), 162–181. doi:10.1111/j.1540-4781.2011.01241.x

Révész, A., & Gurzynski-Weiss, L. (2016). Teachers' perspectives on second language task difficulty: insights from think-alouds and eye tracking. *Annual Review of Applied Linguistics* 36 182–204. doi:10.1017/S0267190515000124

Révész, A., Michel, M. C., & Gilabert, R. (2016). Measuring cognitive task demands using dual task methodology, subjective self-ratings, and expert judgments: a validation study. *Studies in Second Language Acquisition*, 38(4), 703–737. doi:10.1017/S0272263115000339

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.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. doi:10.1093/applin/22.1.27

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, 43(1), 1–32. doi:10.1515/iral.2005.43.1.1

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics*, 45, 195–215.

Robinson, P. (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance.* Amsterdam: John Benjamins. doi:10.1075/tblt.2

Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *IRAL-International Review of Applied Linguistics in Language Teaching*, 45(3), 161–176. doi:10.1515/iral.2007.007

Rose, K. (2001). An exploratory cross-sectional study of interlanguage pragmatic development. *Studies in Second Language Acquisition*, 22(1), 27–67.

Safont, P. (2005). Third language learners: Pragmatic production and awareness (Vol. 12). Clevedon, England: Multilingual Matters.

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

Sasayama, S. (2013). Is a "complex" task really complex? Measuring task complexity independently from linguistic production . Paper presented at the 5th Biennial International Conference on Task-Based Language Teaching, Banff, Alberta, Canada.

Sasayama, S., Malicka, A., & Norris, J. (2015). Primary challenges in cognitive task complexity research: Results of a comprehensive research synthesis. Paper presented at the 2015 Task-Based Language Teaching Conference in Leuven, Belgium.

Sasayama, S. (2016). Is a 'complex' task really complex? Validating the assumption of cognitive task complexity. *Modern Language Journal*, 100(1), 231–254. doi:10.1111/modl.12313.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–295 doi:10.1023/A:1022193728205

Taguchi, N. (2007). Task difficulty in oral speech act production. *Applied linguistics*, 28(1), 113–135. doi:10.1093/applin/aml051

Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. State-of-the-art article. *Language Teaching*, 48(1), 1–50. doi:10.1017/S0261444814000263

Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford: Oxford University Press.

Takahashi, S. (2001). The role of input enhancement in developing pragmatic competence. *Pragmatics in Language Teaching*, 171–199.

Takimoto, M. (2008). The effects of deductive and inductive instruction on the development of language learners' pragmatic competence. *The Modern Language Journal*, 92(3), 369–386. doi:10.1111/j.1540-4781.2008.00752.x

Walters, F. S. (2007). A conversation-analytic hermeneutic rating protocol to assess L2 oral pragmatic competence. *Language Testing*, 24(2), 155–183. doi:10.1177/0265532207076362

Woodfield, H., & Economidou-Kogetsidis, M. (2010). "I just need more time": A study of native and non-native students' requests to faculty for late submission. *Multilingua*, 1, 77–118.

Appendix A

Example of the <u>simplest</u> version the e-mail tasks used in the experiment with high frequency input, high familiarity/low social distance, low causal reasoning, low intentional reasoning, and a small number of elements.

> You're Pere Fort and you're organizing a Conference on Visual Arts in Barcelona. The press officer at the Guggenheim museum in New York is writing to you about the conference.
> You have known her for several years and even met with her a couple of times:
>
> Hola Pere,
> Here's the attachment with the slides (in JPG format) that you asked me for. Please remember not to publish them until I give you the go-ahead. If you do, they'll fire me (although I sometimes wonder if that may actually not be so bad !).
> Hugs,
> Susan
> P.S; How are the kids doing? And Pat? And you? How about Roibos?
>
> She forgot to include the attachment with the slides. Write back to her and:
>
> – Thank her for her quick response to your request.
> – Ask her to send you the image files again and also confirm that you'll not be publishing them.
> – Say you're sorry you forgot to tell her that Roibos (your parakeet), whom she loved so much, has passed away. He could not take the cold winter.
> – Advise her to stick to her job until the crisis is over.

Appendix B

Example of the most complex version the e-mail tasks used in the experiment low frequency input, low familiarity/high social distance, high causal reasoning, high intentional reasoning, and a high number of elements.

> You're Pere Fort and you're organizing a Conference on Visual Arts in Barcelona. You have just received the following e-mail from Mr. Goodenough, an eccentric diplomat and an important speaker at the conference you're organizing. You've been e-mailing back and forth and things are getting complicated. You have a difficult job to do so be ready for it !
> Mr. Fort,
>
> This is simply unacceptable. Why is it that my 1-hour plenary talk
> on the 27th of June? I told you that I have bustling agenda
> and that I can only get to the conference on the 28th. Will
> Dr. Alltrouble be attending after all? You know what the consequences
> may be if she is.

(Continued)

Appendix B.   (*Continued*)

> Besides that, you have said nothing about my request for special
> vegetarian food for Chopper, my Rottweiler whom I need by my side as
> I give my talk (I feel uneasy and terribly lonesomeif he's not around)
> and you know loneliness is a tricky business.
>
> Thank God you managed to get me a room at the Sheraton, which I hope
> will compensate for all these grievances!
>
> Goodenough
>   Sir Anthony Goodenough
>   Cultural Aggregate to the South Amerindian Embassy
>   PS: by the way, is it cold in Barcelona? Should I bring warm clothes?
>   Write back and:
>
> – Thank Mr. Goodenough for his quick response.
> – You need to tell him that it will not be a plenary talk (he misunderstood you in a previous message) but just a 20-minute talk in a panel with other participants.
> – Inform him that not only is Mrs. Alltrouble,his eternal foe and rival, coming but he'll also be in the same panel as him. Find a satisfactory explanation and ask him to accept the change.
> – It's impossible to change the date of his talk because there are other people scheduled for the 28th. Politely ask him to change his plans.
> – Dogs (or any other kind of pet) are not allowed in the conference room. Politely ask him not to bring his dog or advise him to find an alternative solution
> – The reservation at the Sheraton has been cancelled because Mr. Goodenough forgot to confirm his reservation (as requested by the hotel). The only available room is at the Sailor's Inn, an old 4-star hotel by the port where they do not accept dogs. Ask for and thank him his understanding and patience.
> – It's cold and rainy in your city, so warm clothes are not a good idea.

Appendix C

Criteria employed in task design springing from needs analysis

	Simple	−Simple	Complex	+ Complex
+/−FREQUENCY OF INPUT	Frequent/known input (e.g. thanks for)	Less frequent or known input	Infrequent/ unknown input	Very infrequent/ unknown input (e.g. how could we possibly thank you for everything you've done)
+/−FAMILIARITY or +/−EQUAL STATUS	Familiarity with interlocutor/ same status (colleague you normally write to)	Less familiarity/ high status (e.g. boss)	Unfamiliar/ high status (unknown person in an institution)	Very unfamiliar/ very high status (a personality)

Appendix C. (*Continued*)

	Simple	**−Simple**	**Complex**	**+ Complex**
+/−CAUSAL REASONING	Low reasoning demands (simple info transfer and independent instructions) (e.g. ask for and offer info)	Moderately low reasoning demands (info transfer, minor reasoning and independent instructions) (e.g. ask for something in an unfavorable situation)	High reasoning demands (complex reasoning and interconnected instructions) (e.g. in an unfavorable situation, try to persuade by providing convincing reasons)	Very high reasoning demands (highly complex reasoning and interconnected instructions) (e.g. try to justify a major mistake and persuade interlocutor to still help you)
+/− INTENTIONAL REASONING	Explicit message (e.g. straightforward)	Less explicit message (e.g. some irony)	Implicit message (e.g. irony and an implicit threat)	Very implicit message (e.g. a hard-to-figure out intention)
+/−ELEMENTS	Very few instructions	Few Instructions	Many instructions	A lot of instructions

Appendix D

Pragmatics grid

0	1	2	3	4	5	6
Pragmatically incorrect.	• The opening and closing markers are not used. • The addressee is not acknowledged. • The degree of formality is very low and inadequate. • Pragmatic expressions (such as requests and apologies) are awkward and inappropriate. • Mitigation is not used in the pragmatic expressions produced (such as in requests or apologies).	• The opening and closing markers are not adequate for the context. • The addressee is poorly acknowledged and thus the formality of the speech act is awkward. • The structure does not meet the appropriate level of formality. • Pragmatic expressions (such as requests or apologies) are used but they are linguistically wrong. • Pragmatically expressions are rarely mitigated.	• There is some use of opening and closing markers although they do not fully meet the expected standards. • The addressee is partially acknowledged but the speech act still does not meet the required expectations. • The structure follows the standards of formality in a good way but still there are some formal mistakes. • Pragmatic expressions (such as requests or apologies) are used, but some inadequacies can still be identified. • Pragmatic expressions are not often mitigated.	• The use of opening and closing markers is correct and adequate but could be improved. • The addressee is acknowledged and the speech acts meets the expectations of formality to an acceptable extent. • The structure follows the expected formality but there are some minor mistakes. • Pragmatic expressions are appropriately used with minor linguistic inadequacies. • Some mitigation is used in pragmatic expressions.	• The use of opening and closing markers is correct and adequate. • The addressee is acknowledged and the speech acts meets the expectations of formality. • The structure follows the expected formality with sporadic minor mistakes. • Pragmatic expressions are appropriately used with sporadic linguistic inadequacies. • Mitigation is commonly used in pragmatic expressions.	• The use of opening and closing markers is correct for the context. • The addressee is fully acknowledged and is addressed accordingly to his/her position. • The structure of the email follows the expected level of formality. • Pragmatic expressions are fully pragmatically and linguistically appropriate. • Mitigation is always used in pragmatic expressions.

CHAPTER 8

Pragmatics, tasks, and technology

A synergy

Marta González-Lloret & Lourdes Ortega
University of Hawai'i / Georgetown University

This chapter explores the relationship among pragmatics, tasks, and technology when the goal is to support the development of pragmatics in a new language (L2). We view L2 pragmatic competence as culturally and situationally specific and inseparable from authentic communication, which encompasses both the face-to-face and digital worlds of people. Taking as a premise that optimal blends of tasks and new technologies can provide a programmatic framework for L2 instruction (González-Lloret & Ortega, 2014), including L2 pragmatics instruction, we first survey key existing task-based, technology-mediated approaches that incorporate a focus on the development of L2 pragmatics. We then choose two well-known areas of TBLT – cognitive task complexity (Robinson, 2011; Skehan, 2003) and needs analysis (González-Lloret, 2014) – to point at factors that can shape interactional pragmatics, and we show how they can be relevantly applied to technology-mediated tasks for the teaching and learning of L2 pragmatics. From a TBLT stance, we argue that the pragmatics-related demands of tasks might impact cognitive task complexity as an independent variable. From a technology-mediated TBLT stance, we propose that the systematic analysis of tasks in terms of pragmatics is necessary to carve a formal curricular place for L2 pragmatics instruction. We close the chapter by pointing at two fruitful future areas for investigating how the synergy among pragmatics, tasks, and technology can be optimized.

Introduction

Research on pragmatics has long demonstrated that pragmatic competence is essential to communication, maintaining rapport, and avoiding negative evaluative judgments. In the context of learning a new language (henceforth an L2), such competencies take on special importance. While language errors are apparent and can easily be forgiven when the speaker is clearly a learner, this is often not the case for pragmatic errors. As Thomas (1983) states:

> If a non-native speaker appears to speak fluently (i.e. is grammatically competent), a native speaker is likely to attribute his/her apparent impoliteness or unfriendliness, not to any linguistic deficiency, but to boorishness or ill-will. While grammatical errors may reveal a speaker to be a less than proficient language-user, pragmatic failure reflects badly on him/her as a *person*. Misunderstandings of this nature are almost certainty at the root of unhelpful and offensive national stereotyping.   (pp. 96–97, italics in original)

Within the field of pragmatics there is also a recognition that task achievement plays an important role in how humans design and conduct their interactions. The accomplishment of a task is essential for the success of the interaction. At the same time, the interactional norms and people's face must be balanced in relation to the task needs and, hence, task completion affects the way we interact with others. Several factors contribute to how people interact with others to complete a task. Pragmatics scholars have long identified contextual factors such as social distance, relative power, and ranking of imposition (Brown & Levison, 1987). Other equally important factors underlie sociocultural interactive principles that help manage people's basic interactional concerns such as face, rights, obligations, and task achievement (Spencer-Oatey & Jiang, 2003). These interactive forces are socioculturally based and influence how people interpret and produce language according to a culture and/or a specific context.

For the purposes of L2 learning, culture and technology greatly complicate what it means to interpret and produce language in pragmatically competent, recognizable ways in a new language. Although all cultures have their own interactive principles, these are not the same across cultures. For example, Spencer-Oatey and Jiang (2003) discovered that British speakers seem to associate task achievement with speaking clearly and directly, while this is not the case for Chinese participants; however, for both groups the specific context of interaction seems to influence their decision as to whether it is more important to complete a task or to prioritize a concern for face/rapport. These cross-cultural differences and similarities pose a challenge for language learners, who may have little exposure to that new culture, especially if pragmatic information is reduced to a few notes on a textbook, as is typically the case.

The challenges are even more complicated to manage when communication is mediated by technology, which in itself creates its own cultures and expectations. For example, Computer Mediated Communication (CMC) requires a different set of interactional characteristics to maintain communication (Walther et al., 2015). It follows, as Thorne (2003, 2016) has argued, that many forms of communication that are mediated by technology, including CMC, can be seen as a case of double intercultural communication. When users meet digitally to practice a foreign language, they not only bring different language and cultural backgrounds to the task, but their virtual communication gives rise to new, unexpected, digital cultures on the fly. Moreover, technology not only shapes the pragmatics associated with the successful completion of a task,

but in itself constantly creates new digital target tasks. Many of our students learning a language are already seamless users of social and leisure technologies when they arrive in our classrooms. Social technologies are so intimately integrated into their everyday lives that shifting social practices are happening digitally, and the boundaries between digital and real are blurred for them. For many of our students, constantly emerging new digital tasks are no less "real" than traditional tasks (Chapelle, 2014).

In our previous work (González-Lloret & Ortega, 2014), we have argued that technology and tasks serve language education best when two conditions are met. First, tasks and technologies have to be blended by reference to the programmatic educational principles of TBLT (Norris, 2009), that is, conducting a needs analysis, selecting and sequencing tasks, developing materials, teaching, assessing students' learning outcomes, and evaluating programs. Second, the transformative nature of technologies needs to be recognized and harvested so that tasks are more than mere digital translations of activities that could be equally done in traditional formats. It is when these two optimal conditions are met that we concur with Ziegler's (2016) claim that "tasks and technology are ideal partners in a reciprocal relationship" (p. 137).

In this chapter, we apply this thinking to the blending of technology and tasks for the specific purpose of supporting L2 pragmatics development in the classroom. We concentrate on interaction as the main locus of pragmatics development, a focus that aligns well with the interactive nature of typical pedagogic tasks in TBLT. In this view, which we call interactional pragmatics, pragmatic meanings "do not inhere in linguistic conventions but result from participants' ongoing, contingent interpretive work during jointly pursued practical activities" (Kasper, 2009, pp. 278–279). Our choice of interactional pragmatics is therefore intentionally different from other research that incorporates technologies and tasks for the study of pragmatics from a more cognitive-psychological perspective (e.g., Taguchi, Kaufer, Gomez-Laich, & Zhao, 2016). We endeavor to show that focusing on pragmatics as part of the language required to accomplish a task would help learners succeed in that specific interaction, as well as create and maintain the rapport necessary to continue future interactions and establish social relations. When these tasks leverage the mediation of new technologies, the interactional pragmatics required for successful task completion may become even more complex, but these technology-mediated tasks will also become all the more authentic and motivating for students, in turn supporting valuable pragmatics learning.

Supporting pragmatic development with technology and tasks: A précis

The literature integrating technology, tasks, and pragmatics is rich. In this section, we will review studies that feature technology-mediated tasks conceived as

real-world tasks, with a main focus on meaning, and with interactional pragmatics learning built into the task design so as to be experienced by learners as mostly incidental. This is in keeping with the notion in TBLT that tasks must offer some goal-oriented, "language-and-action experience driven by some communicative purpose independent from (although ideally compatible with and even essential to) language learning purposes" (Ortega & González-Lloret, 2015, p. 74). Technology-mediated TBLT in the service of pragmatics learning simultaneously strives to promote some of the basic principles of technological applications such as motivation, authenticity, choice, creativity, and community (Itō et al., 2010). Thus, the tasks are learner-centered and allow for diversity, flexibility, and learner agency, encouraging students to use their digital skills as well as their own linguistic and non-linguistic resources.

Telecollaborations are among the most popular choices for supporting the development of interactional pragmatics through blends of technology and tasks. These are pedagogical multimodal online exchanges, in which two groups of students in geographically distant places meet and interact in the L2. Early examples are explored in Belz and Kinginger (2002, 2003), Kinginger and Belz (2005), and González-Lloret (2008). All of them tracked the development of addressivity, or the linguistic and stylistic resources used to indicate to whom any utterance is directed, which are shaped by expectations of communication with imagined and real audiences. Pragmatic development was therefore operationalized as the evolution of learners' uses of formal and informal pronoun choices (e.g., *Du/Sie* in German and *tú/usted* in Spanish) from less to more target-like (as measured by the use of the same forms by their native speaker interlocutors). The studies by Belz and Kinginger investigated a telecollaborative setting between students of L2 German in the U.S. and students of L2 English in Germany who engaged in several tasks through text chat and email (e.g., text discussions and the creation of a website for a month). The researchers found that the developmental paths for forms of address in German varied greatly among learners but were responsive to features of online social interaction (e.g., amount of interaction, quality of interaction, etc.). González-Lloret's (2008) L2 Spanish learners in the U.S. engaged in a ten-week telecollaborative project with first language (L1) Spanish speakers in Spain. The task, completed via text-based chat, was designing an itinerary for a trip. The results of a longitudinal microanalysis revealed that students not only paid attention to the completion of the task but also engaged in negotiation of pragmalinguistic rules for how to address each other in the L2. A close case analysis of one of the learners showed that, although she displayed early understanding of sociopragmatic rules, it required several interactions before she could adjust her pragmalinguistic use of address forms to the target-like form. This confirmed that, as Alcón Soler (2002) noted, the "relationship between collaborative dialogue and learners' development of pragmatics is not immediate" (p. 371).

Another group of studies also employed telecollaboration but focused on how tasks and technology can be used for the development of intercultural competence, a subfield of pragmatics (see Byram, 1997). Using the virtual environment *Second Life*, where people interact with others in an online space through the use of an avatar, as the platform for the telecollaboration, a team of researchers at Utrecht University (Canto, de Graff, & Jauregi, 2014; Jauregi, Canto, de Graaff, Koenraad, & Moonen, 2011; Jauregi & Canto, in press) developed several tasks to promote intercultural competence in L2 Spanish. In these studies, pragmatic development was closely related to intercultural development. Following Byram's model, intercultural communicative competence was defined as the ability to learn about other cultures, apply skills to unknown situations, and respect and tolerate other worldviews. With this in mind, the tasks developed by the team elicited conversations about everyday implicit cultural habits and beliefs. The researchers also carefully chose topics that would allow for contrast and comparison of learners' own beliefs and those of their interlocutors. Through telecollaboration in *Second Life*, participants from different institutions in different countries with different L1s met remotely, and through their avatars negotiated a place to go out with their roommates, visited and decided on an apartment to rent, planned a holiday, impersonated characters and observed people's reactions, or played a cultural game. Canto et al. (2014) describe some of the tasks more in depth and present positive results confirming that learners did in fact engage in intercultural negotiation and that the potential of such tasks was realized and supported learners' understanding of others. Research that incorporated tasks and telecollaborative communication for developing intercultural pragmatics is prolific, including the oft-cited studies by Belz (2005) or Vinagre (2010). The design principles of tasks that support intercultural competence in telecollaborative environments have also been evaluated by O'Dowd and Ware (2009).

Some task-based, technology-mediated pragmatics applications have featured learning objectives that are quite traditional, such as speech acts, but they have been coupled with task designs that take full advantage of recent immersive technologies, known as synthetic immersive environments (SIEs). As a result, the pragmatics learning achieved in these environments is considerably more sophisticated and authentic than that in traditional pragmatic teaching formats.

One of the early examples in this line of work is *Croquelandia*, a SIE developed by Sykes (2008, 2014) to engage students in how to appropriately perform requests and apologies in Spanish through different tasks. Pragmatic development was operationalized as the increase of correct requests and apologies performed by students (as measured by pre- and post-tests). The game begins with learners winning a travel abroad trip to *Croquelandia* and needing to successfully navigate apology and request interactions with their peers, host family, and professors in the virtual environment. The tasks involve fairly mundane scenarios for these speech acts, such as borrowing a

book, asking for help with a party, apologizing for not cleaning the house, breaking a vase, and so on. The scenario-based tasks were designed considering social variables in the context of interaction, such as power, social distance, rank of imposition, and severity of the offense. Although the tasks are mundane, the unique feature of *Croquelandia* is the use of immersive technologies. The SIE is non-linear in that students can complete tasks, which are much like adventure games, by completing subtasks that help them accomplish the main task. Students can also use aids coming from the environment (clues, eavesdropping) and classmates (through chat). The non-player characters react to learners' pragmatic appropriateness of their responses (from proposed choices) by, for example, showing a furious reaction when a learner breaks a vase but she/he does not apologize appropriately. In addition, if the learner decides to skip some requests, they have to do additional repair work, making the next interaction more difficult.

Sykes (2014) evaluated the learners' experiences in the game, including their willingness to use the SIE's in-built function of restarts (i.e., trying out a situation multiple times). Although pre- and post-test comparisons showed only small improvement, qualitative data collected via interviews, surveys, and in-game behavior observation data suggested an increase in learners' pragmatic abilities with issuing requests. In her evaluation of the environment, Sykes points out the importance of designing this type of task-based SIE environments as learner-centered, as opposed to learner-driven. By this she meant that it is essential to go beyond the task design itself and evaluate whether and how learners take advantage of the designed functionalities (e.g., the specific quests, restarts, type of feedback, complexity of task delivery, potential for player agency, etc.). Indeed, her data confirmed that students did not necessary follow the ideas planned by the SIE designers. Sykes therefore proposes (following Breen, 1989) that any evaluation of SIEs must consider multiple dimensions at once, including the task-as-workplan (the intended design experience), the task-in-process (learner and teacher involvement while completing the task), and the task outcome (in *Croquelandia*, the pragmatics learning achieved). She also proposes to move away from the idea of "playing to learn" and instead adopt a "learning to play" perspective (p. 176). This perspective, borrowed from the gaming literature, views tasks as more than just practices; tasks become the reason to continue playing. Learners learn skills to be able to advance in the game, but in the process, they also learn language skills (lexicon, structures, pragmatic forms, etc.), strategies (collaboration, use of resources, problem solving, etc.), and technological skills (how to move and communicate with avatars in a SIE).

Task-based immersive and gaming technologies have also addressed other less traditional learning goals beyond speech acts. For example, focusing on interactional pragmatics and featuring a SIE called *Final Fantasy*, Piirainen-Marsh and Taino (2014) explored how knowledge asymmetries and changes in game knowledge impact L2

participants' interaction and their identities, as well as the epistemic positions they develop and express within the group. The same researchers have also explored other interactional resources in earlier studies, including coproduction of talk (Piirainen-Marsh, 2011) and other-repetition (Piirainen-Marsh & Taino, 2009). They reached the interesting conclusion that social gaming creates opportunities for joint action and negotiation that make interactional competences relevant. Participants "draw on multiple sources of knowledge and a range of interactional resources when negotiating game tasks or situations and solving emerging problems" and learner experience "changes in epistemic organization [that] lead to more equal opportunities to participate and a cooperative stance" (2014, p. 1035). We could thus posit that these increments in negotiation and participation should lead to an increase in overall language competencies.

A recent immersive and gaming technology draws on place-based augmented reality, which is known in everyday life applications such as GPS. At the University of New Mexico, Holden and Sykes (2012) created a game called *Mentira* for L2 Spanish learners <www.mentira.org>. Students collaboratively interact in groups with game characters who belong to several Spanish-speaking families in a community (Los Griegos in Albuquerque) with the goal of solving a crime that occurred in that neighborhood in the past. Students gather the evidence that they need from the game characters and the neighborhood. The game activity is then solved in the regular classroom via students' face-to-face discussion by deciding who actually committed the crime. This jigsaw type of game aims not only at teaching Spanish, but also incorporates dialectal elements from the Albuquerque area, hence targeting sociolinguistic learning in Spanish. Moreover, the game incorporates Spanish pragmatics by ensuring that a successful interaction between the learners and the characters in the game is highly dependent on learners' levels of politeness. Namely, learners' inappropriate responses elicit less interesting information from the characters and may even end the game. Over several iterations of the game design, the authors implemented enhanced practices that are typical of video gaming and are essential for L2 interactional pragmatics learning such as "impromptu collaboration, risk taking, role playing, learning to play vs. playing to learn, and taking ownership of their experiences within the game world" (p. 123). Holden and Sykes point out how games can add more productive learning behaviors that are missing from most educational contexts and thus help students become aware of the importance of appropriate pragmatic choices when interacting with others (non-player characters in this case). The incorporation of place-based augmented reality in these immersive environments adds authenticity and motivation to the already very engaging digital world of SIEs.

Although the studies presented above focus on the presentation of pragmatics mostly incidentally through the use of the technology, an interesting question is

whether offering explicit pragmatics instruction in the classroom while participating in meaning-oriented online modules can boost pragmatics learning. This interest in knowing whether explicit or implicit pragmatics instruction is more effective is not new (e.g., Kasper, 1997; Rose 2005; Jeon & Kaya, 2006). Cunningham (2016) is one the first studies to investigate whether pragmatics learning can be achieved incidentally (albeit by design) with meaning-focused digital tasks alone or whether pragmatic instruction is preferable, or even necessary for some pragmatic targets, such as the use of internal mitigation to requests (e.g., using the modal "could"), that may be more difficult to learn based on sheer experience than other targets, such as externally mitigated requests (e.g., "I need to ask you a favor").

The research we have reviewed in this section offers robust evidence that the symbiosis of technology and tasks can effectively support L2 pragmatic development. Nevertheless, the evidence also shows that pragmatic development is not equally achieved by all participants. Learning pragmatics requires active interaction on the part of the participants to produce 'critical incidents' (Belz & Kinginger, 2002), repeated iterations (González-Lloret, 2008), or a willingness to take advantage of the restart design in many SIE games (Sykes, 2014). Pragmatics learning is also dependent on other factors such as learners' agency, their willingness to adopt pragmatic norms or disregard them according to their identity and personal choice, the expectations that are imposed on them, and the leeway they are granted as non-native speakers of a language (Belz & Kinginger, 2003). Many of these conditions and moderating factors are equally important in the development of L2 (and L1) pragmatics in face-to-face interactional contexts as well (Hasall, 2015; Ishihara & Tarone, 2009; Kim, 2014; Siegal, 1996). The unique edge of technology-mediated tasks for the learning of pragmatics is that the best blends of tasks and technology can greatly enhance traditional classroom practices, which cannot offer the same level of interactional and social contexts for the development of L2 pragmatic competence.

All the studies reviewed above integrate tasks, technologies, and interactional pragmatics learning, and thus are practical demonstrations of how to merge pragmatics and technology-mediated TBLT in the service of L2 pragmatics learning. In the next two sections, we turn the lens of L2 interactional pragmatics onto two well-known areas of TBLT: cognitive task complexity (Robinson, 2011; Skehan, 2003) and needs analysis (González-Lloret, 2014). Our goal is to show how well-known insights from interactional pragmatics can be relevantly applied to technology-mediated tasks for the teaching and learning of L2 pragmatics. In a nutshell, we will try to argue that the pragmatics-related demands of tasks might impact cognitive task complexity as an independent variable, and that the systematic analysis of tasks in terms of pragmatics is a necessary element in order to carve a formal curricular place for L2 pragmatics instruction.

Shall pragmatics and cognitive task complexity meet?

In the TBLT literature, understanding the cognitive complexity of tasks has proven essential for developing tasks and sequencing them to form a cogent L2 curriculum. The research agenda in this regard has focused mainly on operationalizing cognitive task demands and has led to the consolidation of proposals by Skehan (2003) and Robinson (2011). Both are by now well-known to TBLT communities. Attempts at investigating the relationship between pragmatics-related demands of tasks and cognitive task complexity have just begun with the pioneering work of Kim and Taguchi (2016) and the chapter by Gilabert and Barón in the present volume. Here we would like to offer some theoretical justification for our own position that the pragmatics-related demands of tasks might in fact impact their cognitive demands, and that technology mediations compound to shape these demands.

In their partially competing and partially overlapping views on task demands, both Skehan (2003) and Robinson (2011) acknowledge the existence of interactional or communicative factors that affect learners' production in the areas of complexity, accuracy, and fluency, or CAF. In fact, many of the proposed factors can be related to issues of interactional pragmatics, such as who the participants are (i.e., not just the number of participants but their relationships and their dynamics of familiarity and power), what the context and the medium of communication are, and so on. In particular, Robinson differentiates between task complexity, task difficulty, and task conditions. *Task complexity* is affected mainly by cognitive factors like reasoning or planning, while *task difficulty* is influenced by learner abilities including proficiency or aptitude. *Task conditions* result from interactional demands of the task such as whether the interaction flows one or two ways, the number of participants included, participants' characteristics such as gender, level of proficiency, status, etc. This third dimension clearly involves pragmatics. Specifically, it most closely engages sociopragmatic interactional principles, which in turn necessitate pragmalinguistic choices that may be specific to each target language. This is in keeping with the classic distinction in pragmatics between sociopragmatics, which refers to the norms and principles that affect the behavior of the participants, and pragmalinguistics, which refers to the choices in the language use to realize those norms (Thomas, 1983).

Table 1 offers our translation and expansion of Robinson's (2011) task conditions into different aspects to be considered in interaction, which we take to reveal sociopragmatically grounded variables that might be manipulated in order to increase or decrease the cognitive demands of a given task. The +− symbol in the table expresses the scalar nature of sociopragmatic behavior and thus of some of these variables which, as we will explain below, vary according to cultural and situational specific contexts. This is in contrast with the polar (±) nature of task conditions that is assumed in Robinson's model.

Table 1. Comparison of proposed sociopragmatic interactional variables with Robinson's (2011) *task conditions*

Based on Robinson's (2011) task conditions	Sociopragmatic interactional variables
a. **Participation variables making interactional demands** ± open solution ± convergent solution ± one-way flow ± few participants ± few contributions needed ± negotiation not needed	a. **Interactional variables (culturally and situationally specific)** number of participants, flow of interaction, amount of contribution and negotiation = sequential organization of the interaction (Schegloff, 2007) • turn taking • topic initiation/ending • preference/dispreference +− politeness principles (Lakoff 1973) • avoid imposition (give options) • avoid negative evaluation by the hearer • avoid hurting feelings +− relevance (Sperber & Wilson, 1986/2001)
b. **Participant variables making interactant demands** ± same gender ± familiar ± shared content knowledge ± equal status and role ± shared cultural knowledge ± same proficiency	b. **Participant variables (culturally and situationally specific)** +− same gender +− familiarity +− shared background knowledge (social, situational, personal) +− power status
	c. **Artifact and environment mediation** (Marsh & Onof, 2008)

First, as seen in Table 1, under *participation* variables, Robinson (2011) stipulates that the number of participants, how the interaction flows, and the quantity of contributions are important in the design of tasks. We would argue that these factors shape the sequential organization of an interaction (Schegloff, 2007), including the turn-taking sequence (how are turns allocated, appropriated, simultaneously produced, repaired, etc.), who initiates, directs, and changes topic, and how people organize their language when they are producing a preferred or dispreferred turn. In addition, other interactional aspects play a role in the complexity achieved by the participation variables of a task. Following Lakoff's (1973) Cooperative Principle, we know that in interaction people avoid imposition on their interlocutors, avoid being evaluated negatively by what they say, and try not to hurt others' feelings. Therefore, a task that demands from the speaker to impose him/herself, possibly projecting a negative image, will be interactionally more difficult to accomplish for a speaker and may result in avoidance. It is important to also consider Sperber and Wilson's (1986, 1986/2001) Relevance Theory, which in the field of L1 pragmatics is currently one of the most influential models about how social beings interact with others. According to Relevance Theory, when we

hear (or read) an utterance directed to us and intended to communicate something, we interpret it for its relevance to us. That is, we assume it is relevant enough to be worth our processing and it is the most relevant the speaker could have produced. From the different possible meanings an utterance may have, given the physical context, background assumptions, gestures, and so on, we derive the most plausible interpretation. Therefore, a task that requires the speaker to produce maximally relevant information will require less processing on the part of the hearer and therefore be less interactively complex. It is important to note that this interpretation is not a one-time affair but rather is culturally and situationally grounded. We interpret and reinterpret meaning and intentions constantly as an interaction unfolds (Gibbs, 2001).

Second, Robinson (2011) acknowledges in his TBLT model of task complexity *participant* variables (gender, familiarity, shared knowledge, and power status). Their influence on the kind of interactions engendered by a given task has been studied in SLA (e.g., Azkarai, 2015; Gass & Varonis, 1984). What we would emphasize from a pragmatic-theoretical perspective is that their impact on interaction is not constant or fixed, but instead is situationally and culturally bounded. For instance, the same task (e.g., purchasing a book) conducted by a non-religious man and woman in the United States would substantially change (and might be much more challenging if not impossible) when happening between a man and a woman in some Muslim countries (but not between two women), and it may be less challenging between that same Muslin man (or woman) when interacting in the United States. To further illustrate, a task including a request between a boss and her employee may require a different interaction if done inside the office during a work meeting than if done in a bar during an after-work social event, and the interaction will also be different if boss and employee are in a country where socializing over drinks after work is a common practice versus a country where it is not.

Finally, Table 1 includes one more element which is essential when interacting and that is not present in Robinson's *task conditions*: the presence of any artifacts that may mediate the interaction. An artifact can be many things, from a small business card that mediates the introduction in a business interaction to a technological tool such as a computer. Marsh and Onof (2008) famously characterized humans as culturally bounded, rationality-bounded, and environmentally located agents. Their characterization is reflected in Table 1 in our stipulation of cultural and situational specificity. Humans almost always communicate indirectly because of mediations, or artifacts, that are a result of modifications of the environment. In today's world, and in the context of the discussion in this chapter, digital technologies provide what is perhaps the most powerful modification to our environments and a ubiquitous set of mediating artifacts. Artifacts structure how people engage in social action and in language activity (Marsh & Onof, 2008). Thus, for example, a text-based chat tool that will help us connect to a customer service representative when purchasing or fixing a computer

product, will structure the communication differently than a phone call interaction with the same customer service.

For example, in the computer-based chat interaction with a customer representative (for a purchase, technical trouble, etc.) turn taking might be considerably slower than in the same activity on the phone, since the technician may be engaged in more than one chat at the same time. The turns can be expected to be shorter as well, and the information more to the point. In addition, some sequences that are found in the phone counterpart activity, such as small talk, may not occur. More generally, and as we have argued elsewhere (González-Lloret & Ortega, 2014), within a technology-mediated TBLT framework the technology does not only mediate but modifies the task as well as the language involved in it, and in itself constantly creates new real-world target tasks. Knowing the pragmatics involved in interactions that are mediated by those artifacts should bear on the complexity of that task.

In a nutshell, we propose that the pragmatics-related demands of tasks might in fact impact their cognitive demands. If so, we argue that pragmatics task characteristics can be a source of fruitful experimental manipulation that remains untapped in the current TBLT research program. Put more concretely, the interactional variables outlined in Table 1, belonging to *task conditions* in Robinson's framework, need to be attended to when diagnosing and/or manipulating *task complexity*.

We recognize that this is uncharted territory that has hardly ever been the focus of research. But the possibilities for empirical and pedagogical implementation are compelling. The empirical implementation would not be very different from already existing classroom studies that have explored cognitive task complexity in CMC in contrast to face-to-face tasks (Adams & Alwi, 2014; Baralt, 2013, 2014; Collentine, 2010, 2013; Nik, Adams, & Newton, 2012). Rather than manipulating cognitive factors of the task, such as prior knowledge or reasoning demands, we propose to manipulate sociopragmatic interactional variables of the task (e.g., gender, power difference, degree of politeness), while maintaining constant cognitive factors that affect the task complexity. Technology has a great potential to conduct this type of research. For example, in immersive and virtual environments, variables such as gender, origin, familiarity, and power can be assigned to different avatars when students need to interact with those avatars in a task that is constant otherwise. From a pedagogic point of view, interactional spaces mediated by technologies can offer an opportunity for real-life interaction that can be archived, retrieved, and brought back to class for further analysis and feedback. This reflective element, as Samuda and Bygate (2008) note, is a staple of TBLT that issues from the Deweyan vision of experiential learning, and it can be done as explicitly as the teacher may deem necessary (see Cunningham, 2016). From an interactional pragmatics perspective, we would emphasize that such a reflective feedback loop would focus on issues that may bring about cross-cultural and pragmatic awareness of the way people interact and construct their ways of being.

In sum, we are hopeful that in the near future researchers interested in exploring synergies among tasks, technologies, and pragmatic learning might turn their empirical attention to the study of how the pragmatics-related demands of tasks might impact their cognitive demands. The benefits, we suggest, lie not only in the potential for this new research program to invigorate a line of investigation that is at the center stage of TBLT interests, but also because the manipulation of pragmatics task demands – if indeed they can be shown to have an impact on the cognitive demands and on the language produced – may ultimately help modulate L2 learning as well. In the next section, we explore another important line of research in TBLT, systematic needs analysis, under the lens of interactional L2 pragmatics.

Needs analysis

A programmatic design of L2 pragmatics instruction from a technology-mediated TBLT perspective would require the incorporation of sociopragmatic and pragmalinguistic elements at all stages of the curriculum design (Norris, 2009), from the needs analysis all the way to evaluation and assessment. In our discussions thus far, we have paid attention to L2 pragmatic instructional design and implementation issues. Issues surrounding the assessment of L2 pragmatics have also received considerable attention by researchers (e.g., Roever, 2011, 2013; Youn, 2015), and Part II of the present volume is devoted to the task-based assessment of pragmatics, including chapters by Youn and Timpe-Laughlin. By comparison, work is scant that addresses needs analysis to identify the requirements of real-world tasks for the teaching and learning of L2 pragmatics. We therefore reflect on what would be involved in analyzing needs for a task-based, technology-mediated approach to teaching and learning L2 pragmatics.

In order to understand what L2 pragmatic knowledge and resources would be required to successfully perform a task via a formal needs analysis, we would want to specify a number of features related to the three dimensions of actions, language, and technology. The dimensions are outlined in Table 2, using the act of purchasing a mobile phone as illustration.

First, we would need to elicit information about what *actions* need to be accomplished, including the participants' behaviors and the sociopragmatic features that would determine the interaction. The latter would include analyzing cultural norms, contextual norms, the relationship among participants, the background knowledge of the participants, and whether any artifacts mediate the interaction. For example, in order to buy a mobile phone in the L2 we need to greet the sales person, request information about one or more mobile phone(s), express the desire to buy (or not to buy) the phone, complete the payment transaction, etc. In order to do this successfully, we need to respect the social norms of the culture for a transactional context (i.e., the

Table 2. Components of the needs analysis of a task to teach and learn L2 pragmatics

	Components	Objective: To find out…	An example: purchasing a mobile phone
Actions	**Actions**	What needs to be done in order to successfully complete the task – participants' behavior	– Greet the sales person – Request information about one or more mobile phone(s) – Express the desire to buy (or not to buy) the phone – Complete the payment transaction
	Sociopragmatic features	What are the interactional norms surrounding the task – cultural norms, – contextual norms, – relationship among participants, – background knowledge – mediating artifacts	– What is the body language permitted/appropriate in that context – Who initiates/ends the interaction – Do we bargain or not? – Are there any mediating artifacts (business card, credit card)
Language	**Target language**	What language elements are essential to successfully complete the task – lexical choices – structures – routines	– Ask questions – Respond to questions – Understand and use specific vocabulary – Display agreement and disagreement
	Pragmalinguistic features	What language is essential to appropriately accomplish the task according to cultural and contextual norms	– Formulaic greetings and farewells in a service context – Formulas to express wants – Language for accepting/ rejecting offers – Ways to initiate and change topic – Politeness strategies – Turn-taking norms
Technology	**Technologies**	What technologies are needed, Whether the learners have access to them, What is the learners' level of expertise with these technologies, What are the sociopragmatic norms of the context (the netiquette)	– How to navigate the site – Ask questions to an online representative – Use the online payment system – Save/print prove of purchase – Appropriate norms of interaction in the medium

physical gestures to greet, whether to bargain or not, whether part of the interaction calls for any artifacts as a business card or a credit card, who initiates and ends the interaction, etc.) and behave in ways that are appropriate to our relationship with the sales person (i.e., the sales person can be a complete stranger, a neighbor, or a family member that works in the shop).

Second, we would need to determine the *language* needed to accomplish the task. Here one would want to identify not only specific vocabulary, structures, and language routines but also the pragmalinguistic choices important for the success of the interaction, which are directly related to the corresponding sociopragmatic norms. For instance, depending on the sociopragmatics of an action, different choices of politeness markers may be called for when formulating questions (e.g., "I would like" versus "I want to", or the use of hedges and mitigators such as "I don't really think that I can afford it" instead of "I don't want it"). It is important to emphasize that both culture and the situated context are essential in shaping the sociopragmatic and pragmalinguistic choices available, since it is possible that in different speech contexts within the same culture certain behaviors and language choices are different. For example, although we know that politeness is preferred over impoliteness and indirect requests preferred over direct requests, when preventing someone from getting run over by a car crossing the street, a very direct command ("stop!") with a not-so-polite yank of the arm is an accepted pragmatic reaction in most cultures.

Finally, as shown in Table 2, we need to determine what *technology* (if any) mediates the task, and how any technologies involved will likely affect the interaction. Included here are the norms and regulations of interaction through that technology. The interaction may be slightly or very different from the same interaction without the technology. For example, if the phone is being purchased online, we need to know whether participants are able to navigate the site, ask questions to an online representative, use the online payment system, and save/print proof of purchase. During an online purchase, a possible online chat with a representative will be quite short and focused, and will consist mainly of brief questions and answers. This may be different from the more extended exchange with a representative in a face-to-face encounter.

In sum, as we have also argued elsewhere (González-Lloret & Ortega, 2014), within a technology-mediated TBLT framework, the technology does not only mediate but modifies the task, and from a pragmatics perspective it also structures the language involved in it. In these ways, technology in itself constantly creates new real-world target tasks and thus new needs for students wanting to use the target language to do things.

If our sketch in Table 2 rightly captures how the three dimensions of actions, language, and technology shape the sociopragmatic and pragmalinguistic requirements and choices for a given interaction, then systematic needs analysis along these three

dimensions can become a powerful tool for technology-mediated TBLT. For instance, it can be used to create a bank of real-world tasks analyzed for their usefulness for particular learner populations or educational contexts and specified for the actions (including sociopragmatic features), language (including pragmalinguistic features), and technology (including digital competencies) they call for. This may be a good investment for researchers interested in the study of how L2 pragmatics develops in technology-mediated task-based environments, and it would facilitate the sequencing and cognitive task manipulation steps that would be leading to a full curriculum design. We know of no published needs analysis of this kind, although a few examples exist of needs analysis conducted for L2 pragmatic teaching not including a technological component (e.g., Frenz-Belkin, 2015; Hertel, 2015). Yet for us, the synergies among tasks, technology, and pragmatics can only be optimally exploited in a TBLT curriculum if preceded by a systematic needs analysis (see González-Lloret, 2014).

Future research directions

We would like to close this chapter by suggesting two areas for future research. A first area is the investigation of which tasks are best suited for what kinds of L2 pragmatic learning and which pragmatic demands are most central to what types of technologies. This is because – as we hope our discussion has made clear – not all tasks lend themselves automatically to pragmatics learning, and not all technologies entail the same types of pragmatic demands.

There are many opportune gaps to be filled in pursuit of this first future line of research. We submit three examples here. First, within the gaming for learning community, the use of tasks is inherently present since most games are designed as spaces where quests or tasks need to be accomplished in order to keep moving within the game. But not all gaming researchers interested in L2 learning necessarily think of pragmatics as an area for learning gains during gaming. What exactly is gained, when implicit L2 pragmatics learning is targeted by design, versus when the learning is entirely incidental to the game? As another example, different technological platforms will inevitably shape at least in part different pragmatics of digital communication, but we have little knowledge of exactly how this happens, because different technologies have yet to be compared systematically for their pragmatic-interactional demands. The opportunities to do so are ripe. The earliest telecollaboration experiences, for instance, drew heavily on text-based chat for the support of interactional competence learning, intercultural learning, or both. But telecollaboration platforms have greatly diversified in the last decade. A recent review by Çiftçi (2016) suggests that for the purposes of intercultural telecollaboration alone the four most commonly employed technologies currently are online message boards, text-based chat, blogs, and email exchanges,

whereas a few other studies have also tried microblogging such as Twitter, podcasting, video recording, and video conferencing. Finally, as our review of the relevant literature has shown, most researchers interested in documenting the L2 pragmatic benefits of using technology have turned to telecollaboration, SIEs, and gaming in place-based augmented reality. By comparison, digital social networks (e.g., Facebook or Twitter) have attracted markedly less attention from an L2 pragmatics learning perspective. It may be justified to assume that other dimensions of L2 learning are better foregrounded in social network use, such as the development of L2 identities or literacies. As Blattner and Fiori (2011) have shown, digital social networked communication also shapes interesting pragmatic demands that can help foreign language educators address student needs for sociopragmatic learning in the L2.

In sum, future research is needed to elucidate what technological platforms can be suitable for what specific tasks and at the service of what diverse pragmatic learning objectives. It is only through a research-informed alignment among technologies, tasks, and pragmatics objectives that the potential for supporting L2 pragmatic development through technology-mediated TBLT will be achieved. Moreover, future research needs to compare the kinds of pragmatics learning that may be best fostered in technology-mediated TBLT proposals when the digital tasks have and have not been designed with the purpose of teaching (explicitly or implicitly) some aspect of L2 pragmatics.

A second direction for future research that can be led by L2 pragmatic researchers interested in tasks and technology is more meta-disciplinary: This could broaden the technology-mediated TBLT lens and help problematize and ultimately illuminate the notions of "tasks" and "real-world" tasks: What constitutes a task and what makes for "real world" authenticity in any given task experience?

Somewhat in conflict with traditional notions of task that TBLT researchers may entertain, many L2 pragmatic researchers – and certainly all discourse analysts – would consider conversation a task, since this is one of the authentic common things people do every day. The same is true of other tasks common in academic settings, such as discussion of a book or movie, debate about a topic, peer-editing sessions, and so on. If conversations and academic activities are considered tasks, then we would have to include quite a few more studies that have investigated a variety of topics related to interactional pragmatics through diverse technologies such as text, audio and video chats, forums, social networks, and so on. One example of the potential of conversational tasks for pragmatic development is found in González-Lloret (2011), who investigated the changes through an eight-week telecollaborative project of a L2 Spanish learner's speech act of complaining or commiserating, or what is known by discourse analysts as troubles-talk (Jefferson, 1988). An example of L2 pragmatics learning supported by academic tasks is Tsai and Kinginger (2014), who investigated the development of the speech acts of advice giving and receiving among ESL learners

engaged in peer-review tasks. In both studies learners demonstrated a clear development in their interactional practices. In addition, both studies support the idea that CMC environments can be designed so as to be conducive to pragmatics learning, and that the selection of not only appropriate technology but also appropriate tasks can go a long way in succeeding with this goal. Thus, it is important not to overlook the potential for pragmatic development of embedding technologies into tasks that may be less traditional than the TBLT mainstream community usually has in mind. This line of research, reviewed in depth by González-Lloret (2014), empirically demonstrates that technology allows students to engage with other speakers in pragmatic sequences and using interactional resources that are essential for the development of interactional competence in regular face-to-face communication and that are often neglected in L2 classrooms. These resources include negotiation of face (e.g., Golato & Taleghani-Nikazam, 2006), identity construction (e.g. Stommel, 2008), addressivity (e.g. Kim & Brown, 2014), and turn-allocation (e.g. Jenks & Brandt, 2013). Here, the main potential of the technology is to provide sufficient and sustained engagement to lead to interactional development in the L2, which in digital environments as much as in face-to-face environments is a matter of recalibration of resources for social action (Pekarek Doehler & Pochon-Berger, 2015).

In sum, what constitutes a "task" and what makes for "real world" authenticity are vexing definitional problems for TBLT researchers and teachers (Chapelle, 2014; Ellis, 2009; Long, 2016; Van den Branden, 2016). The lenses that L2 pragmatics scholarship offers can be effective as a tool for exploring expanded notions of task and authenticity. Namely, a task can be considered a task if it is treated as social action by the learners themselves, in accordance with diverse emic perspectives of qualitative researchers and Conversation Analysts (Markee, 2013). Likewise, technologies can be treated as decidedly (and ubiquitously) authentic, since all technologies are cases of modifications to the human environment through mediations and artifacts (Marsh & Onof, 2008). Virtual communication creates its own cultures and expectations, opening spaces for double intercultural communication, where not only different language and cultural backgrounds come into contact through the task, but also new, unexpected, digital cultures may emerge on the fly through the technology (Thorne, 2003, 2016). This is how the culturally and socially bounded perspective of L2 pragmatics can illuminate the notion of tasks.

Conclusion

In this chapter, we have explored how the blending of technology and tasks can serve the specific purpose of supporting L2 pragmatic development in classroom learners. We have argued that, from a pragmatics perspective, the learning that gets achieved in

a new language is necessarily culturally and situationally specific as well as inseparable from authentic communication, and authentic communication encompasses both the face-to-face and digital worlds of people. Thus, if in the future scholars invest effort into elucidating (1) the relationship between pragmatics-related demands of tasks and cognitive task complexity, and (2) the L2 pragmatics needs that new tasks and new technologies create, we will be closer to understanding how to design TBLT curricula that sample the full range of social actions people, and particular groups of students, need to carry out in a new language, from the most mundane to the most task-like.

We have also argued, and the research has amply shown, that the unique edge of technology-mediated tasks for the learning of pragmatics is that they greatly improve on what is possible with traditional classroom materials. They can be designed so as to afford classroom learners some of the necessary interactional and social contexts for the development of L2 pragmatic competence. A focus on pragmatics as part of the language required to accomplish a task would help learners succeed in interaction and create and maintain the rapport necessary to continue future interactions and cultivate social relations. When these tasks leverage the mediation of new technologies, the pragmatic specifications may become even more complex, but the tasks will also become all the more authentic and motivating, hence supporting rich pragmatic learning, and rich overall language learning.

References

Alcón Soler, E. (2002). Relationship between teacher-led versus learners' interaction and the development of pragmatics in the EFL classroom. *International Journal of Educational Research*, 37, 359–377. doi:10.1016/S0883-0355(03)00010-7

Adams, R., & Alwi, N. A. N. M. (2014). Prior knowledge and second language task production in text chat. In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT: Researching technology and tasks* (pp. 51–78). Amsterdam: John Benjamins.

Azkarai, A. (2015). Males and females in EFL task-based interaction: Does gender have an impact on LREs? *Vigo International Journal of Applied Linguistics*, 12, 9–35.

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. doi:10.1017/S0272263113000429

Baralt, M. (2014). Task complexity and task sequencing in traditional versus online language classes. In M. Baralt, R. Gilabert, & P. J. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 59–122). London: Bloomsbury Academic.

Belz, J. A. (2005). Intercultural questioning, discovery and tension in internet-mediated language learning partnerships. *Language and Intercultural Communication*, 5, 1–37. doi:10.1080/14708470508668881

Belz, J. A., & Kinginger, C. (2002). The Cross-linguistic development of address form use in telecollaborative language learning: Two case studies. *The Canadian Modern Language Review*, 59, 189–214. doi:10.3138/cmlr.59.2.189

Belz, J. A., & Kinginger, C. (2003). Discourse options and the development of pragmatic competence by classroom learners of German: The case of address forms. *Language Learning*, 53(4), 591. doi: 10.1046/j.1467-9922.2003.00238.x

Blattner, G., & Fiori, M. (2011). Virtual social network communities: An investigation of language learners' development of sociopragmatic awareness and multiliteracy skills. *CALICO Journal*, 29, 24–43. doi: 10.11139/cj.29.1.24-43

Breen, M. (1989). The evaluation cycle for language learning tasks. In R. Johnson (Ed.), *The second language curriculum*. Cambridge: Cambridge University Press.
doi: 10.1017/CBO9781139524520.014

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

Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Clevedon: Multilingual Matters.

Canto, S., de Graff, R., & Jauregui, K. (2014). Collaborative tasks for negotiation of intercultural meaning in virtual worlds and video-web communication. In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT*: *Researching technology and tasks* (pp. 183–212). Amsterdam: John Benjamins.

Chapelle, C. A. (2014). Afterword: Technology-mediated TBLT and the evolving role of the innovator. In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT*: *Researching technology and tasks* (pp. 323–334). Amsterdam: John Benjamins.

Çiftçi, E. Y. (2016). A review of research on intercultural learning through computer-based digital technologies. *Educational Technology & Society*, 19, 313–327.

Collentine, K. (2010). Measuring complexity in task-based synchronous computer-mediated communication. In M. Thomas & H. Reinders (Eds.), *Task-based language learning and teaching with technology* (pp. 105–128). London: Continuum.

Collentine, K. (2013). Using tracking technologies to study the effects of linguistic complexity in CALL input and SCMC output. *CALICO Journal*, 30, 46–65.

Cunningham, D. J. (2016). Request modification in synchronous computer-mediated communication: The role of focused instruction. *Modern Language Journal*, 100, 484–507. doi: 10.1111/modl.12332

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19, 221–246. doi: 10.1111/j.1473-4192.2009.00231.x

Frenz-Belkin, P. (2015). 'Teacher! You need to give me back my homework:' Assessing students' needs for a pragmatics curriculum in an academic ESL program. In S. Gesuato, F. Bianchi, & W. Cheng (Eds.), *Teaching, learning and investigating pragmatics: Principles, methods and practices* (pp. 33–56). Newcastle upon Tyne: Cambridge Scholars.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, 34, 65–87. doi: 10.1111/j.1467-1770.1984.tb00996.x

Gibbs, R. W. (2001). Intentions as emergent products of social interactions. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 105–124). Cambridge, MA: The MIT Press.

Gilabert, R., & Barón, J. (2018). Independently measuring cognitive complexity for task sequencing and interlanguage pragmatics development. In N. Taguchi & Y. Kim (Eds.), *Task-based approaches to teaching and assessing pragmatics*. Amsterdam: John Benjamins. (this volume) doi: 10.1075/tblt.10.07gil

Golato, A., & Taleghani-Nikazm, C. (2006). Negotiation of face in chats. *Multilingua*, 25, 293–322. doi: 10.1515/MULTI.2006.017

González-Lloret, M. (2008). Computer-mediated learning of L2 pragmatics. In E. A. Soler & A. Martinez-Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching and testing* (pp. 114–132). Clevedon: Multilingual Matters.

González-Lloret, M. (2011). Conversation analysis of computer-mediated communication. *CALICO Journal*, 28, 308–325. doi:10.11139/cj.28.2.308-325

González-Lloret, M. (2014). The need for needs analysis in technology-mediated TBLT. In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT: Researching technology and tasks* (pp. 23–50). Amsterdam: John Benjamins.

González-Lloret, M., & Ortega, L. (Eds.). (2014). *Technology-mediated TBLT: Researching technology and tasks*. Amsterdam: John Benjamins.

Hassall, T. (2015). Individual variation in L2 study-abroad outcomes: A case study from Indonesian pragmatics. *Multilingua*, 34, 33–59.

Hertel, K. (2015). Integrating American English pragmatic instruction in tourism training programs in Latin America: A materials portfolio (MA). University of Oregon. Retrieved from <http://linguistics.uoregon.edu/wp-content/uploads/2015/09/Hertel-Kelsey-Aug-15.pdf>

Holden, C., & Sykes, J. M. (2012). Mentira: Prototyping language-based locative gameplay. In S. Dikkers, J. Martin, & B. Coulter (Eds.), *Mobile media learning: Amazing uses of mobile devices for teaching and learning* (pp. 111–131). Pittsburg, PA: ETC Press.

Ishihara, N., & Tarone, E. (2009). Subjectivity and pragmatic choice in L2 Japanese: Emulating and resisting pragmatic norms. In N. Taguchi (Ed.), *Pragmatic competence in Japanese as a second language* (pp. 101–128). Berlin: Mouton de Gruyter.

Itō, M., Baumer, S., Bittanti, M., & Antin, J. (Eds.). (2010). *Hanging out, messing around, and geeking out: Kids living and learning with new media*. Cambridge, MA: The MIT Press.

Jauregi, K., & Canto, S. (in press). Mundos virtuales en la enseñanza de lenguas: Hacia un aprendizaje significativo a través de la interacción, la acción y el juego. In M. González-Lloret & M. Vinagre (Eds.), *Comunicación mediada por tecnologías: Aprendizaje y enseñanza de la lengua extranjera*. Sheffield: Equinox.

Jauregi, K., Canto, S., de Graaff, R., Koenraad, T., & Moonen, M. (2011). Verbal interaction in Second Life: Towards a pedagogic framework for task design. *Computer Assisted Language Learning*, 24, 77–101. doi:10.1080/09588221.2010.538699

Jefferson, G. (1988). On the sequential organization of troubles-talk in ordinary conversation. *Social Problems*, 35, 418–441. doi:10.2307/800595

Jenks, C. J., & Brandt, A. (2013). Managing mutual orientation in the absence of physical copresence: Multiparty voice-based chat room interaction. *Discourse Processes*, 50, 227–248. doi:10.1080/0163853X.2013.777561

Jeon, E. H., & Kaya, T. (2006). Effects of L2 instruction on interlanguage pragmatic development: A meta-analysis. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 165–211). Amsterdam: John Benjamins. doi:10.1075/lllt.13.10jeo

Kasper, G. (1997). Can pragmatic competence be taught? *(NetWork #6) Second Language Teaching and Curriculum Center*, University of Hawai'I, Retrieved August 20, 2004, from http://www.nflrc.hawaii.edu/NetWorks/NW06/.

Kasper, G. (2009). L2 pragmatic development. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (2nd ed., pp. 259–284). Bingley: Emerald.

Kim, E. -Y., & Brown, L. (2014). Negotiating pragmatic competence in computer mediated communication: The case of Korean address terms. *CALICO*, 31, 264–284 doi:10.11139/cj.31.3.264-284

Kim, H. Y. (2014). Learner investment, identity, and resistance to second language pragmatic norms. *System*, 45, 92–102. doi:10.1016/j.system.2014.05.002

Kim, Y., & Taguchi, N. (2016). Learner-learner interaction during collaborative pragmatic tasks: The role of cognitive and pragmatic task demands. *Foreign Language Annals*, 49, 42–57. doi:10.1111/flan.12180

Kinginger, C., & Belz, J. A. (2005). Sociocultural perspectives on pragmatic development in foreign language learning: Case studies from telecollaboration and study abroad. *Interlanguage Pragmatics*, 2(4), 369–421.

Lakoff, R. (1973). The logic of politeness; or, minding your P's and Q's. In C. Corum, T. Cedric Smith-Stark, & A. Weiser (Eds.), *Papers from the ninth regional meeting of the Chicago Linguistics Society* (pp. 292–305). Chicago, IL: Department of Linguistics, University of Chicago.

Long, M. H. (2016). In defense of tasks and TBLT: Nonissues and real issues. *Annual Review of Applied Linguistics*, 36, 5–33.

Markee, N. (2013). Emic and etic in qualitative research. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Malden, MA: Blackwell. doi:10.1002/9781405198431.wbeal0366

Marsh, L., & Onof, C. (2008). Stigmergic epistemology, stigmergic cognition. *Cognitive Systems Research*, 9(1–2), 136–149. doi:10.1016/j.cogsys.2007.06.009

Nik, A. N. M. A., Adams, R., & Newton, J. (2012). Writing to learn via text chat: Task implementation and focus on form. *Journal of Second Language Writing*, 21, 23–39. doi:10.1016/j.jslw.2011.12.001

Norris, J. M. (2009). Task-based teaching and testing. In M. H. Long & C. J. Doughty (Eds.), *Handbook of language teaching* (pp. 578–594). Malden, MA: Wiley-Blackwell. doi:10.1002/9781444315783.ch30

O'Dowd, R., & Ware, P. (2009). Critical issues in telecollaborative task design. *Computer Assisted Language Learning*, 22, 173–188. doi:10.1080/09588220902778369

Ortega, L., & González-Lloret, M. (2015). Staking out the territory of technology-mediated TBLT. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (pp. 59–86). Amsterdam: John Benjamins. doi:10.1075/tblt.8.03ort

Pekarek Doehler, S., & Pochon-Berger, E. (2015). The development of L2 interactional competence: Evidence from turn-taking organization, sequence organization, repair organization and preference organization. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 233–268). Berlin: Mouton de Gruyter. doi:10.1515/9783110378528-012

Piirainen-Marsh, A. (2011). Enacting interactional competence in gaming activities: Coproducing talk with virtual others. In J. K. Hall, J. Hellermann, & S. P. Doehler (Eds.), *L2 interactional competence and development* (pp. 19–44). Bristol: Multilingual Matters.

Piirainen-Marsh, A. P., & Tainio, L. (2009). Other-repetition as a resource for participation in the activity of playing a video game. *The Modern Language Journal*, 93(2), 153–169. doi:10.1111/j.1540-4781.2009.00853.x

Piirainen-Marsh, A. P., & Tainio, L. (2014). Asymmetries of knowledge and epistemic change in social gaming interaction. *The Modern Language Journal*, 98, 1022–1038. doi:10.1111/modl.12153

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), *Researching task complexity: Task demands, task-based language learning and performance* (pp. 3–38). Amsterdam: John Benjamins.

Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28, 463–481. doi:10.1177/0265532210394633

Roever, C. (2013). Technology and tests of L2 pragmatics. In N. Taguchi & J. M. Sykes (Eds.), *Technology in interlanguage pragmatics research and teaching* (pp. 215–234). Amsterdam: John Benjamins. doi:10.1075/lllt.36.11roe

Rose, K. R. (2005). On the effects of instruction in second language pragmatics. *System*, 33, 385–399. doi:10.1016/j.system.2005.06.003

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Houndmills: Palgrave Macmillan.

Schegloff, E. (2007). *Sequence organization in interaction*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511791208

Siegal, M. (1996). The role of learner subjectivity in second language linguistic competency: Western women learning Japanese. *Applied Linguistics*, 17, 356–382. doi:10.1093/applin/17.3.356

Skehan, P. (2003). Focus on form, tasks, and technology. *Computer Assisted Language Learning*, 16, 391–411. doi:10.1076/call.16.5.391.29489

Spencer-Oatey, H., & Jiang, W. (2003). Explaining cross-cultural pragmatic findings: Moving from politeness maxims to sociopragmatic interactional principles (SIPs). *Journal of Pragmatics*, 35(10–11), 1633–1650. doi:10.1016/S0378-2166(03)00025-0

Sperber, D., & Wilson, D. (1986/2001). *Relevance: Communication and cognition*. Oxford: Blackwell.

Stommel, W. (2008). Conversation analysis and community of practice as approaches to studying online community. *Language@Internet*, 5. Retrieved from <http://www.languageatinternet.de/articles/2008/1537>

Sykes, J. (2008). *A dynamic approach to social interaction: Synthetic immersive environments & spanish pragmatics*. Unpublished PhD dissertation, University of Minnesota. Retrieved from ProQuest Dissertations & Theses Global.

Sykes, J. M. (2014). TBLT and synthetic immersive environments: What can in-game task restarts tell us about design and implementation? In M. González-Lloret & L. Ortega (Eds.), *Technology-mediated TBLT: Researching technology and tasks* (pp. 149–182). Amsterdam: John Benjamins. doi:10.1075/tblt.6.06syk

Taguchi, N., Kaufer, D., Gómez-Laich, M. P., & Zhao, H. (2016). A corpus linguistics analysis of on-line peer commentary. In K. Bardovi-Harlig & J. C. Félix-Brasdefer (Eds.), *Pragmatics and language learning* (pp. 357–370). Honolulu, HI: University of Hawai'i, National Foreign Language Resource Center.

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4, 91–112. doi:10.1093/applin/4.2.91

Thorne, S. L. (2003). Artifacts and cultures-of-use in intercultural communication. *Language Learning & Technology*, 7(2), 38–67.

Thorne, S. L. (2016). Cultures-of-use and morphologies of communicative action. *Language Learning & Technology*, 20(2), 185–191.

Timpe-Laughlin, V. (2018). Pragmatics in task-based language assessment: Opportunities and challenges. In N. Taguchi & Y. Kim (Eds.), *Task-based approaches to teaching and assessing pragmatics*. Amsterdam: John Benjamins. (this volume) doi:10.1075/tblt.10.07gil

Tsai, M. -H., & Kinginger, C. (2014). Giving and receiving advice in computer-mediated peer response activities. *CALICO Journal*, 32, 82–112.

Van den Branden, K. (2016). The role of teachers in task-based language education. *Annual Review of Applied Linguistics*, 36, 164–181.

Vinagre, M. (2010). Intercultural learning in asynchronous telecollaborative exchanges: A case study. *The Eurocall Review*, 17. Available at: <http://www.eurocall-languages.org/uploaded/EUROCALL_Review/review17.pdf>.

Walther, J. B., Van der Heide, B., Ramirez, A., Burgoon, J. K., & Peña, J. (2015). Interpersonal and hyperpersonal dimensions of computer-mediated communication. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 3–22). Malden, MA: Wiley-Blackwell.

Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, 32, 199–225. doi:10.1177/0265532214557113

Youn, S. J. (2018). Task design and validity evidence for assessment of L2 pragmatics in interaction. In N. Taguchi & Y. Kim (Eds.), *Task-based approaches to teaching and assessing pragmatics*. Amsterdam: John Benjamins. (this volume) doi:10.1075/tblt.10.07gil

Ziegler, N. (2016). Taking Technology to Task: Technology-Mediated TBLT, Performance, and Production. *Annual Review of Applied Linguistics*, 36, 136–163.
doi:10.1017/S0267190516000039

Task-based assessment of pragmatics

Task design and validity evidence for assessment of L2 pragmatics in interaction

Soo Jung Youn
Northern Arizona University

This study examines how clear communicative goals and authentic interaction are ensured in task-based pragmatic assessment practices, particularly in designing role-play assessment tasks and developing task-appropriate rating criteria. Employing a concurrent mixed methods design, conversation analysis (CA) and multi-faceted Rasch measurement were used to investigate whether task-independent interactional features are elicited from role-play interactions and how interaction-specific rating categories function quantitatively. The CA findings indicated that various interactional organizations emerged from role-play task interactions as examinees oriented to the different degree of imposition and contextual variables embedded in the role-plays. Further, the interactional features included in the rating criteria created a substantial amount of variance in distinguishing varying degrees of pragmatic performance. The findings are discussed in terms of designing valid pragmatic assessment tasks and what features need to be included in developing rating criteria to ensure the validity evidence of task-based pragmatic assessment.

Introduction

An increased interest in assessing second language (L2) pragmatic competence has been observed during the last several decades (e.g., Hudson, Detmer, & Brown, 1992, 1995; Grabowski 2013; Liu, 2007; Roever, 2006, 2011; Ross & Kasper, 2013; Walters, 2007; Youn, 2015; for a review see Taguchi & Roever, 2017). However, the task-based approach to assessing L2 pragmatics is still scarce except for a few notable exceptions (e.g., Youn, 2015). This study reports how clear communicative goals and meaningful interaction can be ensured in task-based pragmatic assessment practices, focusing on designing assessment tasks and task-appropriate rating criteria. Because research on task-based pragmatic assessment is still limited, challenges remain regarding how to ensure valid assessment practices. Particularly challenging is to ensure construct validity, such as eliciting language phenomena appropriate to the construct definition of L2 pragmatics.

At the heart of these issues, the test design and scoring processes that are appropriate to the target construct involve eliciting key characteristics of observed language performances. This process is referred to as ensuring the *evaluation* inference in an argument-based approach to validity (Chapelle, 2012; Kane, 1992, 2006). Considering the wide range of pragmatic knowledge required for successful task-based pragmatic performance (i.e., across different communicative tasks), the evaluation inference becomes more complex, involving several steps towards collecting valid test scores, such as developing valid task-appropriate rating criteria and training raters in awarding scores conforming to descriptions in rating criteria. Consequently, we face various challenges in order to ensure validity evidence for the evaluation inference in assessing pragmatic competence in interaction, that is, a learner's ability to accomplish various pragmatic actions in interaction.

Among various dimensions of pragmatic knowledge, this study focuses on types of interactional organization required to achieve pragmatic actions in interaction. The purpose of this study is to explicate the nature of interactional phenomena elicited from role-play-based pragmatic assessment tasks. I argue that various features involved in role-play task design ensure elicitation of meaningful pragmatic interactional features. The study also examines whether the interactional organizations reflected in rating criteria function well to differentiate examinees' abilities. These issues are closely related to ensuring the evaluation inference for task-based pragmatic assessment.

This study employed a concurrent mixed methods design (Teddlie & Tashakkori, 2006) by combining conversation analysis (CA) and multi-faceted Rasch measurement (MFRM). First, I examined various types of interactional organization that emerged in role-play interaction from the perspective of CA. In tandem with the CA findings, using MFRM, I further investigated whether rating categories reflecting interactional phenomena functioned well in differentiating 102 examinees into a wide range of levels of ability. This paper discusses how findings serve as the backing for various assumptions underlying the evaluation inference. In the following literature review section, I will first briefly discuss the CA approach to defining and assessing pragmatic competence in interaction along with various validity challenges. I will then discuss role-play assessment task design and eliciting validity evidence for L2 pragmatic assessment.

Background

Assessing pragmatic competence in interaction and validity challenges

L2 pragmatics has been conceptualized under various models of communicative competence (e.g., Celce-Murcia, 2007; Celce-Murcia, Dörnyei, & Thurrell, 1995),

and ontologically distinct theoretical frameworks have been available in defining L2 pragmatics (Kasper, 2009). Among them, this study adopts the discursive approach to pragmatics (Kasper, 2006), which emphasizes understanding how pragmatic meaning and actions are accomplished in a sequential organization of participants' turns-at-talk from the CA perspective. In this study, I define pragmatic competence in interaction as the abilities of achieving various pragmatic meanings and actions jointly in organized sequences by employing a wide range of pragmatic and interactional resources. Based on this definition, this study focuses on the interactional organizations that examinees use in achieving pragmatic actions. In addition, this study follows the assumption that distinct sequential organizations and interactional features are indicative of various levels of pragmatic competence in interaction (e.g., Al-Gahtani & Roever, 2012, 2014; Kasper & Ross, 2013; Youn, 2015).

Although pragmatic competence in interaction is grounded in a strong theoretical and analytical framework, actual assessment of task-based pragmatic performance faces various validity issues. Validity in language testing and assessment has traditionally been defined as investigating whether a test "measures what it is intended to measure" (Hughes, 1989, p. 22). However, as the nature of targeted constructs and intended score uses become more complex, approaches to validity have continued to evolve. One of the recent validation frameworks is an argument-based approach, which requires researchers to specify various inferences entailed in score interpretation and use and then collect evidence supporting or countering those inferences within a given assessment (for a review, see Chapelle, Enright, & Jamieson, 2008; Kane, 2006).

Among various inferences, the *evaluation* inference refers to the detailed procedures used to obtain scores from observed performances (Kane, 1992, 2006). The *evaluation* inference is fundamental to task-based performance assessment, as ways in which task-based performances are scored essentially determine whether obtained scores are valid for given assessment purposes. At least two assumptions need to be examined for the evaluation inference: (a) assessment tasks do elicit valid features of the targeted construct and (b) rating criteria function appropriately in reflecting meaningful levels of construct ability. Satisfying these assumptions is particularly challenging for task-based pragmatic assessment for the following reasons. Regarding the first assumption related to assessment tasks, it is crucial to design tasks in which examinees actually accomplish a pragmatic action with an explicit communicative goal. In addition, elicited pragmatic performance needs to be reflective of authentic language use. These are not necessarily guaranteed in commonly employed speaking task types. For example, language proficiency interviews, such as that used by IELTS, a widely employed speaking test format, may not be appropriate to assess task-based pragmatic performance, as examinees mainly have to respond to an interviewer's scripted questions rather than accomplishing a communicative goal.

With regard to the second assumption related to rating criteria, developing criteria that measure L2 pragmatic competence involves various challenges. This is because rating criteria in task-based assessment have to function in several ways, that is, generating a substantial amount of variance in differentiating examinees' abilities while tapping into various dimensions of the targeted construct. Since task-based performances are usually elicited from a small number of assessment tasks, variance has to be ensured from well-designed rating criteria rather than from assessment tasks alone. When considering the construct of pragmatic competence in interaction, isolating concrete interactional resources (e.g., turn-taking devices) that an examinee employs requires careful attention in relation to intended interpretations and uses of a test (Kasper & Ross, 2013), as these resources essentially furnish the descriptions in rating criteria.

This study intends to respond to these validity challenges in order to help strengthen the evaluation inference of assessing pragmatic competence in interaction. In terms of the assessment task type, this study uses role-plays, a commonly adopted instrument in L2 pragmatic research at large. The next section discusses the characteristics of role-plays and what it means to elicit validity evidence from role-play interaction.

Role-plays as assessment tasks and eliciting validity evidence

In L2 pragmatics research, role-plays have been widely employed in eliciting pragmatic performance (Kasper & Ross, 2002; Taguchi & Roever, 2017). Role-plays enable participants to take on certain social roles within a predetermined scenario. Thus, using role-plays, we can create various opportunities to tap into the wide range of pragmatic knowledge. In addition, compared to naturally occurring interaction, the controlled nature of role-play interaction adds to advantage because role-play performance can be comparable across participants. For these reasons, role-plays have been adopted widely as a test instrument in pragmatic assessment research (Hudson, Detmer, & Brown, 1992, 1995).

However, role-plays as assessment tasks, from the task-based perspective, need to be critically examined and improved in many ways. While the role-play design is widely understood, the role-play itself does not necessarily satisfy the characteristics of a task defined in task-based language teaching and assessment research (e.g., Norris, 2016). For example, participants can simply engage in role-play interaction for small talk or a telephone opening conversation that do not necessarily require a clear communicative goal. In addition, if participants share a role-play script and anticipate what another person will say, authentic interaction and meaningful negotiation of meaning cannot be ensured. In short, the role-play cannot be automatically categorized as an authentic task, unless a clear communicative goal is specified with careful task design.

Another concern of using role-plays in pragmatic assessment includes the challenge of eliciting interaction that reflects some degree of authenticity. For example, one of the limitations in the role-plays designed in Hudson, Detmer, and Brown's (1992, 1995) pragmatic test battery was that interactional outcomes were imposed and shared among examinees, which does not reflect the nature of authentic interaction. In addition, lack of real-life social consequences coming from examinees' performance and lack of generalizability of learners' performance across research contexts have been voiced as limitations of role-plays (e.g., Bardovi-Harlig & Hartford, 2005). Such concerns attached to role-plays may jeopardize the quality of validity evidence, particularly in terms of the characteristics of elicited interaction.

Responding to this concern, the capacity of eliciting authentic interactional behaviors via role-plays has been increasingly investigated from the CA perspective (e.g., Huth, 2010; Kasper & Youn, 2017; Okada, 2010; Okada & Greer, 2013). These studies show that the elicited role-play interactions exhibit the features of real-life interaction. Nonetheless, what remains unknown is how role-play task design ensures meaningful pragmatic interactional features and whether the types of interactional organization observed in role-play-based pragmatic performance can be generalizable across different role-play situations. Thus, we need additional validity evidence related to the nature of the interactional features emerging from role-play interaction in order to strengthen the validity argument for task-based pragmatics assessment in the context of spoken interaction.

Even if role-plays can elicit valid interactional behaviors, the question of how the role-play performance is assessed requires its own attention for assessment purposes. The role of task-appropriate rating criteria is essential for not only explaining the targeted construct but also for helping raters in awarding scores consistently corresponding to various characteristics of elicited performance. In the previous pragmatic assessment research, relatively little attention has been paid to developing valid rating criteria reflective of concrete features of role-play-based pragmatic performance. To fill this gap, Youn (2015) developed interaction-sensitive rating criteria that reflect various features of role-play-based pragmatic performance, such as requesting a recommendation letter from a professor or responding to a professor's request during office-hour interaction. The rating criteria included five categories that reflect a wide range of pragmatic knowledge based on the qualitative analysis of sample role-play performances. Using the rating criteria, trained raters were able to assess learners' role-play performances in a satisfactory manner. Of the five rating categories, three categories (*Sensitivity to Situation, Turn Organization*, and *Engaging with Interaction*) were closely related to various types of interactional organization critical to achieving pragmatic actions successfully, while *Language Use* and *Content Delivery* categories were more related to an examinee's grammar knowledge and fluent content delivery. Despite their satisfactory quality, what remains unknown is whether the rating

categories that are specifically related to interactional organizations function well in distinguishing among varying levels of targeted performance. Therefore, additional backing necessary for ensuring the quality of rating criteria is the question of whether the interaction-specific rating categories alone can generate a substantial amount of variance. Such information will help us further argue for the importance of considering interactional features for task-based pragmatic assessment practices. As a follow-up study of Youn (2015), one possible way to address this issue is to compare the function of two types of rating criteria, one with all five categories and the other with only three categories related to interactional organizations (i.e., Sensitivity to Situation, Engaging with Interaction, Turn Organization).

In summary, it is clear that task design critically influences the nature of elicited language performance and that task-appropriate rating criteria should generate the amount of variance that appropriately differentiates examinees' abilities. These two aspects together contribute to the evaluation inference. Because this study focuses on examinees' abilities to interact with each other in achieving pragmatic actions, it is critical to identify various interactional phenomena that serve as validity evidence of pragmatic competence in interaction and to examine how those phenomena are generalizable across different role-play situations. Given the concerns attached to the role-play interaction (e.g., lack of a clear communicative goal and limited authenticity), the nature of interactional phenomena elicited from role-plays needs to be clearly explicated to enhance our understanding of what role-plays can and cannot assess validly. Another validity challenge is to examine the functioning of interactional features included in rating criteria that are supposed to function in differentiating examinees' abilities. Thus, to obtain validity evidence, this study examines the role of both task design and rating criteria in assessing L2 pragmatic competence in interaction.

The present study

In this study, I examined interactional phenomena that emerged from examinees' pragmatic performances in role-play assessment tasks. As validity evidence, I investigated the extent to which various features embedded in role-play task design elicit meaningful pragmatic interactional features and whether types of interactional organization are generalizable across role-plays. I also investigated whether or not rating criteria can produce variance in discriminating among examinees' role-play performances. Using a concurrent mixed methods design, CA was used to explicate the types of interactional organization emerging from role-play tasks. In order to address the functionality of the rating criteria and the degree of variance, multi-faceted Rasch measurement (MFRM), a commonly employed method in performance assessment research, was used. The following research questions guided the present study:

1. In what ways do role-play tasks elicit interactional features of interaction-involved pragmatic performance?
2. In what ways are the interactional features of observed task-based performances reflective of a varying level of pragmatic competence in interaction?

Methods

Participants

Examinees
102 ESL students enrolled in US universities voluntarily participated as examinees. Their TOEFL iBT scores ranged from 65 to 111 and their L1s included Chinese, Korean, Japanese, and Indonesian. They consisted of almost an equal number of graduate and undergraduate students. Of the 102 examinees, 70% were females and 30% were males.

Interlocutors
For the role-play situations that involved a professor interlocutor, four native speakers who were familiar with university-level academic advising participated in the role of professors. All four interlocutors received training for standardized conversation with the 102 examinees. For the role-plays with a classmate, two examinees were paired together to play the role of two classmates. These two examinees only interacted with each other during the test.

Raters
12 trained raters scored examinees' role-play task performances using analytic rating criteria. Of the 12 raters, seven were either native English speakers or bilingual English speakers; five raters were non-native English speakers with a high level of academic English ability. Non-native raters' L1s included Chinese, Japanese, Korean, and Vietnamese. All of them had two to five years of ESL teaching experience.

Test instruments

Role-plays
Five role-play tasks that involved two interlocutors (professor, classmate) and two speech acts (request, refusal) with a different degree of imposition were developed and completed by 102 ESL learners. Table 1 summarizes the role-play scenarios (see Appendix A for an example).

Table 1. Descriptions of five role-play tasks

Role-play	Description
Role-play with a professor during the office hour	1. Requesting a recommendation letter for scholarship application with a short letter due date 2. Requesting additional advising time regarding a class project 3. Responding to a professor's request to change a presentation schedule
Role-play with a classmate working on a class project	4. Deciding an agreeable meeting time when the third group member is absent 5. Deciding a meeting mode (face-to-face vs. online discussion) to discuss an upcoming group project

In consideration of the guidelines that qualify tasks from a task-based educational perspective (e.g., Norris, 2009), the following characteristics justify the role-plays developed in this study as tasks. First, the role-play situations reflected learners' real-life pragmatic needs in an academic context based on the findings from a large-scale pragmatic needs analysis in an academic context (Youn, 2010). For example, Youn found that one of the prominent spoken language use situations in an academic context was interacting with a professor to achieve specific communicative goals (e.g., asking for a recommendation letter request, refusing professor's request) during office-hour interaction. In addition to the student-professor role-plays, role-plays involving interactions with a classmate were also developed, because such a language use situation (e.g., expressing disagreement with a classmate) was identified as an important pragmatic task in Youn's needs analysis. The role-plays then reflected a wide range of contextual variables (e.g., different degrees of imposition required in making requests to a professor) and various speech acts (i.e., request, refusal, agreement, disagreement). Secondly, in order to elicit meaningful interaction, examinees each held a different portion of information in completing the role-plays, and the information was not shared between them (see Appendix A for an example). In this way, although the role-play interaction is simulated, examinees are expected to engage in authentic interaction employing various interactional organizations. For example, in one of the role-plays with a professor, examinees did not expect to hear a professor's conference trip schedule, which affects the likelihood of timely recommendation letter submission. In this way, the ways in which examinees respond to an unexpected situation were captured in a turn-taking structure, which serves as a ratable sequence. Finally, one acceptable communicative goal was specified in each role-play situation, rather than providing the range of acceptable outcomes, so the role-play performances from examinees were comparable.

Rating criteria
In order to score examinees' role-play performances, data-driven, interaction-sensitive rating criteria were developed based on the qualitative analyses of examinees' role-play

performances (Youn, 2015) (see Appendix B). As summarized in Table 2, the analytical rating criteria included five rating categories on a three-point scale, of which the last three categories addressed interaction-specific features. Youn (2015) reported that all five rating categories functioned well statistically in distinguishing examinees' abilities.

Table 2. Descriptions of five rating categories

Rating category	Description
Contents Delivery	The ability to deliver an individual turn clearly and fluently
Language Use	The ability to employ various linguistic resources in achieving various pragmatic actions
Sensitivity to Situation	The ability to provide appropriate reasons and explanations when achieving pragmatic actions in appropriate sequential organizations
Engaging with Interaction	The ability to engage in interaction and to establish a shared understanding when interacting with a classmate and a professor
Turn Organization	The ability to take turns following turn-taking conventions in achieving pragmatic actions.

Procedures

Test administration

Examinees completed all role-plays in two separate sessions, and their performances were audio-recorded. The first session involved two examinees completing the role-play with a classmate; the second session involved an examinee completing the role-play with a professor interlocutor. It took each examinee about one hour to complete all role-plays.

Rater training

All 12 raters completed extensive training sessions with the researcher. The training session involved three sequential steps: (a) familiarize raters with assessment tasks and the descriptions in the rating criteria; (b) norm raters so they develop a shared understanding of three distinct levels of role-play performances corresponding to the three-point scale descriptions in the rating criteria; (c) have raters practice scoring and discuss tricky cases, such as borderline performances. Various training materials (i.e., a rating manual and CA transcripts) that described various linguistic and interactional features in the rating criteria descriptions were prepared.

Rating

The trained raters completed ratings alone. They were asked to score the examinees' performances on one role-play and then move on to the next one. For logistical

reasons, the twelve raters were randomly divided into four rater groups, with each group scoring a different set of performance data. All raters scored at least 30 examinees' role-play performances of varied proficiency levels, which serve as anchored data, a necessary condition for the MFRM analysis (Schumacker, 1999).

Data analysis

Conversation analysis

For the first research question, 20 examinees' role-play performances randomly selected from all role-play situations were analyzed by using the conversation-analytic (CA) method to examine how the examinees employed various interactional resources and oriented to the normative nature of talk-in-interaction to achieve pragmatic actions. Due to page limitations here, only 12 participants' data are presented in this study. The role-play performance data was transcribed according to the CA notation system (see Appendix C).

Multi-faceted Rasch measurement

For the second research question, whether two types of rating criteria, one that reflects all five rating categories developed in Youn (2015) and the other that included three interaction-specific rating categories only (this study), functioned similarly were examined using multi-faceted Rasch measurement (MFRM) (Bond & Fox, 2007; McNamara, 1996). The computer program FACETS, version 3.71.4 (Linacre, 2014), was employed for the MFRM analysis. MFRM has been commonly used in performance assessment (McNamara & Knoch, 2012), and its use has been advocated to ensure the validity of research instruments in applied linguistic research in general (Purpura, Brown, & Schoonen, 2015). Rather than relying on raw scores that do not account for different degrees of raters' severity and task difficulty, MFRM offers a more robust method for analyzing the characteristics of various facets involved in performance assessment (e.g., examinee ability, task difficulty, raters' severity level, rating category difficulty). Given that the original rating criteria functioned well (Youn, 2015), I compared two types of rating criteria (five rating categories combined vs. three categories with interaction-specific criteria) to see whether or not they functioned similarly. Thus, in this study, two separate MFRM analyses corresponding to each rating criteria type were conducted and the findings were compared.

Results

The CA findings focusing on various types of interactional organization are presented first, followed by the MFRM findings on how two types of rating criteria functioned in estimating 102 examinees' abilities of pragmatic competence in interaction.

Conversation Analysis (CA) findings

This section presents various types of interactional organization that emerged across all five role-play situations at various levels of examinees to answer the first research question. In particular, ways in which the features considered in role-play task design ensure the elicitation of meaningful pragmatic interaction were examined.

Recommendation letter request role-play

Excerpt 1 presents a typical sequence found in higher-level examinees' performances on making a recommendation letter request role-play. The greeting exchange in lines 1 to 4 is first presented. Then, P (Professor) orients to the institutional nature of this conversation by projecting *what can I do for you* in line 4, inviting J (Jessie, a name given to the examinee) to project the course of actions. With the prefaces of *uhm* and *actually*, J first projects the pre-sequence *I have uhm little of big favor for you* before the main action. The pre-sequence refers to a sequence that comes before a specific pair type, such as request or invitation, which projects the contingent possibility of the base sequence (Sacks, 1992; Schegloff, 2007). Here, J's pre-sequence serves as an alert to the listener that a request is to follow. Here, the pre-request creates the room for J to explain the background information before launching a request sequence in lines 10 to 11. By doing so, J orients to the higher degree of imposition involved in requesting a recommendation letter from a professor. After a 0.8 second pause in line 12, P completes the turn construction unit by accepting J's request in line 13.

Excerpt 1.
P: Professor, J: Jessie (Examinee ID46, High-level)

```
 1    J:   hi
 2    P:   hi how are you
 3    J:   I'm good how are you?
 4    P:   good thank you what can I do for you today?
 5    J:   uhm I actually have uh uh uhm little of a
 6         big favor for you (.) [uhm (.) I'm: uh applying for this
 7    P:                         [°mmhmm°
 8    J:   uhm department scholarship↑ (.) [and uhm I need a (.)
 9    P:                                   [°mmhmm°
10    J:   letter of recommendation and I was wondering if you
11         are .hh uhm able to write one for me
12         (0.8)
13    P:   sure I'd be happy to write the letter for you=
```

Excerpt 2 demonstrates another high-level examinee's performance, but this time the quality of adjacency pair is not ensured. The adjacency pair is the basic unit of sequence organization, which is composed of two ordered turns produced by different speakers (Schegloff, 2007). The first pair part projects a prospective action, which

limits possible options for the second pair parts. For example, the second pair parts of an invitation sequence are typically limited to either an acceptance or a decline. The appropriateness of the second pair part is displayed in the next turn. In lines 6 to 9, J projects an account and begins a request sequence. While P accepts J's request with no inter-turn delay in line 10, what J says in the next turn is *okay* (line 11). The expected second pair part to the acceptance of request (i.e., *thanking*) is absent here. The 1.2 second pause in line 12 indicates that P orients to the absence of the expected second pair part (thanking). P then initiates another sequence by asking when the letter is due (lines 13 to 14), followed by J's response in line 15.

Excerpt 2.
P: Professor, J: Jessie (Examinee ID64, High-level)

```
 6    J:   uhm I need a recommendation letter for the
 7          department scholarship↑ that I'm going
 8          to apply↑ (0.5) so I wondered if you could uhm
 9          give me (0.7) a recommendation letter?=
10    P:   =>Yes of course< (.) no problem
11    J:   okay
12          (1.2)
13    P:   Uh:: was there >when is the< when- when do you
14          need a letter (.) by=
15    J:   =uhm: I need it (.) in one week↑=
```

Excerpt 3 below shows a lower-level examinee's performance on the same role-play situation with a quite distinct sequential organization compared to those shown in Excerpts 1 and 2. What is distinct in Excerpt 3 is the noticeable amount of intra-turn and inter-turn delays throughout the interaction. J first projects a pre-sequence in line 7, alerting P that a question is to follow. After the 0.5 second delay in line 8, P projects a go-ahead response in line 9. What is interesting after this sequence is J's question in lines 11 to 12, which asks P what needs to be done to apply for the scholarship. Here, J formulates a question to a professor, rather than a request (i.e., target pragmatic action). Then, P orients to this unexpected sequence by repeating the question in line 14. Responding to P's question, J then states the need of letter in a declarative sentence in lines 18 to 19. The request sequence in Excerpt 3 indicates that this examinee shows little evidence toward orienting to the imposition of the request (asking a professor for a recommendation letter with short notice).

Excerpt 3.
P: Professor, J: Jessie (Examinee ID65, Low-intermediate-level)

```
 7    J:   a:h (0.4) I have some question
 8          (0.5)
 9    P:   sure
10          (0.6)
```

11   J: what should I (0.3) d<u>o</u> I (0.9) reply (0.3) re-
12   (0.6) scholarship
13   (0.9)
14   P: a:h (0.2) what should you do?
15   (1.1)
16   J: °a:h°=
17   P: =to (.) apply for a scholarship?
18   J: yeah (.) I need (0.2) recommendation <u>le</u>tter for (0.4)
19   about (.) you

In the recommendation letter request role-play, the professor interlocutor was instructed to ask an examinee whether the recommendation letter can be submitted a bit late due to an upcoming trip, which was noted in the role-play card. When an examinee responds to such an unexpected proposition, his/her turn is often designed to avoid overt disalignment, such as rejection, which is a dispreferred response. For example, in Excerpt 4, P asks whether the letter can be submitted a bit late in lines 32 and 33. After a 0.9 second pause (line 34), J's response starts with a hesitation, evidenced by *uhm* and *I'm not sure*, as an attenuation of overt 'no' and J's response ends with an alternative solution to the professor's request. This is in accord with the preference organization sequence (Pomerantz, 1984; Pomerantz & Heritage, 2013). On the other hand, Excerpt 5 involves a rather short second pair part (*I don't know*, line 42) to the professor's request, which is prefaced with *uh* and the 1.6-second inter-turn delay to the first pair part. In addition, there is no post-expansion in rejecting the professor's proposition.

Excerpt 4.
P: Professor, J: Jessie (Examinee ID12, Mid-level)

32   P: =yeah is- B<u>UT</u> uh is it possible to (0.7) uh submit the
33   letter (.) uh a little bit l<u>a</u>te (0.4) after the deadline
34   (0.9)
35   J: uhm (1.1) I'm not sure but I- I will ask I mean (.) I will
36   check with the (1.1) the professor in my (0.5) in the
37   department that they (0.9) yeah like they manage their
38   (0.4) scholarship↑ (1.4) a:nd

Excerpt 5.
Professor, J: Jessie (Examinee ID16, Low-intermediate-level)

39   P: =but would it be okay if I turn it in a little bit
40   late (.) could it be a day or two late?
41   (0.4)
42   J: u:h (1.6) .hh I don't know
43   P: okay

Excerpts 1 to 5 represent the recommendation letter request role-play performance at various levels. It is worth noting that the contextual variables embedded in the role-play ensured elicitation of various pragmatic and interactional resources. The examinees oriented to the higher degree of imposition of recommendation letter request differently which is evident in sequential organizations, such as the use of pre-request (e.g., line 5 in Excerpt 1) and the absence of formulatic request expression (e.g., line 14 in Excerpt 3). Furthermore, the contingency in the role-play design (i.e., a professor's conference trip) ensured additional opportunity to elicit examinees' abilities to respond to a dispreferred action (lines 35–38 in Excerpt 4; line 42 in Excerpt 5).

Additional advising time request role-play
The examinees engaged in another request role-play situation with a professor. They were asked to request additional advising time during the office hour role-play interaction. Excerpt 6 shows the high-level examinee's performance. What is noticeable here is the use of pre-sequence in lines 1 and 2 (*I have another question*) that prefaces with a conditional (*if you have time*), displaying J's orientation to the imposition of requesting additional advising time to a professor. After receiving a confirmation from P in line 3, J first establishes a shared understanding (*you know we have to do assignment*) and offers the explanation of the situation before asking the professor's availability for advising. Namely, J employs various types of interactional organization that clearly demonstrate that J orients to the degree of imposition.

Excerpt 6.
P: Professor, J: Jessie (Examinee ID7, High)

```
1   J:   ahm professor Brown if you have (.) time ah (.) I
2        have another question
3   P:   u:hm
4   J:   ahm (0.8) you know we have to do this assignment↑
5        (0.4) still a:h (0.5) I haven't been able to choose
6        a topic (0.5) would you be able to advise me:
7        (0.7)
8   P:   tsh .hhh=
9   J:   =ah (0.4) to select the topic?
```

In contrast to Excerpt 6, Excerpt 7 shows a distinct turn positioning of a lower-level examinee's request sequence. J first provides an account in lines 1 and 3, and then projects a time availability question in lines 5 and 6. This request sequence is quite different from those shown in Excerpt 6. While the examinee in Excerpt 6 first projects a pre-sequence and establishes a shared understanding before the request sequence, the examinee in Excerpt 7 projects the inability of executing the given task up front.

Excerpt 7.
P: Professor, J: Jessie (Examinee ID5, Low-intermediate-level)

```
 1   J:   °hi° (0.5) a:h I cannot decide (0.6) good (0.3) topic↑
 2   P:   uh huh=
 3   J:   =so: (1.0) I'm confusing↑
 4   P:   uh huh↑ =
 5   J:   =hh (0.7) so (0.3) do you have a (0.5) time to (.)
 6        answer my questions?
```

Similar to the recommendation letter request role-play with a professor, the higher-level examinees also oriented to the imposition of request in the advising time request role-play via the use of pre-request and a turn to establish a shared understanding and an account (lines 1–2 and lines 4 to 6 in Excerpt 6).

Refusing a professor's request role-play
In the last role-play situation with a professor, the examinee responded to the professor's request of changing a presentation schedule. Here, the examinees did not know what the professor would ask, and the role-play card did not specify what students were supposed to say. What was provided to the examinees was the schedule of a midterm exam in a different class. Thus, the examinees typically refused the professor's request because of the schedule conflict. Excerpt 8 exemplifies how the examinee orients to refusing as the dispreferred action in sequential organization. In lines 10 to 16, P first establishes the shared understanding of the scheduled presentation and offers a detailed explanation behind P's request for changing the presentation schedule, orienting to the of imposition of this request. After a 0.5 second pause in line 17, J refuses the professor's request using a clear account of a pre-scheduled midterm schedule that prefaces with hesitation (*uh:*) and a discourse marker (*actually)* in lines 18 and 19, followed by an explicit apology in line 21.

Excerpt 8.
P: Professor, J: Jessie (Examinee ID22, High-level)

```
10   P:   and I kn<u>ow</u> that you're scheduled for two weeks from now
11        (1.1) however one of your classmates (.) who's scheduled
12        for next w<u>ee</u>k .hh is really sick (0.4) an[d needs to: (0.3)
13   J:                                                   [oh:
14   P:   uh (0.3) reschedule for the week <u>a</u>fter (0.9) .hh: would it
15        be possible to trade (.) presentation times with him and
16        you do your presentation next week?
17        (0.5)
18   J:   uh: actually I have a midterm exam next week (.) next
19        Friday↑ [for history class so .hh (.) maybe I have no time
20   P:           [mmm
21   J:   for extra class w<u>or</u>k [so I'm sorry about that=
22   P:                             [mmm
```

Not all examinees accomplished refusal as the dispreferred action in their sequential organization. For example, Excerpt 9 demonstrates a mid-level examinee's performance, which entails a distinct turn-taking structure in refusing the professor's request. In response to P's schedule change request in lines 9 and 10, J projects a question that demands a professor's answer before P finishes his utterance with a noticeable overlap in lines 11 and 12. Further, J rejects the professor's request in lines 16 to 17 without a pause or hesitation, displaying little evidence of sensitivity toward the dispreferred response (refusal).

Excerpt 9.
P: Professor, J: Jessie (Examinee ID3, Mid-level)

```
 9    P:   .hh (.) I was wondering if you'd be able to do your
10          presentation a week early to do it next week [and uhm
11    J:                                             [oh can I
12          ask (0.4) why?
13    P:   because the student who's supposed to present next week
14          (.) is sick and she won't be here at all (.) so would
15          you be willing to go a week [early?
16    J:                                 [that's sad but I cannot
17          because (.) I have an: .hh exam next weeks?
18    P:   mmm=
19    J:   =so: (0.3) .tsh I cannot (0.3) .hh: it- it- it (0.4)
20          next time I am too busy to (0.3) do the presentation
```

Deciding a meeting time

Moving on to the role-play interaction between two classmates (P: Phoenix, J: Jessie), Excerpts 10 and 11 illustrate the sequential organization when two examinees interact to determine an agreeable meeting time to discuss a class project. In Excerpt 10, the turn produced by each speaker is immediately followed by another turn with no inter-turn delays, indicating that the examinees orient themselves as affiliated group members. In other words, the examinees oriented to finding a mutually convenient meeting time as the preferred action. On the other hand, in Excerpt 11, P first initiates to discuss their upcoming presentation in line 1. After a 0.7 second pause in line 2, however, J does not give a go-ahead response (lines 3 to 5), which blocks the progress of subsequent sequences. The 2.5 second pause in line 6 further confirms that P orients to this absence of an expected go-ahead response.

Excerpt 10.
P: Phoenix (ID9, Low-intermediate-level), J: Jessie (ID10, High-level)

```
1    P:   Jessie↑ I think we need to talk about the time for
2          our presentation
3    J:   yeah ri[ght
```

```
4    P:           [so: when you- when (.) are you available?
5    J:    a:h [I would be available (0.6) whole Saturday
6    P:           [>°some time this week°<
7    P:    <whole Saturday?>=
```

Excerpt 11.
Phoenix (ID15, Mid-level), J: Jessie (IDI6, Mid-level)

```
1    P:    let's talk about our presentation
2          (0.7)
3    J:    O:h (.) presentation we should to make appointment and
4          (0.5) I don't have enough time (0.3) now because I'm
5          .hh I: have a class:
6          (2.5)
7    J:    so (0.3) when: (1.7) what time (.) ah
8          (0.7)
9    P:    so (1.0) when e- (1.0) a:h so you do you ti:me o:n
10         (0.8) Tuesday?
```

Deciding a meeting mode

Upon deciding on an agreeable meeting time, the examinees moved on to deciding how they will meet, a face-to-face meeting or an online discussion (see Excerpt 12). Here, J first expresses a preference for the online discussion option, which is evident in his listing online meeting options, such as Skype or Google talk, in line 9. However, P disagrees with J's suggestion and instead suggests a face-to-face discussion option in line 11. What is noticeable in this sequential organization is the absence of the expected account as part of the second pair part. What is normally expected in a disagreement sequence is an account and a mitigation. The 2.7 second pause in line 12 and a pursuit to get an answer in line 13 further confirm that J orients to the absence of a normally expected account. J explicitly pursues the expected account in line 13.

Excerpt 12.
Phoenix (ID19, Mid-level), J: Jessie (ID20, High-level)

```
4    J:    eh: (0.6) I personally prefer (0.4) tsh (0.7)
5          face-to-face discussion or online discussion
6          °maybe° (0.3) using [some kind of (0.6) Goo:gle
7    P:                         [oh:
8          (0.6)
9    J:    or Sky:pe (0.5) eh: (0.5) Google talk or something
10         like that?
11   P:    °him:° I prefer to face-to-face↑ (0.8) discussion↑
12         (2.7)
13   J:    ah: wh- why do you prefer face-to-face discussion?
```

As seen in Excerpts 10 to 12, two examinees who played the classmates oriented to the normative turn-taking structure when expressing opinions while trying to find an agreeable decision. From the role-play design perspective, since each examinee did not know what each other will say in terms of the meeting time and the preferred choice of meeting mode, the examinees exchanged turns to express and seek each other's opinions. In doing so, they oriented to the normative turn-taking structure when exchanging opinions.

In sum, the CA findings indicate that the examinees oriented to the varied degree of contextual features embedded in the role-play situations and the recipients oriented to the unexpected interactional organizations, which is captured via CA's analytical apparatus. Ways in which the examinees at various levels achieved pragmatic actions in sequential organization differed as well. Further, various types of interactional organization, including adjacency pairs, pre-sequences, and preference organization, emerged across the role-play situations, which are reflective of naturally occurring conversation. These interactional features that emerged at various levels were integrated into developing task-specific rating criteria descriptions in Youn's (2015) previous study. However, whether interaction-specific rating categories function in distinguishing varying levels of pragmatic interaction remains a question. The next section presents findings for this question.

Multi-faceted Rasch Measurement analysis

The quantitative functioning of the interaction-specific rating categories is examined as additional support for the evaluation inference. Since the analytical rating criteria that reflect various characteristics of role-play interaction were developed and their satisfactory quality was warranted in an earlier study (Youn, 2015), the quality of two types of rating criteria (one with five rating categories and the other with three interaction-specific rating categories only) was compared using MFRM (see Table 2 for the rating categories). Corresponding to each set of rating criteria, two separate MFRM analyses were conducted.

Both MFRM analyses included four facets (examinee, rater, role-play task, rating category). For each set of rating criteria, Figures 1 and 2, respectively, present a FACETS summary, also called *a variable map*, which locates the four facets on the same logit scale. The logit scale is a true interval scale, which transforms the likelihood of a particular response by an examinee at a certain level (McNamara, 1996). The logit scale is presented in the first column in Figures 1 and 2. By convention, an examinee with ability at zero logits has a 50 percent chance of succeeding on an item of average difficulty. Each column in Figures 1 and 2 presents the information for elements within each facet. The second column displays the distribution of 102 examinees' varying abilities. Each asterisk represents one examinee. Those located in the upper part

of the figure have higher abilities, followed by lower-level examinees located at the bottom. The third column presents 12 raters in terms of their degree of severity. More severe raters are located higher in the column, while less severe raters appear in the lower column. As seen in Figures 1 and 2, the rank order of rater's severity estimated by the two types of rating criteria is slightly different. The fourth column lists the five role-plays in terms of their difficulty. More difficult role-plays appear higher in the column and the less difficult role-plays appear lower. The difficulty rank order of the five role-plays was almost identical in the two figures, except for the *discussing a meeting time* role-play with a classmate, that was estimated as more difficult with the full rating categories (Figure 1). The fifth column lists each rating category's difficulty level. Figure 2 only lists three interaction-specific rating categories. The rank order of these three rating categories' difficulty levels was identical in both Figures 1 and Figures 2. The remaining columns demonstrate how scores for the five role-plays were utilized.

```
|Measr|+Examinee  |−Rater    |−Task                                   |−Category   |  S.1 |  S.2 |  S.3 |  S.4 |  S.5 |
+  6 +            +          +                                        +           + (3) + (3) + (3) + (3) + (3) +
|    |            |          |                                        |           |     |     |     |     |     |
|    |            |          |                                        |           |     |     |     |     |     |
|    |            |          |                                        |           |     |     |     |     |     |
+  5 + *          +          +                                        +           +     +     +     +     +     +
|    |            |          |                                        |           |     |     |     |     |     |
|    |            |          |                                        |           |     |     |     |     |     |
+  4 +            +          +                                        +           +     +     +     +     +     +
|    |            |          |                                        |           |     |     |     |     |     |
|    | *          |          |                                        |           |     |     |     |     |     |
+  3 | ***        +          +                                        +           +     +     +     +     +     +
|    | ***        |          |                                        |           |     |     |     |     |     |
|    | ****       |          |                                        |           |     |     |     |     |     |
|    | **         |          |                                        |           |     |     |     |     |     |
|    | *****      |          |                                        |           |     |     |     |     |     |
|    | ***        |          |                                        |           |     |     |     |     |     |
|    | ***        |          |                                        |           |     | --- | --- | --- | --- |
+  2 + ******     +          +                                        +           +     +     +     +     +     +
|    | *****      |          |                                        |           |     |     |     |     |     |
|    | **         |          |                                        |           |     |     |     |     |     |
|    | ****       |          |                                        |           |     |     |     |     |     |
|    | ********   |          |                                        |           |     |     |     |     |     |
|    | ********** | R1       |                                        |           | --- |     |     |     |     |
+  1 + ***        +          +                                        +           +     +     +     +     +     +
|    | ***        |          |                                        | Language  |     |     |     |     |     |
|    | *          | R12      |                                        |           |     |     |     |     |     |
|    | ***        | R2       | Request Ltr                            | Content   |     |     |     |     |     |
|    | *****      | R3N      |                                        |           |     |     |     |     |     |
|    | **         | R10  R6N  R8 | Time                               |           | 2   | 2   | 2   | 2   | 2   |
*  0 * ***        * R9N      * Mode      Refuse    Request time * Situation *       *     *     *     *     *     *
|    | *****      | R4       |                                        | Turn      |     |     |     |     |     |
|    | ***        |          |                                        |           |     |     |     |     |     |
|    | ****       |          |                                        | Engaging  |     |     |     |     |     |
|    | *****      |          |                                        |           |     |     |     |     |     |
+ −1 +            +          +                                        +           +     +     +     +     +     +
|    | ***        |          |                                        |           |     | --- | --- | --- | --- |
|    |            |          |                                        |           | --- |     |     |     |     |
|    |            |          |                                        |           |     |     |     |     |     |
+ −2 +            +          +                                        +           + (1) + (1) + (1) + (1) + (1) +
|Measr| * = 1     |−Rater    |−Task                                   |−Category   |  S.1 |  S.2 |  S.3 |  S.4 |  S.5 |
```
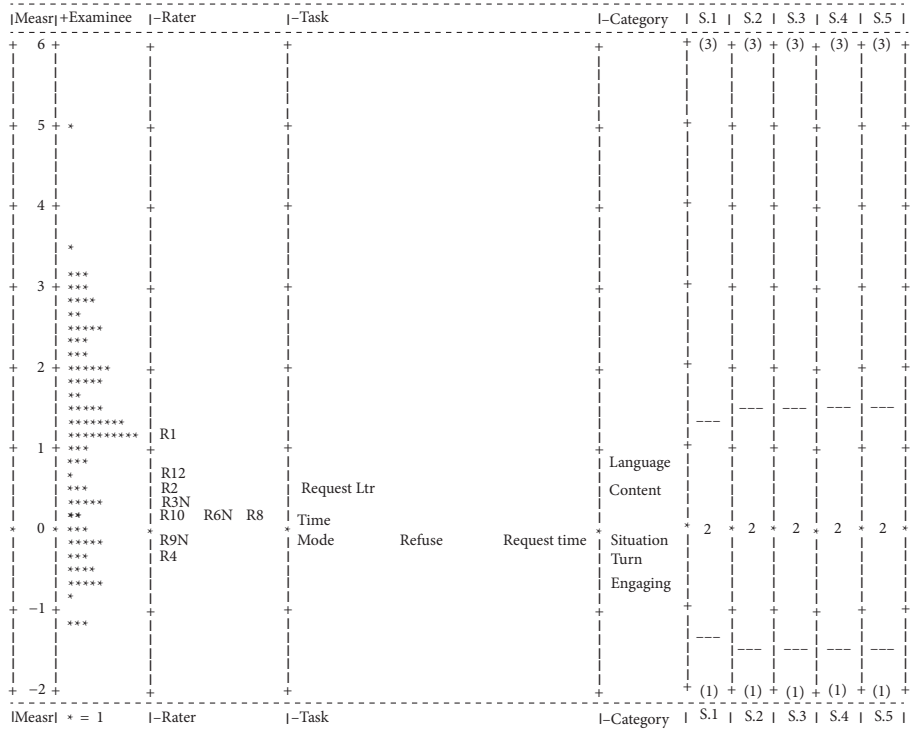
Figure 1. FACETS Summary of All Five Rating Categories

The second column of Figures 1 and Figures 2 displays the distribution of 102 examinees' pragmatic abilities estimated via the five role-plays and the different sets of rating criteria. In both Figures 1 and Figures 2, more examinees appear above the zero logit,

indicating that they were above intermediate ability levels. The fact that the examinees in this study had at least low-intermediate language ability explains such an ability range.
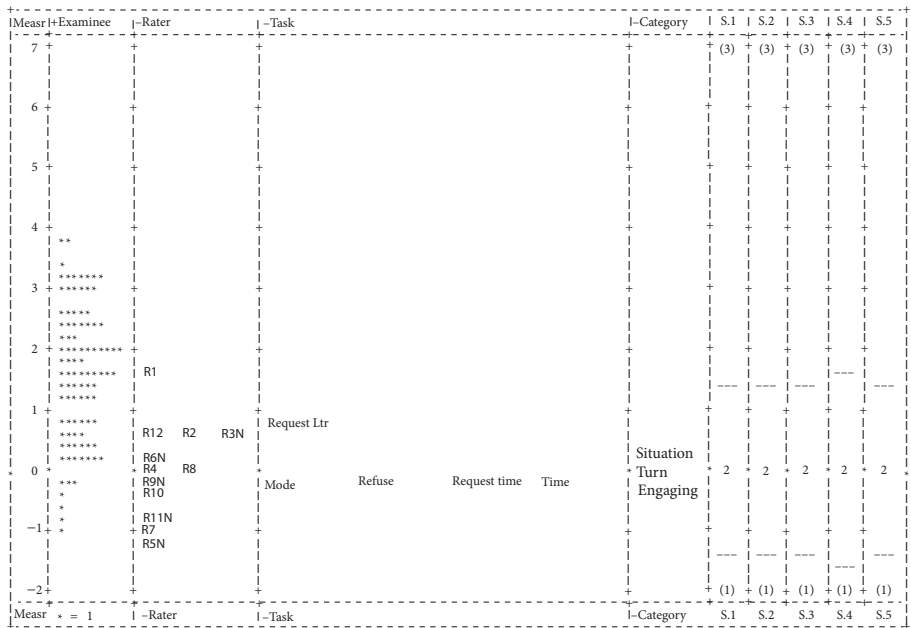

Figure 2. FACETS Summary of Three Interaction-specific Rating Categories (Sensitivity to Situation, Engaging with Interaction, Turn Organization)

MFRM provides various statistics for each facet, such as strata index, reliability of separation, chi-square, and infit statistics (see Bond & Fox, 2007; Eckes, 2011). Table 3 compares these indices to examine how two types of rating criteria functioned in separating the examinees' pragmatic abilities. Firstly, in both sets of rating criteria, the examinees were widely spread out with a similar logit spread (6.32 and 7.05 logits spread respectively). Furthermore, the strata, which refer to the number of distinct levels within a given facet, for both sets of rating criteria (8.35 and 5.57 respectively) were large enough for distinguishing the 102 examinees' pragmatic abilities. In other words, there are about eight levels of examinee pragmatic ability estimated with the full rating criteria, while there are about six levels of examinee pragmatic ability estimated with the interaction-specific rating categories. A separation value indicates how many measurement strata could be statistically distinguishable among the measures. The reliability of separation indicates the degree to which the measures distinguish between different strata for the examinee facet. The separation value (6.01) and strata (8.35), with its reliability index of .97, indicate that all five rating categories better

separated examinees' varying abilities, compared to those of the three interaction-specific rating categories. Finally, the correlation between the examinees' ability logits resulting from both sets of rating criteria was .97, indicating the rank order of each examinee's ability logit was highly compatible.

Apart from the strata and reliability indices, MFRM also estimates infit statistics that indicate whether the examinee ability was appropriately measured or not. The mean fit values for both sets of rating criteria were close to the expected value of 1.0. Generally accepted infit values range from 0.5 to 1.5 (Lunz, Wright, & Linacre, 1990). An infit value over the upper-control limit refers to *misfit*, which indicates too much unpredictability in estimating the examinees' ability logits. An infit value below the lower-control limit is called *overfit*, which displays a lack of variation (Eckes, 2011). Both misfit and overfit indicate that the extent to which examinee ability is measured appropriately. Applying this criterion, all examinees estimated using both sets of rating criteria were within the acceptable range, indicating that no examinee's performance was erratically unpredictable.

Taken together, the high strata indices for both sets of rating criteria indicate that the variance among the examinees was substantially larger than the error of estimates, which means that the rating criteria separated the 102 examinees into at least six statistically distinct levels of pragmatic competence in interaction. The high reliability statistic confirms that the findings are replicable as well. In short, the three interaction-focused rating categories on the five role-plays (see Table 1) discriminated among the examinees in terms of pragmatic competence in interaction, although not as much as when all rating criteria were included.

Table 3. Comparison of two sets of rating criteria on the examinee facet

Rating criteria	Max ~ Min (Logit range)	Strata	Separation	Reliability	Chi-square statistic	Mean infit
All rating categories	5.08 ~ −1.24 (6.32 logits)	8.35	6.01	0.97	4892.6 (*d.f.* = 101; *p* < .00)	.99 (*SD* = .16)
Interaction-focused three rating categories	6.12 ~ −0.93 (7.05 logits)	5.57	3.91	0.94	2101.4 (*d.f.* = 101; *p* < .00)	1.01 (*SD* = .18)

On the other hand, the raters applied the rating criteria slightly differently to each role-play. The last five columns in Figures 1 and Figures 2 indicate how the three-point scale of the rating criteria was used for the five role-play tasks. For example, the lowest score 1 was more frequently used for the recommendation letter request to a professor task (S.1.) as seen in the sixth column in Figure 1. Of the five role-plays, the recommendation letter request to a professor task was most challenging, which might be due to the task feature of a relatively high degree of imposition (i.e., requesting a letter with

a short due date) in a more formal situation. In terms of the second set of rating criteria that only included the interactional features, the lowest score of 1 was least used for the meeting time role-play with a classmate, which means the raters used the scores of 2 and 3 more often. It can be inferred that the raters might have been more lenient when awarding scores on the interactional features elicited from the classmate role-play, which likely involves more informal interaction between classmates compared to student-professor role-play interaction.

Discussion

Employing a concurrent mixed methods design, CA and MFRM analyses were used independently to investigate the role of interactional organizations as validity evidence for assessing L2 pragmatic interaction. The purpose of CA was to examine whether the role-plays could elicit valid interactional behaviors that may be task-independent at varied levels of achievement. The quantitative functioning of the interactional features embedded in rating criteria descriptions in differentiating the examinees' abilities is also critical. Thus, in tandem with the CA findings, the MFRM analyses were conducted to compare the quantitative functioning of the interaction-specific rating categories to those of the full rating categories.

Each role-play situation was different in terms of various contextual variables (e.g., degree of imposition, interlocutor relationship). However, as shown in the CA findings, similar types of interactional organization that the examinees used in accomplishing the given communicative goals emerged recurrently across all role-play situations. First of all, the adjacency pair, one of the most obvious types of interactional organization, was observed across all role-plays. Heritage (1984) describes adjacency pairs as "the basic building-blocks of intersubjectivity" (p. 256). During the role-play interaction, the examinees followed the norms of achieving the adjacency pair. When this norm was violated by one of the speakers, such as the next speaker not producing the relevant next action, the absence of the expected action was treated as noticeable and accountable. For example, when the expected second pair part of granting a request (i.e., thanking) was absent (see Excerpt 2), the recipient oriented to such absence, as evidenced by a noticeable inter-turn delay. This finding, in turn, emphasizes that the degree to which participants produce the adjacency pair, and whether they orient to the abnormal adjacency pair construction need to be investigated as part of the validity evidence for assessing L2 pragmatics in interaction.

Another common feature of sequential organization that emerged across the role-plays was the pre-sequence that establishes the context of an upcoming action. When the higher-level examinees launched a course of action, the pre-sequence (e.g., pre-request) was first projected before projecting the main action sequence (e.g., request).

On the other hand, the lower-level examinees often jumped to the main action immediately without a pre-sequence, as shown in Excerpt 7. Further, the participants projected actions at a different sequential order, displaying a varying degree of sensitivity toward the role-play situations. For example, when making a recommendation letter request, the higher-level examinees first provided an account and background information before an actual request sequence (Excerpt 1). However, the lower-level examinees' turn positioning of the request sequence first started with a question to a professor, demanding an answer to the question, rather than providing the necessary information about the request to the professor up front (Excerpt 3). In another role-play situation with a professor, in which an examinee requests additional advising time, the high-level examinee established a shared understanding first before making a request, whereas the low-level examinee first stated the inability of accomplishing an assignment. In short, the extent to which turns were positioned in relation to the course of actions differed across the examinees' levels.

The organization of preference and dispreference was also prevalent across all role-plays. Various contextual features were embedded in the role-play design, including the varying degree of imposition, the unexpected information from a professor, and different speech acts (e.g., request, refusal, disagreement, agreement). The examinees themselves oriented to those contextual variables, which was evident in sequential organization. For example, some actions in sequential organization are preferred (e.g., agreements, granting a request, accepting an invitation), while other actions are dispreferred (e.g., rejections, declining an offer, and disagreements). As shown in the excerpts presented in the results section, the evidence for preference organization was present across all role-plays. However, not all examinees oriented to the normative nature of dispreferred action in sequential organization. This was seen in the lack of mitigation when refusing the professor's request to change the schedule in Excerpt 9. The fact that different types of interactional organization were present in role-play interaction further confirms the potential of role-plays in eliciting various features reflective of naturally occurring interaction. These findings directly address the concerns attached to the elicited role-play interaction, thereby satisfying the validity assumption that the role-plays indeed elicit authentic interaction of task-based pragmatic performance.

In addition to the CA findings, the quantitative functioning of interactional features embedded in the rating criteria was examined as a follow-up analysis of the rating criteria developed in Youn (2015). The role of rating criteria in ensuring the evaluation inference is critical since the descriptions of rating criteria can translate the theoretical construct into more concrete examples that enable raters to assign scores consistently. This study found that the three interaction-specific rating categories functioned similarly in distinguishing the examinees' abilities in comparison to the findings estimated via the full five rating categories. The MFRM analysis indicated that approximately six distinct levels of examinee ability were estimated via the role-plays using the three

interaction-specific rating categories. At the same time, the raters applied the rating criteria differently for the role-play tasks that differ in terms of formality determined by the contextual features in each role-play. Nonetheless, the compatibility of the two types of rating criteria does not indicate that we should only consider the interaction-specific rating categories because of the construct-representation issue. Other types of pragmatic knowledge, such as grammatical resources, should be reflected in rating criteria descriptions for task-based pragmatic assessment to represent the entire construct adequately. The purpose of comparing two sets of rating criteria was to examine how the interactional features alone embedded in the rating criteria function quantitatively and how they are used differently by the raters across the role-play tasks.

Taken together, the findings of the current study offer the following implications for task-based pragmatic assessment using role-plays. First, various contextual variables need to be considered in designing role-plays for eliciting task-based pragmatic performance. While the role-play itself is widely employed, ensuring a clear communicative goal and authentic interaction is not necessarily guaranteed. In this study, as shown in the CA findings, the examinees oriented to such contextual variables, which are evident in ways in which turns are organized (e.g., pre-request, meaningful pauses between turns) that are similar to those of ordinary conversation. In part, this was possible as the role-plays in this study were designed to simulate real-life interaction where speakers do not know what another speaker will say. The examinees in this study had to respond to various contingencies embedded in the role-plays, such as responding to the professor's unexpected proposal of submitting the recommendation letter late. Further, ways in which the examinees accomplished the pragmatic actions in interaction were different depending on their levels. Such types of sequential organization functioned as ratable data for assessment purposes. In other words, eliciting validity evidence of various interactional behaviors from role-play-based pragmatic performance is certainly possible. Apart from the situations used in the role-plays in this study, various situational variables could be considered depending on the intended uses of task-based pragmatic assessment. Secondly, descriptions related to various types of interactional organization need to be explicitly included when developing rating criteria as long as task-based pragmatic performance involves talk-in-interaction. I argue that embedding the interactional organizations in rating criteria descriptions will not only represent a critical aspect of the targeted construct but also generate enough variance in distinguishing among varying levels of language performance.

Conclusion

In order to address the validity challenges in assessing pragmatic competence in interaction, this study focused on the interactional phenomena elicited from role-play

interaction and the functionality of interaction-specific rating categories for assessment purposes. The CA findings indicated that the examinees oriented to various contextual variables embedded in the role-plays. In addition, the extent to which the examinees achieved their courses of action in sequential organization was distinct across levels. Furthermore, several interactional organizations (e.g., pre-sequence, adjacency pair, preference organization) recurrently emerged in the role-play interaction. These findings support the assumption that carefully designed role-plays that reflect contextual features of authentic tasks can indeed elicit valid interactional features of pragmatic performance. In terms of the assumption related to rating criteria, the interaction-specific rating categories also contributed to differentiating the examinees' abilities in a satisfactory manner. The findings confirm the importance of considering interactional organizations as validity evidence when assessing task-based pragmatic performance in the context of spoken interaction. Nonetheless, the limitations of the study should be noted. The backing for the assumptions underlying the evaluation inference was limited to the sequential organization of select role-play performances and the MRFM analyses. Thus, additional backing to strengthen the degree of generalizability is necessary. Future research should include analysis of the extent to which the interactional features are generalizable to pragmatic performances on different tasks and how they can predict examinees' abilities.

Ensuring valid assessment practices fundamentally stems from the comprehensive understanding of the targeted construct specific to assessment task types and intended interpretations and uses of a test. This study focused on the types of sequential organization that emerged from role-play-based pragmatic performance. Sequences can function as vehicles for getting certain actions accomplished in spoken interaction. Clearly, ways in which the actions were accomplished were distinct in sequential organizations depending on the examinees' levels. Given that the interactional phenomena examined in this study were also reflective of the normative nature of talk-in-interaction, I can conclude that the role-play interaction, even though it is elicited, can indeed be socially consequential (Huth, 2010). This conclusion, in turn, indicates that the role-play interaction, as long as it is carefully designed, can ensure the validity evidence necessary for inferring examinees' task-based pragmatic performance. The potential of role-plays as valid assessment task types for task-based pragmatic assessment appears promising.

References

Al-Gahtani, S., & Roever, C. (2012). Proficiency and sequential organization of L2 requests. *Applied Linguistics*, 33, 42–65. doi:10.1093/applin/amr031

Al-Gahtani, S., & Roever, C. (2014). Insert and post-expansion in L2 Arabic requests. *System*, 42, 189–206. doi:10.1016/j.system.2013.10.018

Atkinson, J., & Heritage, J. (Eds.) (1984). *Structures of social action*. Cambridge: Cambridge University Press.

Bardovi-Harlig, K., & Hartford, B. S. (Eds.) (2005). *Interlanguage pragmatics: Exploring institutional talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model* (2nd ed.). New York, NY: Routledge.

Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In E. Alcón Soler & M. P. Safont Jordà (Eds.), *Intercultural language use and language learning* (pp. 41–57). Dordrecht: Springer. doi:10.1007/978-1-4020-5639-0_3

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6, 5–35.

Chapelle, C. A. (2012). Validity argument for language assessment: The framework is simple… *Language Testing*, 29, 19–27. doi:10.1177/0265532211417211

Chapelle, C., Enright, M., & Jamieson, J. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York, NY: Routledge.

Eckes, T. (2011). *Introduction to many-Facet Rasch measurement*. Frankfurt: Peter Lang.

Grabowski, K. (2013). Investigating the construct validity of a role-play test designed to measure grammatical and pragmatic knowledge at multiple proficiency levels. In S. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 149–171). Houndmills: Palgrave Macmillan. doi:10.1057/9781137003522.0013

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In M. J. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 299–345). Cambridge: Cambridge University Press.

Hudson, T. D., Detmer, E., & Brown, J. D. (1992). *A framework for testing cross-cultural pragmatics* (Technical Report #2). Honolulu, HI: University of Hawaiʻi, Second Language Teaching and Curriculum Center.

Hudson, T. D., Detmer, E., & Brown, J. D. (1995). *Developing prototype measures of cross-cultural pragmatics* (Technical Report #7). Honolulu, HI: University of Hawaiʻi, Second Language Teaching and Curriculum Center.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Huth, T. (2010). Can talk be inconsequential? Social and interactional aspects of elicited second-language interaction. *Modern Language Journal*, 94, 537–553. doi:10.1111/j.1540-4781.2010.01092.x

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527–535. doi:10.1037/0033-2909.112.3.527

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (pp. 17–64). Westport, CT: American Council on Education.

Kasper, G. (2006). Speech acts in interaction: Towards discursive pragmatics. In K. Bardovi-Harlig, C. Félix-Brasdefer & A. S. Omar (Eds.), *Pragmatics and language learning* (pp. 281–314). Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawaiʻi.

Kasper, G., & Rose, K. R. (2002). *Pragmatic development in a second language*. Malden, MA: Blackwell Publishing.

Kasper, G. (2009). L2 pragmatic development. In W. C. Ritchie & T. K. Bhatia (Eds.), *New handbook of second language acquisition* (pp. 259–295). Leeds: Emerald.

Kasper, G., & Ross, S. (2013). Assessing second language pragmatics: An overview and introductions. In S. J. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 1–40). Houndmills: Palgrave Macmillan. doi:10.1057/9781137003522.0007

Kasper, G., & Youn, S. J. (2017). Transforming instruction to activity: Roleplay in language assessment. *Applied Linguistics Review*. Advance online publication. doi:10.1515/applirev-2017-0020

Linacre, J. M. (2014). Facets Rasch measurement computer program, version 3.71.4. Chicago, IL: Winsteps.com.

Liu, J. (2007). Developing a pragmatic test for Chinese EFL learners. *Language Testing*, 24, 391–415. doi:10.1177/0265532207077206

Lunz, M. E., Wright, B. D., & Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3, 331–345. doi:10.1207/s15324818ame0304_3

McNamara, T. (1996). *Measuring second language performance*. New York, NY: Addison Wesley Longman.

McNamara, T., & Knoch, U. (2012). The Rasch wars: The emergence of Rasch measurement in language testing. *Language Testing*, 29, 555–576. doi:10.1177/0265532211430367

Norris, J. M. (2009). Task-based teaching and testing. In M. H. Long & C. J. Doughty (Eds.), *Handbook of language teaching* (pp. 578–594). Oxford: Blackwell. doi:10.1002/9781444315783.ch30

Norris, J. M. (2016). Current uses for task-based language assessment. *Annual Review of Applied Linguistics*, 36, 230–244.

Okada, Y. (2010). Role-play in oral proficiency interviews: Interactive footing and interactional competencies. *Journal of Pragmatics*, 42, 1647–1668. doi:10.1016/j.pragma.2009.11.002

Okada, Y., & Greer, T. (2013). Pursuing a relevant response in oral proficiency interview role plays. In S. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 288–310). Houndmills: Palgrave Macmillan. doi:10.1057/9781137003522.0019

Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 57–101). Cambridge: Cambridge University Press.

Pomerantz, A., & Heritage, J. (2013). Preference. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 210–228). Chichester: Wiley-Blackwell.

Purpura, J. E., Brown, J. D., & Schoonen, R. (2015). Improving the validity of quantitative measures in applied linguistics research. *Language Learning*, 65, 37–75. doi:10.1111/lang.12112

Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing*, 23, 229–256. doi:10.1191/0265532206lt329oa

Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28, 463–481. doi:10.1177/0265532210394633

Ross, S. J., & Kasper, G. (Eds.) (2013). *Assessing second language pragmatics*. Houndmills: Palgrave Macmillan. doi:10.1057/9781137003522

Sacks, H. (1992). *Lectures on conversation* (2 Vols.). Oxford: Blackwell.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511791208

Schumacker, R. E. (1999). Many-facet Rasch analysis with crossed, nested, and mixed designs. *Journal of Outcome Measurement*, 3, 323–338.

Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford: Oxford University Press.

Teddlie, C., & Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *Research in the Schools*, 13(1), 12–28.

Walters, F. S. (2007). A conversation-analytic hermeneutic rating protocol to assess L2 oral pragmatic competence. *Language Testing*, 24, 155–183. doi:10.1177/0265532207076362

Youn, S. J. (2010). *Task-based needs analysis of L2 pragmatics in an EAP context: Implications for teaching*. Unpublished manuscript, University of Hawai'i.

Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, 32, 199–225. doi: 10.1177/0265532214557113

Appendix A. Role-play with a professor and role-play cards

Role-play 1 (with a professor)

Situation. You have an appointment with a professor Morgan Brown today to ask for a recommendation letter for a scholarship for international students from your department and to ask a few questions about a course project. Your professor is meeting with you outside of the office hour since you have a class during the office hour. Now you're about to visit your professor. You just enter to a professor's room.

Task. You will receive role-play cards that describe what you're going to tell your professor. Please have a conversation with your professor naturally.

For undergraduate participants: This professor teaches Economy 101 that you're taking this semester.

For graduate participants: This professor is one of the faculty members in your department. Although he/she is not your advisor, you've known this professor for about 1 year and you're currently taking a course from this professor.

Role-play card for requesting a recommendation letter

Jessie	Professor
1. After greeting, **ask for a recommendation letter** for the department scholarship that you will apply. The letter is due in one week.	

Jessie	Professor
	Respond to the request. Inform the student that you will write a letter and ask when the due date is, if the student doesn't tell you. Inform students that you have a conference next week and you're leaving tonight. Tell him/her that you will do your best to submit the letter by the deadline, but ask the student if the letter can be submitted a bit late.

Jessie	Professor
2. Respond to what the professor says and **tell the professor that you will check with your department office**.	

Jessie	Professor
3. Inform the professor of two options of providing a letter, **hard copy or electronic submissions** through a website. **Ask for the professor's preference**.	Prefer an electronic submission as you will be traveling.

Appendix B. Rating criteria for Role-play with a Professor

Score	Contents delivery	Language use	Sensitivity to situation	Engaging with interaction	Turn organization
3	• Clear, concise, fluent (esp. speech act delivery) • Smooth topic initiations with appropriate transitional markers and clear intonations (i.e., smooth turn initiation) <u>#1</u>: a letter request & letter submission option <u>#2</u>: need for a meeting & decide a time <u>#3</u>: respond to professor's request & explain a situation	• Pragmatically appropriate linguistic expressions (**bi-clausal, conditional: I was wondering if, I don't think I can; modal verbs,** would, could, might) • Good control of grammar and vocabulary that doesn't obscure meaning	• Consistent evidence of awareness and sensitivity to situations exists in contents or tone <u>#1</u>: request along with explanations about the scholarship; acknowledge a short letter due <u>#2</u>: explanations for a meeting request <u>#3</u>: handle a face-threatening refusal with acceptable reasons or accept a request	• A next turn shows understandings of a previous turn throughout the Interaction (i.e., shared understanding) • Evidence of engaging with conversation exists (e.g., clarification questions, backchannel, acknowledgement tokens) <u>*Note*</u>: Non-verbal cues also serve as acknowledgement, so no need to heavily rely on the amount of discourse markers.	• Complete adjacency pairs (e.g., question & answer, granting a request & thank) • Interactionally fluid **without awkward pauses** or **abrupt overlap** <u>*Note*</u>: Interactionally meaningful pauses include those before refusal and between disagreements
2	• Generally smooth, but **occasionally unclear**, or unnecessarily wordy • **Abrupt** topic initiation (in terms of contents) • Unclear transitional cues (e.g., unclear intonation and stress)	• Able to use modal verbs in **mono-clausal** (e.g., could, can, might), but doesn't or inconsistently use complex structures for pragmatic meaning • Linguistic expressions are occasionally inaccurate and a bit limited that sometimes obscure meaning	• Inconsistent evidence of awareness and sensitivity to situations (e.g., explain the letter request, but not acknowledge a short letter due)	• Some evidence of engaging with the conversation, but not consistent • A next turn doesn't sometimes show understandings of previous turns	• Some turns are **delayed** and a next turn is **absent** in adjacency pairs (e.g., absence of answers & thank) • Sometimes **abruptly cut off** a previous turn
1	• Delivery is choppy, fragmented, and minimal (due to lack of language competence)	• Expressions sound **abrupt, direct,** or **not polite enough** (e.g., I need, I want, I can't) • Linguistic expressions are inaccurate and quite limited that obscure meaning	• Little evidence of situational sensitivity (e.g., not acknowledge a short letter due, insist turning in the letter on time, lack of explanations for refusal)	• Noticeable absence of discourse markers • Evidence of not achieving a shared understanding	• Noticeably abrupt overlap or no pauses between disagreements and refusal • **Noticeably long pauses** or **noticeable cutoff** between turns

Appendix C. CA transcription notation (Atkinson & Heritage, 1984)

:	Lenghtening of the preceding sound
–	Abrupt cutoff
(.)	Very short untimed pause
> <	Talk surrounded by this bracket is produced more quickly than neighboring talk
[	Point of overlap onset
=	No gap between adjacent utterances
word	Speaker emphasis
CAPITALS	Especially loud sounds relative to surrounding talk
° °	Utterances between degree signs are noticeably quieter than surrounding talk
(3.5)	Intervals between utterances (in seconds)

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

Monika Ekiert, Sofia Lampropoulou, Andrea Révész &
Eivind Torgersen
City University of New York / University of Liverpool / University College London /
Norwegian University of Science and Technology

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 set-
tings. 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 tra-
ditional 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 certain 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; Iwasaki, 2010; Schauer, 2004). 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 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 incomplete delivery. The participant calls the caterer to complain about it.	Customer – Catering Service Input: Aural 20 seconds planning time 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.	Student – 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.	Friend – 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

---

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).

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).

Results

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.

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.

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

Table 2. Descriptive statistics for discourse ratings

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

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

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, 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 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 a 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

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.
doi: 10.1017/S0272263113000429

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.

Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning*, 63, 68–86.
doi: 10.1111/j.1467-9922.2012.00738.x

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, NY: 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. doi:10.1111/j.1473-4192.2009.00231.x

Gurzynski-Weiss, L., & Révész, A. (2012). Tasks and teacher feedback in the Spanish foreign language classroom. *Language Learning*, 62, 851–879. doi:10.1111/j.1467-9922.2012.00716.x

Haverkate, H. (1984). *Speech acts, speakers and hearers* (Pragmatics and Beyond 4). Amsterdam: John Benjamins. doi:10.1075/pb.v.4

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. doi:10.1093/applin/amn047

Johnson, K., & Johnson, H. (1999). *Encyclopedic dictionary of applied linguistics*. Oxford: Blackwell. doi:10.1111/b.9780631214823.1999.x

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. doi:10.1017/S0272263114000618

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. doi:10.1111/modl.12273

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. doi:10.1177/0267658316684904

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. doi:10.1093/applin/24.4.492

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30, 590–601. doi:10.1093/applin/amp045

Plonsky, L., & Oswald, F. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878–912. doi:10.1111/lang.12079

Purpura, J. (2004). *Assessing grammar*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511733086

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, 31, 437–470. doi:10.1017/S0272263109090366

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. doi:10.1111/lang.12061

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. doi:10.1002/9781444315783.ch29

Ross, S., & Kasper, G. (2013). *Assessing second language pragmatics*. London: Palgrave Macmillan. doi:10.1057/9781137003522

Schauer, G. (2004). 'May you speaker louder maybe?' Interlanguage pragmatic development in requests. *EUROSLA Yearbook*, 4, 253–272. doi:10.1075/eurosla.4.12sch

Searle, J. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511609213

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. doi:10.1177/136216889700100302

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

Taguchi, N. (2007). Task difficulty in oral speech act production. *Applied Linguistics*, 28, 113–135. doi:10.1093/applin/amlo51

Taguchi, N. (2011a). Teaching pragmatics: Trends and issues. *Annual Review of Applied Linguistics*, 31, 289–310. doi:10.1017/S0267190511000018

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. doi:10.1515/iral.2011.015

Taguchi, N. (2011c). The effect of L2 proficiency and study-abroad experience in pragmatic comprehension. *Language Learning*, 61, 904–939. doi:10.1111/j.1467-9922.2011.00633.x

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. doi:10.1017/S0261444814000263

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. doi:10.1016/j.pragma.2012.05.007

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4, 91–112. doi:10.1093/applin/4.2.91

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. doi:10.1515/9783110885286

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.

[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**.

Assessing functional adequacy of L2 performance in a task-based approach

Folkert Kuiken & Ineke Vedder
University of Amsterdam

When assessing L2 performance in task-based research, dimensions of complexity, accuracy, and fluency (CAF) have typically been evaluated. Less attention has, however, been devoted to the functional dimension. This paper argues that it is critical to consider the functional dimension of L2 performance in addition to complexity, accuracy, and fluency. Functional adequacy as a task-related construct is viewed in the present study as a component of L2 pragmatics, referring to the appropriateness and felicity of the utterances of the speaker/writer within a particular context, and evaluated by the listener/reader. The study investigates the applicability of a rating scale developed for the assessment of functional adequacy in the L2, considered from the perspective of task-based language teaching (TBLT) and task-based language assessment (TBLA), as successful task completion. In the rating scale, four components of functional adequacy are distinguished: content, task requirements, comprehensibility, and coherence and cohesion. A group of non-expert raters judged the oral and written samples of two groups of university students of Dutch L2 and Italian L2. The results show that the scale appears to be a reliable and efficient tool for assessing the functional adequacy of written and spoken L2 production.

Introduction

When assessing second language (L2) linguistic performance, dimensions of complexity, accuracy, and fluency (CAF) have typically been evaluated (for an overview see Housen, Kuiken, & Vedder, 2012). Less attention, however, has been devoted to the functional component of L2 performance and to learners' pragmatic abilities. This paper argues that, in addition to complexity, accuracy, and fluency, it is critical to consider the functional adequacy of oral and written L2 performance and the pragmatic strategies required to carry out a certain language task.

The present study was conducted within the framework of task-based language teaching (TBLT) and task-based language assessment (TBLA). Like pragmatics, TBLT is primarily concerned with language use in social context for achieving particular

communicative goals. As Long (2015) and other scholars have argued, pedagogic tasks are beneficial for language learning, as they address learners' real-world communicative needs and entail learners' engagement with language use in socially situated interactions. Hence, the assessment of functional adequacy, viewed from a task-based perspective as the ability of successful and effective task performance, is also a concern of learning and teaching of L2 pragmatics.

The goal of this study is to examine the applicability of a rating scale developed for the assessment of functional adequacy in spoken and written L2 performance. In the rating scale, four distinct components of functional adequacy were distinguished: content, task requirements, comprehensibility, and coherence and cohesion. The study investigates how functional adequacy in an argumentative monologue task can be measured in both L2 speaking and writing. A group of non-expert raters (native speakers of the target language without teaching experience) used the rating scale for functional adequacy to evaluate the oral and written samples of two groups of L2 university students of Dutch and Italian.

In the next section we begin by discussing the challenges of testing L2 pragmatics within the TBLT and TBLA framework, and we delineate the construct of functional adequacy as it has been used in this study. Firstly, we describe the theoretical underpinnings and levels of the rating scale. Secondly the research questions, methodology, tasks, participants, and data analysis are described. We then report the results of the study. In the final section the applicability of the rating scale for the assessment of functional adequacy in oral compared to written performance is discussed, as well as the implications of the research for testing L2 pragmatic competence from a task-based perspective.

Assessment of L2 pragmatics

Pragmatic competence in an L2 has generally been assessed within the theoretical framework of speech act theory (Austin, 1962; Searle, 1969, 1975), maxims of conversation (Grice, 1975), and politeness theory (Brown & Levinson, 1987). Many of the existing studies on L2 pragmatics have investigated the acquisition of speech acts in L2 (e.g., requests, complaints, refusals, and compliments) or the use of modifiers and hedges, in different settings that add to politeness levels of the speech acts (Barron, 2003; Martinez-Flor & Usó-Juan, 2010). Data have been collected primarily by means of Discourse Completion Tests (DCT), role plays, appropriateness judgment tasks, and self-assessments. Other studies have addressed issues like the comprehension in an L2 of conversational implicature (Bouton, 1994, 1999; Roever, 2005; Taguchi 2009, 2011) and the use of formulaic routines and conventional expressions (Bardovi-Harlig, 2009; Roever, 2012; Taguchi 2013).

Application of speech act theory, maxims of conversation, and politeness theory as guiding theoretical frameworks, as well as the use of DCT and other questionnaire-based elicitation tasks, have recently come under criticism because these theories and measures underrepresent the construct of pragmatic competence. Although these measures can assess learner's knowledge of pragmatics (Roever, 2011; Walters, 2013), they underrepresent other aspects of pragmatic competence, such as pragmatic performance in extended discourse, turn-taking, sequence organization, conversational repair, or use of contextualization cues. Though alternative approaches to analyzing pragmatics at discourse-level have been suggested (Ross & Kasper, 2013), studies focusing on learners' interactional abilities or on the perlocutionary effects of the speaker's utterances on other participants of the interaction have been scarce (Taguchi & Kim, this volume).

Pragmatic competence has been conceptualized as pragmatic *knowledge* (i.e., the ability to identify and produce speech intentions appropriately in context) and pragmatic *processing* (i.e., the ability to use pragmatic knowledge efficiently in real time; Taguchi, 2012). However, as pointed out by Ross and Kasper (2013), this conceptualization, in cognitive terms, of pragmatic competence (as individual language users' cognitive processing and representation), offers no theoretical perspective on social context. Since assessment of the appropriateness of oral and written language largely depends on the way in which context has been conceptualized and analyzed in the test task, undertheorizing social context may thus be a serious concern for the testing of L2 pragmatics (McNamara & Roever, 2007). TBLT and TBLA can offer a solution to this problem because they focus both on the relationship between the interlocutors and the specific language task to be carried out within a particular context.

Within the TBLT framework, task-based language assessment has been increasingly discussed. Assessment tasks employed in task-based research, as observed by Taguchi and Kim (this volume), are usually designed based on researchers' intuitions, without considerations of the degree of correspondence between task situations and real-life situations. Rating criteria have likewise been criticized for being too generic and not sufficiently related to the target task. As pointed out by González-Lloret (2016), they should, however, always be derived from the assessment task and the type of language required in order to fulfill the task. Task-based performance assessment has also been challenged with respect to the issues of validity and reliability (Bachman, 2002; Norris, 2002, 2009; Shehadeh, 2012). A potential problem is that inferences may not be made beyond the specific target task and test context, which may seriously weaken the generalizability of the test results (Bachman, 2002), although, as pointed out by Norris (2009, 2016), this may entirely depend on the intended uses of the assessment, as task-based assessments are used for a variety of purposes, each emphasizing distinct qualities for design and validation. One implication of this variability is the reality that distinct frameworks, often in the form of task-dependent rating scales, may be called

for in order to adequately capture performance on different tasks for different assessment purposes (Norris, Brown, Hudson, & Yoshioka, 1998). Special attention in TBLA should thus also be devoted to the issue of interrater and intrarater reliability, that is, the consistency of the ratings between two or more raters and the consistency of one rater for the same test performance at different times (McNamara, 1996, 1997).

To date not many empirical studies have addressed the assessment of L2 pragmatics, or functional adequacy, from a task-based perspective. Our study aims to partially fill this gap. The main goal of the research is to examine the applicability of a rating scale of functional adequacy, as a component of L2 pragmatics and viewed as a task-related construct. The study focuses on the adequacy of the messages that the speaker and writer transmit to a particular interlocutor in a specific social setting, as described in the prompt of the target task. Recognizing the importance of assessing L2 pragmatics at discourse-level (Ross & Kasper, 2013; Roever, 2011), the study addresses the assessment of the functional adequacy of oral and written performance in extended discourse (i.e., a monologue task). Rating criteria, as recommended by González-Llorett (2016) and Norris (2001), were derived from the assessment task itself.

The construction of a rating scale for functional adequacy

In the literature, there has been no consensus so far on how functional adequacy as a construct should be defined (Iwashita et al., 2008). Functional adequacy has been interpreted in various ways: as successful information transfer (Upshur & Turner, 1995), pragmatic appropriateness (McNamara & Roever, 2007), successful task performance (De Jong et al. 2012a, b), or text coherence and cohesion (Knoch, 2009). In our study, functional adequacy, considered within the TBLT framework as a task-related construct, is defined as successful task completion, in line with De Jong et al. (2012a, b). In terms of the conversational maxims proposed by Grice (1975), the felicity and adequacy of the message in the text transmitted by the speaker/writer is judged by the receiver with respect to the quantity, relevance, manner, and quality of the text. The main focus in this definition is thus both on the social context and the target task (e.g., making a phone call to the doctor, writing a short note to a friend, taking part in a discussion) to be carried out by the speaker/writer and the reception by the listener/reader (Kuiken & Vedder, 2014, 2017; Kuiken, Vedder & Gilabert, 2010).

The scale that was used in the study at hand is an adaptation of the holistic rating scale that we employed in a previous study (Kuiken & Vedder, 2014; Kuiken, Vedder & Gilabert, 2010). In that study, we investigated the relationship between functional adequacy and linguistic complexity in L2 Dutch, Italian, and Spanish learners' writing samples. The participants involved in the study were 32 learners of Dutch, 39 learners of Italian, and 23 learners of Spanish, with a proficiency level ranging from A2 to B1. In that study, all participants completed two writing tasks, each consisting of a short

argumentative text. To collect baseline data, the same writing tasks were administered to a group of L1 writers. All texts were rated on a global six-point Likert scale by four native speakers of the languages involved. These raters were experienced instructors of the target language and can therefor be considered expert raters. Interrater reliability ranged from 'acceptable' to 'good' (Cronbach's α ranged between .70 and .90).

Based on the raters' comments on the scale descriptors of some of the rubrics, the holistic scale was split up into four subscales, each assessing one of the four components mentioned in the introductory section: (1) content, (2) task requirements, (3) comprehensibility, and (4) coherence and cohesion. Task requirements, not included among the descriptors of the earlier rating scale, were added as a separate scale dimension, in order to take into account the functionality of language use (i.e., the ability of the learner to accomplish specific tasks under particular conditions).

The requirements of the rating scale for functional adequacy are the following: (i) deconstruction of relevant components of functional adequacy; (ii) independence of descriptors of functional adequacy from linguistic descriptors in terms of complexity, accuracy, and fluency; (iii) objective and countable scale descriptors; (iv) applicability both for expert and non-expert raters; (v) the possibility to use the scale in both L2 and L1. The scale (a six-point Likert scale), inspired by the conversational maxims of Grice (1975), focuses on the quantity, relevance, manner, and quality of the speaker/writer's message transmitted to the listener/reader. The rating scale of functional adequacy, defined in terms of successful task completion by the speaker/writer in conveying a message to the listener/reader, thus comprises the following four scale dimensions (see Appendix A):

1. *Content: Is the number of information units (in terms of ideas, concepts, thematic elaboration) provided in the text adequate and relevant?*
This dimension takes into account (1) the adequacy of the number and type of information units in the text, and (2) their consistency and relevance independent from the specific requirements of the language task to be carried out.

2. *Task requirements: Have the requirements of the task been fulfilled successfully, with respect to text genre, register, and speech act?*
This dimension focuses on the extent to which the task is completed in accordance to the genre, register, and speech acts required in the message transmitted by the speaker/writer to the listener/reader, and the specific instructions and requirements of the task to be completed.

3. *Comprehensibility: How much effort is required of the listener/reader in order to understand the purpose and ideas in the text?*
This dimension takes into account the extent to which the message in the text is comprehensible for the listener/reader (Bridgeman et al., 2012; De Jong et al., 2012a, b).

4.   *Coherence and cohesion: Is the text coherent and cohesive?*

This dimension focuses on the adequacy of the message of the speaker/writer in terms of the occurrence of cohesive ties (presence or absence of deictic elements, anaphoric devices, and strategies), conjunction use, coherence breaks, number of repetitions (Knoch, 2007, 2009, 2011).

Rater behaviour may be influenced by teaching experience and rating practice. As demonstrated by several studies (Cumming, Kantor & Powers, 2002; Kuiken & Vedder, 2014; Schoonen, Vergeer & Eiting, 1997; Thompson, 1991), experienced raters (i.e., in this case teachers) are sometimes either more lenient or more strict compared to more 'naive' native speakers and may therefore be somewhat biased in their judgments. In Kuiken, Vedder and Gilabert (2010), only expert raters participated in assessing the functional adequacy of the written texts. In order to investigate whether the scale of functional adequacy could also be employed by other raters, we asked non-expert raters in a follow-up study to rate the written texts that had already been collected and assessed in our earlier study, with the exception of the texts of the Spanish L2 learners, which for practical reasons were left out (see Kuiken & Vedder, 2017, for a full account of this follow-up study). In the current study, our first aim is to evaluate the reliability and applicability of the rating scale of functional adequacy for oral L2 performance by non-expert raters. Secondly, we compare the use of the rating scale by non-expert raters for the spoken L2 data to the non-expert ratings of the written data (as reported in Kuiken & Vedder, 2017).

The central research question of the present study is: Can the rating scale of functional adequacy be used by non-expert raters for both written and spoken L2 performance? In order to answer this overall question the following sub-questions will be addressed:

– What are the interrater reliability scores of the raters on the four dimensions of functional adequacy?
– How are the judgments of raters on the four dimensions of functional adequacy correlated?
– How do the judgments of raters of functional adequacy correlate in the two tasks performed by the participants?

Methodology

In order to test out the applicability of the rating scale also for oral L2 performance by non-expert raters, spoken data were collected. This section describes the participants involved in the study, data collection, rating procedures, and the ways in which data were analyzed.

Participants

The materials collected in this study, both the spoken and the written data, were assessed by four raters, both for Dutch and for Italian. As mentioned in the previous section the written data had been collected in an earlier study (Kuiken & Vedder, 2017; Kuiken, Vedder, & Gilabert, 2010). Therefore, the raters who assessed the spoken data were not the same as those who had rated the written data. All raters were native speakers of the target language and university students of approximately the same age as the participants that were involved in the study. They did not have any specific experience in judging oral or written L2 production and could therefore be qualified as being non-expert.

L2 learners[1]

Raters were asked to assess samples that were produced by Dutch L2 and Italian L2 learners. The written texts that had been collected earlier (Kuiken & Vedder, 2017; Kuiken, Vedder, & Gilabert, 2010) were produced by 32 learners of Dutch and 39 learners of Italian, who were all university students. Their proficiency level in the target language ranged from A2 to B1. Spoken data were collected from 22 learners of Dutch and 26 learners of Italian, all university students (these learners were different from those who had performed the writing tasks). In terms of the CEFR (Council of Europe, 2001), the proficiency level of the Dutch learners differed from A2 to B2 and that of the Italian learners from A2 to C1.

Tasks

All L2 learners completed two tasks. In the written modality, learners were asked to write two short argumentative texts that involved pragmatic concerns. In the first task, they were asked to provide advice to the board of governors of the university regarding which non-governmental organization should receive a grant from the university. In the second task, they had to give advice to the editorial staff of a newspaper regarding the topic of the article to publish in the newspaper's monthly magazine (for a more detailed description of the tasks the reader is referred to Kuiken, Vedder, & Gilabert, 2010; see Appendix B for an example of one of the two target tasks). The oral tasks completed by these learners were similar to the writing tasks. The only difference was that participants were asked to produce the advice orally in the speaking tasks whereas in the writing tasks they were asked to produce advice in a written text (see Appendix C for two performance samples of the participants for Dutch L2 and Italian L2).

---

1.   Contrary to Kuiken, Vedder, and Gilabert (2010) and Kuiken and Vedder (2017) where both L2 and L1 learners were involved, the analysis of the spoken data did not include L1 learners.

Rating procedure

All texts (both written and spoken) produced by the participants were rated on a six-point Likert scale by non-expert raters based on the four dimensions of functional adequacy mentioned in the previous section: (1) content, (2) task requirements, (3) comprehensibility, (4) coherence and cohesion.[2] In order to become accustomed to working with the four scale dimensions of functional adequacy, two training sessions were organized with the raters of each language, for both modalities (written and oral). During these sessions, the underlying principles and use of the rating scale were explained and raters were trained by means of practice samples.

Data analysis

Interrater reliability was assessed by means of intraclass correlations among both the four raters of Dutch and the four raters of Italian. In order to examine the extent to which raters' judgments on the four separate dimensions of functional adequacy corresponded with each other, Pearson's product-moment correlations were calculated. For each participant, the scores by the four raters were averaged, for each scale dimension. The distribution of these average scores was examined to ensure normality and potential outliers. Next, Pearson's product-moment correlations coefficients were computed to determine the association between the four different dimensions. The average scores of the raters on task 1 and task 2 were then compared by means of Pearson correlations in order to establish the extent to which raters corresponded in their judgments of the two texts produced by the participants.

Results

In order to answer our research questions we will first describe the interrater reliability obtained by the raters of the oral and written samples for Dutch and Italian. We then present the correlations between the four dimensions of functional adequacy for both written and spoken data. Subsequently, results on the extent to which the raters concurred in their judgments of task 1 and task 2 are reported. The main outcomes of a retrospective panel discussion following the rating sessions are then discussed, in order to investigate the ways in which the raters perceived the four dimensions of functional adequacy. Finally, two examples are presented of the actual ratings that were assigned by the raters, one for Dutch and one for Italian.

---

2.   The scales used for rating the written and spoken texts were identical, except for some minor adaptations: 'text', 'writer' and 'reader' used in the scale for written production had been replaced by 'performance', 'speaker' and 'listener' in the scale for oral production.

Interrater reliability

To assess interrater reliability for both judgments on the written and the oral texts, intraclass correlations were calculated among the four raters of Dutch and those of Italian (see Table 1).[3] With respect to the assessment of the written texts, interrater reliability scores were moderate to high across all raters, ranging between .725 (for task requirements in Italian) and .940 (for comprehensibility in Dutch). For spoken samples, the reliability was similar or even higher, ranging from .864 (for comprehensibility in Dutch) to .929 (for coherence and cohesion in Italian).

Table 1. Intraclass correlations among raters for Dutch and Italian written and spoken texts

Dimension	Written texts		Spoken texts	
	Dutch L2/L1	Italian L2/L1	Dutch L2	Italian L2
Content	.841	.838	.878	.924
Task requirements	.824	.725	.890	.908
Comprehensibility	.940	.901	.864	.893
Coherence/cohesion	.860	.867	.876	.929

Correlations between dimensions of functional adequacy

Whereas the Italian L2 participants received overall higher mean scores than the Dutch L2 participants in writing, this was not the case in speaking. Standard deviations were higher for both writing and speaking for the Italian L2 participants, indicating that in Italian L2 there appeared to be more inter-individual variation than in the Dutch L2 group (see Table 2).

Table 2. Mean rating scores and SD's for Dutch and Italian written and spoken texts

Dimension	Written texts				Spoken texts			
	Dutch L2 ($N = 32$)		Italian L2 ($N = 39$)		Dutch L2 ($N = 22$)		Italian L2 ($N = 26$)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Content	2.92	.46	3.87	.70	3.62	.84	3.53	1.17
Task req.	3.04	.59	4.32	.69	4.44	.95	3.82	1.05
Compr.	3.09	.47	3.80	.73	4.07	.68	4.24	.91
Coh./coh.	2.84	.45	4.94	.61	3.65	.69	3.55	1.00

---

3. As in the study on the assessment of functional adequacy in writing both L2 and L1 writers were involved, it is important to note that the correlations reported for writing are based on the rating scores of L2 and L1 writers combined.

Pearson's product-moment correlations were calculated (see Table 3) to examine the extent to which the average judgments of the raters on each of the four separate dimensions of functional adequacy corresponded. All correlations were statistically significant, both for the written and the spoken texts. The correlations ranged from moderate (.544 for task requirements and comprehension in written Italian samples) to strong (.950 for content and coherence/cohesion in spoken Dutch samples). The relations between content and the other scale dimensions were particularly strong, as were those between comprehensibility and the other dimensions of the scale. On the whole, lower correlations were found between task requirements and other dimensions of functional adequacy.

Table 3. Pearson's product-moment correlations between dimensions of functional adequacy for Dutch and Italian written and spoken texts

Dimension	Written texts						Spoken texts					
	Dutch L2/L1			Italian L2/L1			Dutch L2			Italian L2		
	T.R.	Com	C&C	T.R.	Com	C&C	T.R.	Com	C&C	T.R.	Com	C&C
Cont.	.848**	.814**	.880**	.710**	.844**	.877**	.849**	.799**	.950**	.957**	.707**	.899**
T.R.		.694*	.851*		.544**	.559**		.646**	.774**		.706**	.915**
Com			.873*			.938**			.852**			.822**

*p < .05
**p < .01
***p < .001

Correlations between task 1 and task 2

Finally, we investigated the extent to which the average judgments of the raters on the two tasks written by the participants corresponded, both for the written and the oral modality. On the basis of the calculation of Pearson's correlation coefficients, it can be concluded that the correlations between the two tasks are in all cases high for the four dimensions of functional adequacy, both for writing and speaking, indicating that raters judged both texts largely in the same way (see Table 4). The lowest correlation was obtained on task requirements for written Italian (.455), the highest on comprehensibility for spoken Dutch (.906). All correlations were statistically significant.

The results of the study thus show that interrater reliability scores were generally high, both for writing and speaking, in Dutch and in Italian (Table 1). Mean scores in writing appeared to be higher in Italian than for Dutch, but not in speaking. The higher standard deviations in Italian L2, however, indicate that in Italian there appeared to be more inter-individual variation among the participants than in Dutch (Table 2). Correlations between the four scale dimensions in Dutch and Italian were significant in all

Table 4. Pearson's product-moment correlations between task 1 and task 2 for Dutch and Italian written and oral texts

Dimension	Written texts		Spoken texts	
	Dutch L2/L1 (N = 49)	Italian L2/L1 (N = 57)	Dutch L2 (N = 22)	Italian L2 (N = 26)
Content	.623***	.607***	.836***	.808**
Task Req.	.704***	.455***	.709***	.679**
Compreh.	.877***	.766***	.813***	.832**
Coh./Coh.	.719***	.802***	.906***	.830**

*$p < .05$
**$p < .01$
***$p < .001$

cases, both for writing and speaking, although lower correlations were found between task requirements and other dimensions of functional adequacy (Table 3). Rater judgments with respect to the average scores assigned to task 1 and task 2 turned out to be high in all cases in the two languages, both for writing and speaking, indicating that raters generally judged the two texts produced by the participants in largely the same way (Table 4).

Raters' perceptions of functional adequacy

In order to shed light on raters' reflections on the use of the scale and the ways in which they interpreted the four scale dimensions of functional adequacy, a retrospective panel discussion following the rating sessions was organized. During the panel discussion raters indicated that they had no difficulties to keep the various dimensions of the construct apart, as they considered them as being conceptually different. For content they indicated, for instance, that a text contained irrelevant information, such as personal opinions, or they mentioned that the position taken by the writer was not clearly described. Operationalized by considering the number and relevance of information units, content is distinct from the dimension of task requirements, which focuses on the extent to which the task has been carried out appropriately in accordance to genre, register, and speech act. Raters reported, for instance, that learners made use of creative and original arguments, that arguments were supported by means of examples, or that arguments were not convincing, in relation to social context, listener/reader, and task type. Comprehensibility was operationalized in terms of the amount of effort required from the interlocutor to understand the purpose and ideas expressed by the

speaker/writer. Raters indicated that minor mistakes did not hinder the overall comprehensibility of the text, that the listener or the reader needed much effort in order to understand the text, or even that the large number of errors discouraged from further listening/reading. Comprehensibility, as demonstrated by the retrospective panel discussion with the raters, could thus be assessed in a reliable and objective way. This was also true for coherence and cohesion (by considering the presence or absence of cohesive and anaphoric devices, strategies, coherence breaks, and repetitions), which could be assessed in a quantifiable way, as demonstrated by raters' comments that the use of connective devices ensured a good coherence of the text, or that the text contained explicitly established relations between the various components of the text.

The two text samples in Appendix C, derived from the written data for Dutch and the oral data for Italian, finally show that, despite high correlations between the four scale dimensions, in a number of cases raters may assign different ratings to the various components of functional adequacy. In Extract 1 for Dutch L2 (written text, task 2), the writer (D.L.) attempts to convince the members of the editorial board of a newspaper that the leading article for the monthly topic should be dedicated to the pros and cons of 'Animal experiments'. For content, and coherence and cohesion, the scores of the four raters (A, B, C, D) are identical (respectively 4, 2, 3, 3), while higher scores are assigned for the comprehensibility of the text (4, 3, 4, 5) and somewhat lower scores are given for task requirements (3, 1, 3, 3). The ratings of one of the raters (rater B) follow the same pattern, although the scores assigned by this rater to the various scale dimensions (2, 1, 3, 2) are on the whole lower than the judgments of the other three raters (see Appendix C). That raters consider the four components of functional adequacy as conceptually different, and assign different grades to these dimensions is shown also by Extract 2 for Italian L2 (oral text, task 2). In Extract 2, similarly to Extract 1, the speaker (B.C.) argues that the leading article for the monthly topic should be dedicated to 'Animal experiments'. Whereas comprehensibility is judged by the four raters for Italian (A, B, C, D) as being quite good (6, 4, 5, 5), task requirements (3, 3, 4, 4), coherence and cohesion (3, 3, 4, 4), and content (5, 2, 3, 3), with the exception of rater 1, on average receive lower scores (see Appendix C).

Discussion

In this study, functional adequacy as a specific component of pragmatic competence was defined within a TBLT and TBLA perspective as successful task fulfillment, in relation to the conversational maxims of Grice (1975) of quantity, relevance, manner, and quality. In order to assess the functional adequacy of oral and written L2 performance, we developed a six-point rating scale comprising four different scale dimensions: content, task requirements, comprehensibility, and coherence and cohesion.

The requirements of the rating scale were that the descriptors should be objective and countable, independent from CAF measures (i.e., complexity, accuracy, and fluency), and that application of the scale should be possible in L2 and L1, also by non-expert raters.

Kuiken, Vedder, and Gilabert's (2010) study showed that a preliminary version of the scale could be used by expert raters in assessing functional adequacy of written texts produced by L2 and L1 writers of Dutch, Italian, and Spanish, but that some adaptations of the scale were needed, particularly with respect to task requirements, not included in the earlier scale. Based on the findings of this previous study, a revised version of the rating scale was constructed, which was tested out for assessing written L2 and L1 texts of Dutch and Italian by non-expert raters in a second study (Kuiken & Vedder, 2017). As a follow-up study of the investigation of the written samples, the focus of the study discussed in this chapter was on the assessment of functional adequacy of spoken samples from Dutch and Italian L2 learners by non-expert raters, and on the comparison of the oral data with the findings of Kuiken and Vedder (2017) regarding the written data.

The results of the study at hand showed that functional adequacy in L2 can be reliably measured by means of a rating scale containing four different subscales by non-expert raters in writing as well as in speaking (Kuiken & Vedder, 2017, this study). As demonstrated by the present study, overall interrater reliability scores were moderate to high, both for writing and speaking. Investigation of the extent to which the judgments of the raters correlated on the four dimensions of functional adequacy showed that all correlations were moderate to high, both for the written and oral modality, although there are a number of cases in which raters assigned different scores to a particular scale dimension, as demonstrated by the examples discussed in the previous section. As shown also by the comments of the raters during the retrospective panel discussion, it is thus important to distinguish these four dimensions of functional adequacy from each other, as they are conceptually different. Moderate to high correlations were also found between the average ratings of the two tasks produced by the participants, both for writing and for speaking, indicating that raters were generally stable in their judgments.

Our overall research question (that is to say, can a rating scale developed for the assessment of functional adequacy be used by non-expert raters for the assessment of both written and spoken L2 performance?) can be answered affirmatively based on the findings of the study. The non-expert raters involved in our study were successful in familiarizing themselves with the scale in just two training sessions. The assessment of functional adequacy as a task- and context-related construct, by using a rating scale comprising distinct components of functional adequacy, may thus increase our understanding of task-based language assessment of L2 pragmatics. As interrater reliability turned out to be high, our study finally confirms that the rating scale for functional

adequacy is one step forward towards a reliable and objective tool for the assessment of this dimension of L2 pragmatics.

Conclusion

The present study raises a number of questions and issues that call for further investigation. First, it should be kept in mind that in our research the rating scale has been tested out for adult, highly educated language learners, who were assigned to one particular type of argumentative task. The applicability of the scale should also be tested out with other groups of learners (e.g., low-educated learners, adolescents), by means of different types of speaking and writing tasks, and with varying proficiency levels. It may be possible, for instance, that the differences in mean rating scores and standard deviations for Dutch and Italian written and oral texts have to be attributed to differences in the background of the learners: the Italian L2 learners were students of Italian with Dutch as their L1, whereas the Dutch L2 learners all came from different countries and varied in mother tongue. The learners of our study also varied in proficiency level: A2-B1 for the L2 learners of Dutch and Italian who performed the writing tasks; A2-B2 for the Dutch L2 learners who performed the speaking tasks and A2-C1 for the learners of Italian (who were in either their first, second or third year of study). Distinguishing different levels of L2 proficiency may lead to better insights into the acquisition of so-called 'diagnostic' textual features, which correspond to different levels of the Common European Framework of Reference (Hulstijn et al., 2010).

The rating scale for functional adequacy has recently been employed with some minor adaptations in a research project on L1/L2 writing development involving over 30 primary schools in northern Italy (Pallotti, 2017). The scale has also been tested out by means of an argumentative speaking and writing task assigned to native Italian university students (Cortés Velásquez, & Nuzzo, 2017), and to Chinese L2 learners of Italian, in a narrative task compared to an instruction information-gap task (Faone, Pagliara, & Vitale, 2017). In these studies, which were carried out among different types of learners and by means of various target tasks, the rating scale turned out to be a reliable and efficient instrument to be used in different settings, both in L2 and L1.

Another issue worth investigating further is the use of the rating scale in speaking tasks compared to writing. Although our results have demonstrated that the rating scale, which has initially been designed for the assessment of written texts, can also be used for speaking tasks, raters were inclined to think that judging coherence and cohesion in speech is harder than in writing, simply because it is easier to look back at utterances in a written text than in ongoing speech. It may also be the case that – contrary to the written data which were all offered in the same format and font to the raters – the rating of functional adequacy in spoken speech is influenced by

suprasegmental features like intonation, rhythm, and pitch (Vitale, De Meo, & Pettorino, 2012). Furthermore, it should be emphasized that, although (almost) any text is intended to be addressed to an implicitly present interlocutor, the rating scale in its present shape has been developed for the assessment of the functional adequacy of monologue tasks. In order to be able to test out the applicability of the rating scale for the assessment of dialogue tasks, it may be necessary to add one or more extra dimensions to the existing scale concerning the interactional abilities of the speaker, with respect to turn-taking, sequence organization, conversational repair, or use of contextualization cues.

Finally, the scale has a number of pedagogical advantages that allow teachers to provide L2 learners with specific feedback and focused comments on their achievements with respect to the four different dimensions of functional adequacy. This implies that the rating scale can be employed as a diagnostic tool. Whether the scale can also be used for summative performance assessment should be investigated in future research.

Acknowledgements

References

Austin, J. L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19, 453–476. doi:10.1191/0265532202lt240oa

Bardovi-Harlig, K. (2009). Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics. *Language Learning*, 59, 755–795. doi:10.1111/j.1467-9922.2009.00525.x

Barron, A. (2003). *Acquisition in interlanguage pragmatics. Learning how to do things with words in a study abroad context*. Amsterdam: John Benjamins. doi:10.1075/pbns.108

Bouton, L. F. (1994). Conversational implicature in the second language. Learned slowly when not deliberately taught. *Journal of Pragmatics*, 22, 157–167.
doi:10.1016/0378-2166(94)90065-5

Bouton, L. F. (1999). Developing non-native speaker skills in interpreting conversational implicature in English: Explicit teaching can ease the process. In E. Hinkel (Ed.), *Culture in second language teaching and learning* (pp. 47–70). Cambridge: Cambridge University Press.

Bridgeman, B., Powers, D., Stone, E., & Mollaun, P. (2012). TOEFL iBT speaking test scores as indicators of oral communicative language proficiency. *Language Testing*, 29(1), 91–108. doi:10.1177/0265532211411078

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

Cortés Velásquez, D., & Nuzzo, E. (2017). *Assessing L1 functional adequacy: Can we use the same scale as for L2*. Paper presented at TBLT 2017, Barcelona.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Cumming, A., Kantor, R., & Powers, D. E. (2002). Decision making while rating ESL/EFL writing tasks: A descriptive framework. *Modern Language Journal*, 86, 67–96. doi:10.1111/1540-4781.00137

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012a). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA* (pp. 121–142). Amsterdam: John Benjamins. doi:10.1075/lllt.32.06jon

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012b). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34(1), 5–34. doi:10.1017/S0272263111000489

Faone, S., Pagliara, F., & Vitale, G. (2017). *How to access L2 information-gap tasks through functional adequacy rating scales*. Paper presented at TBLT 2017, Barcelona.

González-Lloret, M. (2016). *A practical guide to integrating technology into task-based language teaching*. Washington, DC: Georgetown University Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts* (pp. 41–58). New York, NY: Academic Press.

Housen, A., Kuiken, F., & Vedder, I. (Eds.) (2012). *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins. doi:10.1075/lllt.32

Hulstijn, J. H., Alderson, J. C., & Schoonen, R. (2010). Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them? In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 11–20). (Eurosla Monographs Series 1).

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29(1), 24–49. doi:10.1093/applin/amm017

Knoch, U. (2007). 'Little coherence, considerable strain for reader': A comparison between two rating scales for the assessment of coherence. *Assessing Writing*, 12(2), 108–128. doi:10.1016/j.asw.2007.07.002

Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275. doi:10.1177/0265532208101008

Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, 16(2), 81–96. doi:10.1016/j.asw.2011.02.003

Kuiken, F., & Vedder, I. (2014). Rating written performance: What do raters do and why? *Language Testing*, 31(3), 329–348. doi:10.1177/0265532214526174

Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing. Towards a new rating scale. *Language Testing. Language Testing*, 34(3), 321–336.

Kuiken, F., Vedder, I., & Gilabert, R. (2010). Communicative adequacy and linguistic complexity in L2 writing. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 81–100). (Eurosla Monographs Series 1).

Long, M. H. (2015), *Second language acquisition and task-based language teaching*. Malden, MA: Wiley Blackwell.

Martinez-Flor, A., & Usó-Juan, E. (Eds.) (2010). *Speech act performance: Theoretical, empirical, and methodological issues*. Amsterdam: John Benjamins. doi:10.1075/lllt.26

McNamara, T. (1996). *Measuring second language performance*. New York, NY: Addison Wesley, Longman.

McNamara, T. (1997). 'Instruction' in second language performance assessment: Whose performance? *Applied Linguistics*, 18(4), 446–466. doi:10.1093/applin/18.4.446

McNamara, T., & Roever, C. (2007). *Testing: The social dimension*. Malden, MA: Blackwell.

Norris, J. M. (2001). Identifying rating criteria for task-based EAP assessment. In T. Hudson & J. D. Brown (Eds.), *A focus on language test development: Expanding the language proficiency construct across a variety of tests* (Technical Report #21, pp. 163–204). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.

Norris, J. M. (Ed.) (2002). Special issue: Task-based language assessment. *Language Testing*, 19(4). doi:10.1191/0265532202lt234ed

Norris, J. M. (2009). Task-based teaching and testing. In M. Long & C. Doughty (Eds.), *Handbook of second and foreign language teaching* (pp. 578–594). Cambridge: Blackwell. doi:10.1002/9781444315783.ch30

Norris, J. M. (2016). Current uses for task-based language assessment. *Annual Review of Applied Linguistics*, 36, 230–244.

Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments* (Technical Report #18). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.

Pallotti, G. (2017). Percorsi di educazione linguistica efficace per ridurre le diseguaglianze. In M. Vedovelli (Ed.), *L'italiano dei nuovi italiani*. Atti del XIX Convegno Nazionale GISCEL (pp. 505–520). Roma: Aracne.

Roever, C. (2005). *Testing ESL Pragmatics: Development and validation of a web-based assessment battery*. Berlin: Peter Lang.

Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463–481. doi:10.1177/0265532210394633

Roever, C. (2012). What learners get for free: Learning of pragmatic formulae in ESL and EFL environments. *The ELT Journal*, 66(1), 10–21. doi:10.1093/elt/ccq090

Ross, S. J., & Kasper, G. (Eds.) (2013). *Assessing second language pragmatics*. Houndmills: Palgrave, Macmillan. doi:10.1057/9781137003522

Schoonen, R., Vergeer, M., & Eiting, M. (1997). The assessment of writing ability: Expert readers versus lay readers. *Language Testing*, 14(2), 157–184. doi:10.1177/026553229701400203

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press, doi:10.1017/CBO9781139173438

Searle, J. R. (1975). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.). *Syntax and semantics 3: Speech acts* (pp. 59–82). New York, NY: Academic Press.

Shehadeh, A. (2012). Task-based language assessment: Components, developments, and implementation. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge guide to second language assessment* (pp. 156–163). Cambridge: Cambridge University Press.

Taguchi, N. (2009). Corpus-informed assessment of L2 comprehension of conversational implicatures. *TESOL Quarterly*, 43, 738–749. doi:10.1002/j.1545-7249.2009.tb00202.x

Taguchi, N. (2011). The effect of L2 proficiency and study-abroad experience in pragmatic comprehension. *Language Learning*, 61, 904–939. doi:10.1111/j.1467-9922.2011.00633.x

Taguchi, N. (2012). *Context, individual differences and pragmatic competence*. New York, NY: Multilingual Matters.

Taguchi, N. (2013). Production of routines in L2 English: Effect of proficiency and study-abroad experience. *System*, 41, 109–121. doi:10.1016/j.system.2013.01.003

Thompson, I. (1991). Foreign accents revisited: Factors relating to transfer of accent from the first language to a second language. *Language and Speech*, 24(3), 265–272.

Upshur, J. A., & Turner, C. E. (1995). Constructing rating scales for second language tests. *ELT Journal*, 49(1), 3–12. doi:10.1093/elt/49.1.3

Vitale, M., De Meo, A., & Pettorino, M. (2012). Foreign accent and persuasiveness. Native and non native voices in a radio spot. In A. De Meo & M. Pettorino (Eds.), *Prosodic and rhythmic aspects of L2 acquisition. The case of Italian* (pp. 213–222). Newcastle upon Tyne: Cambridge Scholars.

Walters, F. (2013). Interfaces between a discourse completion test and a conversation analysis informed test of L2 pragmatic competence. In S. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 172–195). Houndmills: Palgrave Macmillan. doi:10.1057/9781137003522.0014

Appendix A. Scale for rating the functional adequacy of oral tasks

Content: Is the number of information units provided in the text adequate and relevant?

1	2	3	4	5	6
The number of ideas is *not at all adequate* and insufficient and the ideas are unrelated to each other.	The number of ideas is *scarcely adequate*, the ideas lack consistency.	The number of ideas is *somewhat adequate*, even though they are not very consistent.	The number of ideas is *adequate* and they are sufficiently consistent.	The number of ideas is *very adequate*, they are very consistent to each other.	The number of ideas is *extremely adequate* and they are very consistent to each other.

Task Requirements: Have the task requirements been fulfilled successfully (e.g. genre, speech acts, register)?

1	2	3	4	5	6
None of the questions and the requirements of the task have been answered.	*Some (less than half)* of the questions and the requirements of the task have been answered.	*Approximately half* of the questions and requirements of the task have been answered.	*Most (more than half)* of the questions and the requirements of the task have been answered.	*Almost all* the questions and the requirements of the task have been answered.	*All* the questions and the requirements of the task have been answered.

Comprehensibility: How much effort is required to understand text purpose and ideas?

1	2	3	4	5	6
The performance is *not at all comprehensible.* Ideas and purposes are unclearly stated and the efforts of the listener to understand the text are ineffective.	The performance is *scarcely comprehensible.* Its purposes are not clearly stated and the listener struggles to understand the ideas of the speaker. The listener has to guess most of the ideas and purposes.	The performance is *somewhat comprehensible,* some sentences are hard to understand at a first listening. A second attempt helps to clarify the purposes of the speech and the ideas conveyed, but some doubts persist.	The performance is *comprehensible,* only a few sentences are unclear but are understood, without too much effort, after a second listening.	The performance is *easily comprehensible* and flows smoothly. Comprehensibility is not an issue.	The performance is *very easily comprehensible* and highly fluent. The ideas and the purpose are clearly stated.

Coherence and cohesion: Is the text coherent and cohesive (e.g. cohesive devices, strategies)?

1	2	3	4	5	6
The performance is *not at all coherent.* Unrelated progressions and coherence breaks are very common. The speaker does not use any anaphoric device. The speech is *not at all cohesive.* Connectives are hardly ever used and ideas are unrelated.	The performance is *scarcely coherent.* The speaker often uses unrelated progressions; when coherence is achieved, it is often done through repetitions. Only a few anaphoric devices are used. There are some coherence breaks. The speech is *not very cohesive.* Ideas are not well linked by connectives, which are rarely used.	The performance is *somewhat coherent.* Unrelated progressions and/or repetitions are frequent. More than two sentences in a row can have the same subject (even when the subject is understood). Some anaphoric devices are used. There can be a few coherence breaks. The speech is *somewhat cohesive.* Some connectives are used, but they are mostly conjunctions.	The performance is *coherent.* Unrelated progressions are somewhat rare, but the speaker sometimes relies on repetitions to achieve coherence. A sufficient number of anaphoric devices is used. There may be some coherence breaks. The performance is *cohesive.* The speaker makes good use of connectives, sometimes not limiting this to conjunctions.	The performance is *very coherent:* when the speaker introduces a new topic, it is usually done by using connectives or connective phrases. Repetitions are very infrequent. Anaphoric devices are numerous. There are no coherence breaks. The performance is *very cohesive* and ideas are well linked by adverbial and/or verbal connectives.	The speaker ensures *extreme coherence* by integrating new ideas in the performance with connectives or connective phrases. Anaphoric devices are used regularly. There are few incidences of unrelated progressions and no coherence breaks. The structure of the speech is *extremely cohesive,* thanks to a skillful use of connectives (especially linking chunks, verbal constructions and adverbials), often used to describe relationships between ideas.

Appendix B. Target task (Task 2; writing)

Every month your favourite newspaper invites its readers to have a say in what will be the leading article for their monthly supplement. This time the Editorial Board has come up with three suggestions:

1. the effects of global warming;
2. the importance of physical education;
3. pros and cons of animal experiments.

Out of these three suggestions one has to be selected. The selection is made by a Readers Committee. Every member of the committee has to write a report to the editors in which s/he states which article should be selected and why. On the basis of the arguments given by the committee members the Editorial Board will decide which article will be placed on the front page.

This month you have been invited to be a member of the Readers Committee. Read the brief descriptions of the suggestions for articles below. Determine which article should be on the front page and why. Write a report in which you give at least three arguments for your choice. Try to be as clear as possible and include the following points in your report:

−   which article should be selected;
−   what the importance of the article is;
−   which readers will be interested in the article;
−   why the editorial board should place this article on the front page of the Special Magazine (give three arguments).

You have 35 minutes to write your text and you need to write at least 150 words (about 15 lines). The use of a dictionary is not allowed.

Appendix C. Written and oral text samples (Dutch and Italian)

Extract 1. Written text (D.L.)
De meeste mensen willen niet over dierproeven weten. Naturlijk is het dan makkelijker om niks tegen deze praktijk te doen. Daarom vind ik het belangrijk om het artikel over dierproeven als hoofdartikel te kiezen. Het is een onderwerp waarover meer aandacht moet besteden worden. Mensen moeten bewuster zijn van de afschuwelijke dingen die iedere dag in de laboratoriums gebeuren. De behandeling van dieren als voorwerpen hoort niet bij het tegenwoordige leven; het is iets van vroeger, en het moet eindelijk voorgoed verdwijnen. Zullen de lezers van de krant in deze onderwerp geïnteresseerd zijn? Misschien niet; de meeste willen het vermijden. Maar dat is precies de reden waarom deze artikel zo belangrijk is. Zonder de steun van de kranten, en de pers in het algemeen, zal de strijd tegen dierproeven steeds moeilijker worden.

Translation of written text (D.L.)
Most people do not want to know about animal experiments. Of course, it's easier to do nothing against this practice. Therefore, I think it's important to choose the article on animal trials as a main article. It is an issue that needs more attention. People need to be more aware of the horrible things that happen in the labs every day. The treatment of animals as objects does not belong to present life;

it's something from the past, and it should disappear forever. Will the readers of the newspaper be interested in this topic? Maybe not; most people want to avoid it. But that's exactly why this article is so important. Without the support of the newspapers, and the press in general, the fight against animal testing will become always more difficult.

Ratings

Rater	Content	Task requirements	Comprehensibility	Coherence and cohesion
A	4	3	4	4
B	2	1	3	2
C	3	3	4	3
D	3	3	5	3

Extract 2. Oral text (B.C.)
Eeh l'articolo che secondo me devono scegliere è quello sugli degli esperimenti sugli animali perché so che ci sono stati diversi esperimenti sugli animali, ma non è ancora molto chiaro, almeno non per me, se è un bene o no. E poi vorrei anche sapere di più su cosa stanno studiando quali animali, credo che sia molto importante non solo per le persone per gli scienziati ma anche per l'uomo comune che sì, in quel caso sapranno di più su questi esperimenti. È anche importante che tipo di esperimenti fanno, perché cioè a me piacciono tanto gli animali ma non voglio che loro li fanno del male. Quindi credo che sia importante per me per le persone che sono interessate a questo argomento ma anche per gli scienziati, magari anche per i bambini, quindi sì può si può sapere di più credo che questo argomento dovrà apparire in prima pagina del del del giornale.

Translation of oral text (B.C)
The article that I think to choose is that of animal experiments because I know there have been several experiments on animals, but it is not very clear, at least not for me, whether it is good or not. And then I would also like to know more about what they are studying about these animals, I think it is very important not only for people for scientists but also for ordinary people who will in that case know more about these experiments. It is also important what kind of experiments they do, because that is, I mean why I like to animals, but I do not want them to hurt them. So I think it's important to me for people who are interested in this topic but also for scientists, maybe even for kids so yes you know you can know more I think this topic will have to appear on the front page of the journal.

Ratings

Rater	Content	Task requirements	Comprehensibility	Coherence and cohesion
A	5	3	6	3
B	2	3	4	3
C	3	4	5	4
D	3	4	5	4

CHAPTER 12

Pragmatics in task-based language assessment

Opportunities and challenges

Veronika Timpe-Laughlin
Educational Testing Service

Several researchers have pointed out synergies between task-based language assessment (TBLA) and L2 pragmatics assessment, insofar as both domains are based on a concern for effective communication in context (Kim & Taguchi, 2015; Norris, 2016; Timpe-Laughlin, Wain, & Schmidgall, 2015). It is therefore surprising that very little research has been carried out on pragmatics in TBLA. Kim and Taguchi (2015), for example, highlighted that "[p]ragmatics has been a particularly underinvestigated area of task-based research" (p. 660), both in teaching and assessment – a paradox given that the task-based assessment paradigm provides a fitting framework for the assessment of L2 pragmatics. Bringing together the two lines of research – L2 pragmatics and TBLA – this chapter will highlight similarities between the basic tenets of both domains, while discussing opportunities and challenges for employing TBLA as a framework for designing assessments that measure L2 pragmatics. To that end, the chapter begins with a brief review of the fundamental concepts of task-based assessment, highlighting the role of the task as a basic unit of analysis in designing tests that include a focus on pragmatics. It discusses challenges related to generalizability, reliability, and validity as well as issues of task design and task difficulty. It then canvasses current uses of TBLA to assess L2 pragmatics, illustrating different types of task-based assessments that have included pragmatic phenomena, both in research and in operational testing. Ultimately, it argues that, as a foundation for assessing L2 pragmatics, tasks offer a number of benefits that, despite the challenges, are worth pursuing.

Introduction

Language test developers are increasingly facing demands to include in their assessments features that allow inferences about a test taker's abilities to *use* the target language in interactive communication, that is, in relation to different interlocutors and situations. According to discourse-based studies, employers and academic institutions have noted that although test-takers may achieve high scores on standardized assessments, they are not always able to communicate effectively and appropriately with

interlocutors in different situations (e.g., Clyne 1994; Holmes, 2000). For example, in a study focusing on migrant workers in New Zealand, Holmes (2000) found that employers believed that all workers had sufficient second/foreign (L2) abilities to perform their jobs (in fact, they had all obtained a fairly proficient average score of 6.5 on the IELTS test). Yet, "they seem unfriendly or uncomfortable at work; they don't seem to fit in smoothly" (Holmes, 2000, p. 9). That is, rather than mentioning transactional aspects of language use, most of the employers identified relational talk (i.e., social action and interaction) as the concern – and these are, of course, central aspects of second and foreign language pragmatics.

Pragmatics, as the study of language use in context (Crystal, 1997), has at its core the ability to map form and function with the goal of using language effectively and appropriately relative to a certain communicative context (see also Timpe-Laughlin et al., 2015). The success of form-function-context-mappings, both in production and comprehension, is defined by expectations of the specific target discourse community. Therefore, pragmatic performance is highly context-specific and variable (i.e., not only across contexts but also cultures, personalities, etc.). Moreover, it constitutes an ability that may depend to a large extent on experience and exposure to relevant contexts.

This protean nature of pragmatics presents a number of challenges for the assessment of pragmatic ability in terms of fairness, test design, and generalizability. For example, can experience and cultural knowledge be assumed – both of which are arguably needed to achieve pragmatic competence, yet are mostly gained from exposure? Or do these factors create an issue of fairness given that not every L2 learner can, for example, afford to travel and experience first-hand interaction with representatives from the target discourse community? Second, in something as variable and context-dependent as pragmatics, how do you make inferences beyond the test task? Finally, large-scale assessments often use multiple parallel forms, which need to produce similar results. Thus, is it feasible, especially for large-scale assessments, to describe large numbers of highly-specific contexts that elicit comparable pragmatic performances?

Given these challenges that L2 pragmatics assessment faces, task-based language assessment (TBLA), with the task as the basic unit of analysis for test design, may constitute a fitting framework to assess this context-dependent, (inter)action-oriented skill. As a subset of integrated and direct language performance assessment, TBLA is built upon communicative language testing (Brindley 2009; Timpe, 2013a) and thus shares the concern for effective communication in context. As a potentially suitable paradigm for assessing language use in context, TBLA may offer opportunities in response to the challenges outlined above. For instance, by focusing on specific tasks, generalization becomes less of an issue, and very specific domains (i.e., a given task) can be identified which in turn simplifies test design (see e.g., Bachman, 2002; Norris, 2009). Also, if a specific task/context is the unit of measure, then it becomes more reasonable to require culture-specific knowledge

that is typically called for in such a task/context. Overall, focusing on the intersection between TBLA and L2 pragmatics assessment, these and other ideas will be discussed and explored further in this chapter.

In the following, I will address three key aspects that are crucial to consider when adopting TBLA as a framework for assessing L2 pragmatic performance: (1) the task and task characteristics, (2) the selection of assessment tasks, and (3) the rating criteria to evaluate task performance. Following the section on tasks and task characteristics, I will review task-based pragmatics assessments in practice, focusing on language policies and standards as well as the assessment of pragmatic phenomena in language for specific purpose assessments, in educational assessments, and in assessment research. The chapter will conclude with a look at large-scale assessments as well as ideas for a research agenda that could advance task-based pragmatics assessment (TBPrA).

Tasks and task characteristics

Bringing together two lines of research – L2 pragmatics and TBLA – first requires some unraveling of terminology, in particular with regard to the fundamental unit of analysis: tasks. Traditionally, pragmatics researchers have assessed different subcomponents of L2 pragmatic knowledge by means of various discrete-skill measures referred to as discourse completion *tasks* (e.g., Roever, 2006; Timpe, 2013a), appropriateness judgment *tasks*, and dialogue choice *tasks* (e.g., Roever, Fraser, & Elder, 2014) (see Chapter 1). The term 'task' is used synonymously with 'item type' in the L2 pragmatics (assessment) literature and thus denotes basically any type of assessment activity. By contrast, the definition of 'task' used in the TBLA domain is much more focused. In task-based research and educational practice, tasks are generally defined as those real-life "activities that people do in everyday life and which require language for their accomplishment" (Norris, Brown, Hudson, & Yoshioka, 1998, p. 33) such as writing an academic essay or email, giving a presentation, hosting a business meeting, or calling a client. Thus, tasks, as understood in task-centered research, foreground authentic language use (across modalities) for interpreting and expressing meaning in real-world contexts (Kim & Taguchi, 2015; Norris, 2016; Timpe-Laughlin et al., 2015) – all key dimensions that also feature prominently in L2 (discursive) pragmatics.

Most approaches to TBLA share this underlying premise that intended inferences about learners' ability to *use* language and to perform form-function-context mappings draw upon skills and knowledge (e.g., Bachman, 2002; Brindley, 1994; Ellis, 2003). For example, Brindley (1994) clearly highlights the language use aspect, describing the TBLA paradigm as "the process of evaluating, in relation to a set of explicitly stated criteria, the quality of the *communicative performances elicited from learners as part of goal-directed, meaning-focused language use requiring the integration of skills and*

knowledge" (p. 74; my italics). Similarly, Ellis (2003) maintains that TBLA constitutes "a way of eliciting and evaluating *communicative performances from learners in the context of language use that is meaning-focused and directed towards some specific goal*" (p. 279; my italics). Hence, TBLA highlights the task as a vehicle for eliciting L2 learners' performances, allowing them to show what they "can do with the language" (Brown, Hudson, Norris, & Bonk, 2002, p. 5).

In order for learners to fully show what they can do, assessment tasks need to be designed in ways that provide them with the opportunity to engage in situated communicative interaction in which they need to use language, including pragmatic moves, in order to accomplish a particular goal. For example, writing an email to a professor requesting an extension of a term paper requires form-function-mapping processes insofar as the L2 learner needs to attend to verbalizing the request in relation to contextual features such as the relationship to the professor, the degree of imposition associated with the request, and the mode of communication. Hence, tasks need to be carefully designed in order to elicit pragmatic phenomena in authentic, goal-directed, and meaning-focused L2 performance. Moreover, they need to be designed systematically in order to provide for comparability between tasks with regard to issues such as task complexity and difficulty. However, this is easier said than done given that we still know very little about task design features and their potential impact on the elicitation of L2 pragmatic moves (but see Youn, this volume).

Especially for assessment purposes it is paramount to investigate systematically and in more detail the complex interplay between task design and demands that affect L2 pragmatic output. Task complexity variables such as linguistic input (range, number of input sources), mode of communication, genre, number of interlocutors, relationships between interlocutors, and cognitive aspects such as input availability create highly contextualized environments that situate language performance (Norris et al., 1998). However, thus far few studies within task-based research have actually focused on the intersection between task design variables or task characteristics and L2 pragmatics output. For example, Taguchi (2007) reported more appropriate requests and refusals in situations of equal power, low imposition, and small social distance than in those that featured large differences in contextual variables between interlocutors, thus revealing an impact of situational variation on task difficulty. Another study by Gilabert and Barón (2013) investigated the impact of task complexity on the number of pragmatic moves. They found that a more complex task in terms of cognitive demands elicited more pragmatic moves than did a simpler one, indicating a clear task effect. However, they did not find a task effect on the variety of moves. That is, complex tasks did not necessarily elicit a greater variety of pragmatic moves. Although these studies have provided first insights into the complex relationships among task design, test-takers, and elicited performance, it is still largely unclear which elements in a context activate and thus elicit pragmatic knowledge and performance (Roever &

McNamara, 2006). In other words, there is still much to be learned about (a) which task design variables motivate pragmatic-related output and (b) whether it is possible to achieve task comparability by operationalizing these variables in a systematic way to elicit pragmatic moves. Or, is it unrealistic to aim for the systematic representation of comparable, complex, real-world communicative environments? To provide responses and insights into these questions and advance the field of TBPrA, it is paramount to further investigate how task characteristics affect test-takers' interaction with the assessment task and impact their pragmatic performance – a relationship that is also fundamental to any type of generalizations that are to be made from a given task-based assessment (Mislevy, Steinberg, & Almond, 2002).

Task selection and development in TBLA

The systematic selection and development of assessment tasks constitutes a crucial link between tasks on the one hand, and the intended uses and purposes of a given assessment on the other. A key question then is: What kind of evidence or observable behavior do we need to elicit to make valid judgments about the pragmatic knowledge and/or ability we want to measure? Thus, it is key to anchor the contexts embedded in assessment tasks to the target domain and to create situations that elicit the same pragmatic knowledge and performance that a person would require in a given criterion situation (Roever & McNamara, 2006). Kasper and Ross (2013), for instance, note that "[s]ocial and cultural appropriateness anchors communicative competence firmly in the social world and enables relevant descriptions of the target use domain in language assessment contexts" (p. 5). Thus, without specification of the context it is not possible to rigorously identify the features of an "appropriate" performance. Accordingly, task selection and development should start with a review of real-world communicative situations in a specific target language use (TLU) domain identified in the test specifications. In other words, it should start with an understanding of the contexts and language use we want to make inferences about in order to attain "a close correlation between the test performance, i.e. what the testee does during the test, and the criterion of performance, i.e. what the testee has to do in the real world" (Ellis, 2003, p. 279).

Given the focus on performance in TBLA, rather than on measurement of a psychological construct, Norris (2000) proposed a framework that can provide conceptual guidance for task development and selection. As shown in Figure 1 below, the central element, intended assessment use, is specified by four interconnected components involved in the assessment process: (a) stakeholders or users of a given assessment, (b) information a certain assessment is supposed to provide, (c) the purpose a given assessment serves, and (d) the impact and consequences of the assessment. Therefore, the development of any type of task-based pragmatics assessment should start with a

review of the TLU context and potentially a needs analysis in order to answer these fundamental questions about intended use.
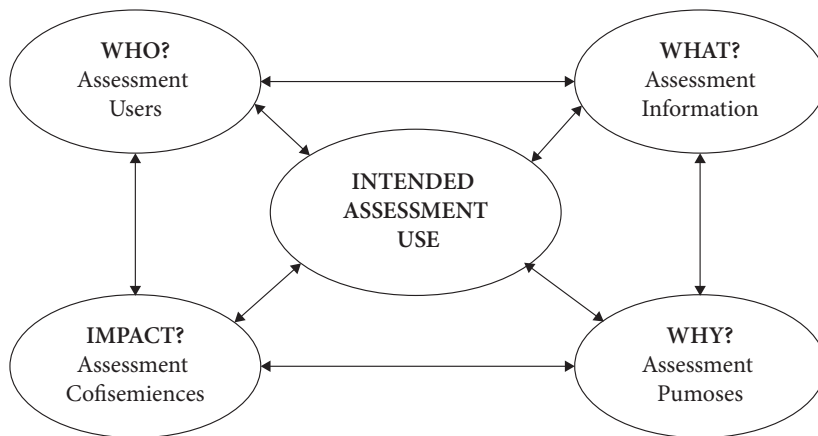
Figure 1. Specification of intended assessment use (adopted from Norris, 2000)

A comprehensive example from the L2 pragmatics literature that has used and operationalized Norris' framework is Youn (2008, 2015). Youn developed role-play tasks for the assessment of L2 discursive pragmatics in an English for Academic Purposes (EAP) setting. With the aim of using these tasks for instruction and assessment in classes offered by the English Language Institute (ELI) at the University of Hawai'i at Mānoa, Youn conducted a large-scale needs analysis on EAP L2 pragmatics learning needs in relation to that context. She first identified groups of stakeholders at the ELI. In a two-step process, she then administered semi-structured interviews and a survey questionnaire, gathering feedback and insights on L2 learners' EPA pragmatic needs from different stakeholder groups such as program administrators, instructors, and L2 learners. The findings informed the identification and development of EAP pragmatic assessment tasks, including decisions about task types, task specifications, task design elements, task difficulty, and intended uses (Youn, 2008). Thus, driven by the specific EAP use case, Youn followed the systematic approach of a needs analysis to inform task identification, selection, and development – an approach that subsequently contributed to the validation argument to justify the use of her assessment (Youn, 2015).

Developing rating criteria to evaluate task performance: Different approaches and challenges

Evaluative criteria constitute a crucial link between performance on a given task and inferences that are to be made based on the elicited performance. Task-based language

assessment serves the primary purpose of eliciting performance and drawing inferences about a test-taker's abilities (Bachman, 2002; Timpe, 2013a). Thus, the elicited task performance needs to be evaluated and rated according to pre-established criteria. The specification of criteria based on intended inferences and uses is key in the process of judging performances. However, as Quellmalz (1991) argued, the identification of evaluative criteria that can be used to distinguish different levels of expertise and ability is "[p]erhaps the greatest challenge facing proponents of performance assessment" (p. 319). While it constitutes a challenge for TBLA in general, it proposes an even greater challenge for the task-based assessment of L2 pragmatic ability – a domain in which the specification of benchmarks and evaluation criteria has traditionally been a thorny issue (Timpe-Laughlin et al., 2015).

First and foremost, it is essential to determine what kinds of behavior we need to elicit and observe in order to draw conclusions about a learner's pragmatic ability in a certain context. Given the protean character of pragmatics, researchers and language testers have traditionally struggled to clearly define what it means to *use* language appropriately and effectively and to operationalize the resulting definition for the purpose of assessment. Kim and Taguchi (2015), for example, argue that tasks need to elicit pragmatics-specific features such as certain speech acts, while also allowing form – function – context mappings given that L2 pragmatics "need to be assessed on these associations" (p. 660). However, these form – function – context mappings vary considerably across communicative situations and thus also across tasks. Moreover, McNamara and Roever (2006) pointed out that "[j]udgements of what is and what is not appropriate differ widely among [speakers] of a language and are probably more a function of personality and social background variables than of language knowledge" (p. 57). Nevertheless, in order to draw inferences about learners' pragmatic ability, yardsticks or evaluative criteria are necessary against which pragmatic meaning and meaning making utterances can be evaluated.

In the literature there has been a debate regarding the focus of assessment, featuring two perspectives on TBLA that have been labeled as a *weak* and a *strong* approach to performance assessment (Messick, 1994; Norris, 2001; Timpe, 2013a). Proponents of the *weak* form of TBLA hold a construct-centered view, arguing that the task is a vehicle that elicits a certain language performance which is then evaluated against criteria based on a given language model or framework. McNamara (1996) described weak performance assessments as those where "the capacity to perform the task is not actually the focus of assessment. Rather, the purpose of the assessment is to elicit a language sample so that second language proficiency, and perhaps additionally qualities of the execution of the performance, may be assessed" (p. 44). By contrast, the *strong* approach to TBLA proposes a task-centered view of assessment that regards language performance as necessary in order to accomplish the task. Norris (2001), for example, argued that "language performance serves as a vehicle for task accomplishment" (p. 167). In other words, the focus of assessment here is task accomplishment rather than

the measurement of a construct. Hence, the criteria used to evaluate task performance also need to be task-dependent.

Despite the theoretical distinction that is made between weak and strong forms of TBLA, the use of one approach over the other should be determined primarily on the use and purpose of a task-based assessment (Norris, 2009, 2016). For example, with respect to generalization, construct-centered approaches try to generalize across tasks to make inferences about an ability such as pragmatics. By contrast, strong TBLA does not attempt this generalization, and thus makes a more limited (but more targeted) inference about language ability; here, pragmatic ability is inferred only to the extent that, and in the specific ways that, it makes a difference in a given target-task performance. Hence, there seems to be an inherent tradeoff between the two approaches, and the selection of one paradigm over the other should ultimately be based on the purpose and use of an assessment.

Taking into account the challenges in identifying and operationalizing criteria to evaluate L2 pragmatic performances, the strong approach to TBLA may provide a number of opportunities. First, models of pragmatic competence notoriously under-theorize interaction (Kasper & Ross, 2013), a core element of performance-based TBLA. Thus, it is challenging to operationalize evaluative criteria based solely on a theoretical framework. Second, if the criteria for judging L2 pragmatic performance are based on a theoretical model instead of real-world needs, they may lack – at least to a certain degree – reference to real-world tasks and how people actually accomplish them (Norris, 2001). Third, the assessment of pragmatics hinges upon contextualization, an environment that a task or a sequence of tasks is more likely to provide. The context also constitutes a clear point of reference for the identification of criteria that determine the success of pragmatic performance. For example, responding via email to a complaint made by a customer, in order to provide the customer with a solution to a problem, could be judged on several criteria. For instance, the following could serve as evaluation criteria: appropriateness of the level of politeness in features such as address terms, salutation, adequate use of genre requirements, appropriate and effective use of pragmalinguistic forms that indicate knowledge of sociopragmatic features, and ulti-mately, the goal of the task, customer satisfaction (i.e., whether the customer feels that the response was provided in a professional and polite manner). Finally, the develop-ment of rating criteria based on the analysis of form-function-context mappings in a specific real-world language use environment would further sway criticism that has oftentimes been raised where evaluative criteria were identified based on either 'expert judgments' or the notorious 'native speaker norm.' That is, criteria could be identified by means of an analysis of language used in the real-world context.

Despite the advantages of task-centered L2 pragmatics assessment, some chal-lenges remain, especially with regard to reliability, validity, and generalizability. In terms of reliability and validity of potential inferences drawn on the basis of pragmatic

performance in a task, it is essential that (a) rubrics describe the relevant pragmatic features that characterize different levels of task performance and (b) raters are able to apply the pragmatics-related aspects of scoring rubrics in a consistent manner. Thereby, potential challenges may lie not only in the development and design of scoring rubrics, but also in a potential lack of raters' awareness of (L2) pragmatic phenomena. Language users – both L1 and L2 speakers – are oftentimes unaware of pragmatics as a language-related phenomenon (Washburn, 2001). While lexical or grammatical mistakes are usually perceived as L2 deficiencies, pragmatic failures tend to be attributed to a person's character or manners (Timpe, 2013b). Similarly, raters who may have a good conceptual understanding of lexico-grammatical and morpho-syntactic aspects of language, may struggle with rating form-function-context mappings. In order to ensure reliable ratings and eventually valid inferences, it is essential that raters have a profound (meta)pragmatic awareness and are well trained in applying rubrics.

In addition to challenges in terms of reliability and validity, generalizability has been a controversial issue in TBLA as well as L2 pragmatics assessment, and it continues to be a challenge when pragmatics is measured by means of TBLA. It can be debated whether TBLA can make warranted inferences on actual pragmatic competence beyond particular tasks or test contexts. If evaluative criteria are relative to a particular task that has been identified, developed, and operationalized in reference to a specific real-world context, can the observed pragmatic abilities be generalized across tasks and contexts? Given these challenges, it may be easy to conceptualize such a task-based approach to L2 pragmatics assessment in highly constrained, low-stakes assessment contexts, such as "for learners and teachers to generate and act upon feedback related to form-function-meaning relationships" (Norris, 2016, p. 241). However, could this approach also be used for large-scale, high-stakes assessments? A potential solution may lie in the careful selection, sampling, and sequencing of tasks representative of communicative demands in the real-world context. In the following section, I will briefly review different types of task-based assessments that have included pragmatic phenomena, both in research and in operational testing. The intent here is not to provide a comprehensive account, but rather to highlight some of the primary uses of TBPrA.

Task-based pragmatics assessments in practice

Although still in its infancy, TBPrA can be found in different contexts of L2 assessment, including in particular language policies and standards, language for specific purpose (LSP) assessments, educational assessments, and L2 assessment research. With regard to standards and language policies, Norris (2016) argued that several large-scale standards and language policies, including the *Common European Framework of Reference*

(Council of Europe, 2001), the *Hong Kong Target-Oriented Curriculum* (Curriculum Development Institute, 2005), and the *Standards for Foreign Language Learning* (National Standards in Foreign Language Education Project, 1996), have been developed and "founded on the basis of, or at least in substantial reference to, tasks" (p. 234). In addition to their task-based reference, they also highlight pragmatic ability as a key aspect of communicative language ability that should be taught and assessed by means of tasks. The CEFR, for example, identifies three key language competences as equally important for communicative language ability: linguistic ability, sociolinguistic ability (i.e., sociopragmatics), and pragmatic ability (i.e., the functional use of language), thus emphasizing the importance of language *use* and meaning in communicative interaction. They highlight that the "emphasis in a communicative task is on successful task completion and consequently the primary focus is on meaning as learners realise their communicative intentions" (Council of Europe, 2001, p. 158). The CEFR emphasizes meaning by citing tasks such as "interacting with a public service official and completing a form; reading a report and discussing it with colleagues in order to arrive at a decision on a course of action" (p. 158). In relation to these tasks, the CEFR outlines key aspects of pragmatic sensitivity such as audience awareness as well as other sociopragmatic and pragmalinguistic competencies that can be operationalized for the purpose of assessment within a given task. Moreover, it provides scales that can inform rubrics to score L2 pragmatic performances (see CEFR, p. 124ff.). Hence, language policy documents such as the CEFR draw upon tasks as a means of depicting language ability in general and pragmatic ability in particular, providing a basis for the development and validation of TBPrA (Norris, 2016).

TBPrA in language for specific purpose assessments

The most comprehensive form of TBPrA – oftentimes designed on the basis of language policy documents – can be found in LSP assessments such as the Canadian English Language Benchmark Assessment for Nurses (CELBAN) and the Australian Occupational English Test (OET). Both, the CELBAN and the OET are firmly grounded in a task-based assessment approach given that they meet the definition of "elicitation and evaluation of language use for expressing and interpreting [pragmatic] meaning within a well-defined communicative context (and audience), for a clear purpose, toward a valued goal or outcome" (Norris, 2016, p. 232). Moreover, they contain a considerable focus on measuring pragmatic phenomena across all skills. The OET, for example, constitutes an English for specific purposes (ESP) assessment. It is available for 12 different healthcare professions, including nursing, dentistry, medicine, optometry, dietetics, occupational therapy, pharmacy, physiotherapy, podiatry, radiography, veterinary science, and speech pathology. It measures listening, reading, speaking,

and writing by means of engaging test-takers in health-oriented tasks with supportive materials (*task realia*), relevant for a specific occupation. In the speaking section, for instance, test-takers are required to engage in two role-play tasks which aim to assess the test-taker's ability to communicate effectively in speech with patients or clients in common workplace situations. Assuming the roles of either patient or healthcare provider (e.g., a dentist, a nurse, a speech pathologist etc.), the role-play features clear (communicative) goals such as rephrasing ideas in different ways to help or persuade a patient or reassuring a worried or angry patient (see Elder, 2016 for a more detailed description). The speaking performances are independently scored by two raters who focus exclusively on the language, due to a law in Australia that requires the separate assessment of language and professional competences (Elder, 2016; Elder et al., 2013). Among the rating criteria for speaking are appropriateness of language use in relation to the goal of a given role-play task. Similarly, test-takers are required to write an email to a patient in the writing section of the assessment. In addition to task fulfilment and lexico-grammatical features, writing ability is scored on the basis of appropriateness of language use (i.e., appropriateness of lexis, register, tone), a rubric that also includes the following pragmatic phenomena: formality, register, tone, salutations, and aware- ness of target audience. Hence, the tasks included in the OET replicate content, proce- dures, interlocutors, and so on that are relevant in the real-life context, thus providing a high-degree of contextualization that allows for the scoring of pragmatic features and eventually the making of inferences about a test-taker's pragmatic competence and sensitivity.

Other examples of task-based LSP assessments include tests that certify job-related language abilities for service jobs such as call center agents and cab drivers (see Lockwood, 2015) as well as diagnostic assessments for teacher trainees and instruc- tors (Elder, 2001). Especially the 'Classroom Language Assessment Schedule' (CLAsS) constitutes an interesting example that includes TBPrA. The CLAsS is a diagnostic assessment that takes the real-life tasks an instructor faces in the classroom as the unit of analysis that is being evaluated. In other words, the L2 teacher works in the real-life classroom context while they are being observed and evaluated with regard to poten- tial English language issues they may face during their school-based teaching practice (see Elder, 2001). In addition to general language proficiency and subject-specific lan- guage use, the evaluation focuses on the language of classroom interaction, including features such as overall communicative effectiveness, appropriate use of a variety of forms of address, appropriate levels of formality/firmness, appropriate responses to student questions, and appropriate rejection and/or acceptance of student contribu- tions. In addition to a holistic rating of classroom interaction, the pragmatic phenom- ena are judged individually in terms of whether "they need work" (Elder, 2001, p. 169). Moreover, the rater can provide commentary with regard to strengths and weaknesses of each pragmatic feature (see Elder, 2001, p. 167–169 for the complete observation

protocol). While other LSP assessments usually rely on the careful selection of representative tasks, the test-task here is the lesson itself and thus "as natural a context as is possible" in which the test-taker's performance is observed (Elder, 2001, p. 155).

TBPrA in educational assessments

In addition to profession-specific LSP assessments, TBPrA can also be found in educational contexts. A key motivation for task-based assessment in language education is the need to align curriculum, instruction, and assessment in a way that they "complement each other in supporting effective language learning" (Norris, 2016, p. 237). A prime example for this approach in which TBPrA is used in a high-stakes, educational context is the evaluation of student learning outcomes in the German Department at Georgetown University. At Georgetown, the German Department has replaced the original form-focused normative approach underlying the undergraduate program with a "content-oriented collegiate foreign language curriculum" including testing in the form of task-based assessment (Byrnes, 2002, p. 419). With regard to the assessment of writing progress and achievement, for example, tasks are sequenced on the basis of course and curricular objectives. Although approaching the issue from the point of view of systemic functional linguistics (Byrnes, Maxim, & Norris, 2010), the evaluation of writing task performances includes, among other criteria, pragmatic aspects such as audience, register, and writing intentions and goals. The information obtained about learners' pragmatic abilities in writing is then used to make decisions about learners (achievement, certification of abilities, advancement, etc.) as well as to inform instruction (diagnosing learner needs, focusing students' learning, improving instruction, etc.).

TBPrA in assessment research

Outside of operational assessment contexts, the development of TBPrA has also received considerable attention in assessment research (e.g., Grabowski, 2009; Okada & Greer, 2013; Timpe, 2013b; Youn, 2008, 2013, 2015). Grabowski (2009), for example, designed role-play tasks and evaluated test-taker performances based on the effectiveness of sociolinguistic and sociocultural appropriateness. She included a range of pragmatic phenomena such as aspects of politeness, power differences between interlocutors, variations in register and modality, and use of formulaic expressions and collocations. She employed a 5-band holistic scoring rubric, ranging from 0 (no effective use) to 4 (effective use). Using many-facet Rasch measurement to evaluate the reliability of the rating scales, Grabowski (2009) found that the tasks elicited in a reliable

manner a range of pragmatic meanings that were scalable across different levels of language ability (see also Grabowski, 2013). In the context of an ESP writing assessment, Youn (2013, 2015) developed a 3-band rubric (1=inadequate, 2=able, 3=good) for the scoring of learner's pragmatic abilities in writing tasks, evaluating tone, genre, directness, politeness, salutation, and register. Similar to Grabowski (2009), Youn reported internal consistency between the ratings based on FACET's stable fit statistics which she interpreted as evidence supporting the generalization inference. She found that two out of 12 raters showed slight unpredictability, a finding which Youn attributed to a lack of detail in the scoring rubrics. For oral L2 abilities, Timpe (2013b) developed four different telephone role-play tasks, conducted via *Skype*, requiring students to interact with two types of interlocutors – a professor and a fellow student – while accomplishing different communicative goals such as arranging a meeting or asking for an extension on a term paper. Drawing upon Hudson, Detmer, and Brown (1995), Timpe scored the role-play performances in terms of interactional and pragmatic phenomena, including formality, directness, politeness, forms of address, closing, and turn-taking. Although Timpe reported inter-rater reliability estimates that ranged from .88 to .97 for pragmatic competence, anecdotal evidence from raters suggested that a more detailed scoring rubric may have been more helpful in further guiding and aligning ratings. Hence, although all researchers conducted rater trainings and obtained fairly high reliability estimates, it seems that detailed descriptions in the scoring rubrics for the rating of pragmatic phenomena are key to maximizing the reliability of ratings and ultimately the validity of inferences.

In addition to developing rubrics for these pragmatic phenomena, Youn (2013) and Timpe-Laughlin and Park (under review) used Conversation Analysis (CA) to examine pragmatic and interactional moves elicited in the task performances. Given the need to further explore the relationship between characteristics of assessment tasks and pragmatic performance elicited by a given task, CA could provide beneficial insights into "interactional details of the testing process" (Okada & Greer, 2013, p. 308), thus shedding light on the products and processes of TBPrA. Thus, CA could help to (a) demonstrate, for validation purposes, that test tasks elicit performance similar to performance in the real world and (b), in order to support scoring, empirically document the qualities of performance at different ability levels.

Taking into account the different examples of TBPrA across all use contexts, certain aspects can be noted. First and foremost, all of the TBPrA examples reviewed above feature well-defined communicative contexts, a clear purpose and (communicative) goal, and specific interlocutors – key features of TBLA and L2 pragmatics. Second, they all share 'appropriateness' of the form-function-context mappings as the benchmark across the different pragmatic phenomena that are being assessed (e.g., directness in speech act realization, salutations and forms of address, register). That is, 'appropriateness' can be seen as the overall criterion for evaluating task accomplish-

ment in terms of pragmatics. Similar to a common denominator, appropriateness can thereby be determined on the basis of performance observed in a range of proficient language users who engage in these situations in the real world – an opportunity that TBPrA provides due its high degree of contextualization. Third, each assessment features somewhat distinct pragmatic phenomena. That is, there is not a pre-determined set of pragmatic phenomena that need to be included in any assessment. Rather, the pragmatic features that are scored need to be determined on the basis of the pragmatic ability that is needed in order to successfully accomplish the real-world criterion task. Finally, pragmatics always plays a key role in each assessment, but it is not the sole focus. Instead, it is part of a broader construct of communicative ability or set of skills necessary to accomplish a given task. Given that task performances always "reveal multiple aspects of language ability and/or development within a single instance" (Norris, 2016, p. 241), the rubric determines which elements of language should be evaluated for what interpretations and purposes. Thus, task-based language assessment allows for a flexible and polyvalent application of the assessment relative to the intended use. If the use is an evaluation of a test-taker's pragmatic ability, then relevant criteria can be foregrounded in the scoring process.

Concluding remarks

Taking the previously discussed considerations into account, several characteristics of TBLA are congruent with key dimensions of L2 pragmatics, making tasks a useful vehicle for assessing pragmatic abilities. Such a Task-Based Pragmatics Assessment (TBPrA) approach facilitates the high degree of contextualization needed to simulate real-life contexts in which test-takers are required to use language in order to accomplish a specific (communicative) goal. Moreover, as a type of integrated performance assessment, TBPrA allows for the assessment of pragmatics in interaction. Hence, TBPrA provides a means of assessing pragmatics as performance, focusing on discursive pragmatics or pragmatics-in-interaction, rather than pragmatics knowledge – a need that has been highlighted repeatedly in the field of L2 pragmatics assessment (e.g., Kasper, 2006; Roever, 2011; Taguchi & Roever, 2017; Timpe-Laughlin et al., 2015).

While a task-based approach to L2 pragmatics assessment has a number of advantages, there is still a pressing need for further foundational research that investigates the relationships among task design features (e.g., task complexity, task difficulty), pragmatic targets, and elicited pragmatic moves. Kim and Taguchi (2015), for example, argued for "a process and product oriented task-based research perspective" (p. 660), highlighting the importance of investigating the quality of task performance by examining student output and potential relationships associated with task design

features. A focus should also be placed on pragmatic phenomena in interaction; that is, the focus of the analyses should go beyond individual speech acts (Kim & Taguchi, 2015). Qualitative analyses, for example, from a CA perspective could provide a more comprehensive picture, by investigating speech acts within discourse sequences and essentially informing task design. By focusing on potential effects tasks may have on the elicitation of pragmatic moves, such a research agenda can provide a more detailed understanding of the effects of task characteristics and their effects on pragmatic performance. A better understanding would in turn allow for more informed decisions about the selection, development, and sequencing of tasks in a given assessment context. Additionally, a more systematic understanding of task effects could also support claims regarding generalizability of performance across different tasks and consistency in task difficulty. Moreover, a more detailed understanding would allow for test specifications with a stronger empirical basis, which would facilitate production of items that more consistently elicit targeted aspects of pragmatic performance – a key issue, especially for large-scale proficiency tests.

Currently, most large-scale, standardized assessments such as the Test of English as a Foreign Language iBT (TOEFL® iBT) and International English Language Testing System (IELTS™) follow an interactionalist view of language ability which is arguably task-oriented insofar as tasks serve as the vehicle to elicit certain knowledge, skills, and abilities (i.e., a weak form of TBLA). Although large-scale proficiency tests like IELTS™ and TOEFL® iBT have been highlighted as "good example[s] of the extent to which simulations of authentic communication tasks have come to be common practice on international tests of English for academic purposes proficiency" (Norris, 2016, p. 235), pragmatics plays a still rather limited role in these assessments. The only instances of a focus on pragmatics in both, the TOEFL® iBT and IELTS, have to do with understanding stance and opinion in the listening sections. While TOEFL® iBT tasks have been carefully selected, sampled, and sequenced in the process of a systematic domain analysis to be representative of typical academic tasks, these tasks would benefit from some features that could be added to the item specifications such as specifying an audience for a writing or speaking task (for a first endeavor into that direction see Cho & Choi, 2017).

To conclude, the TBLA paradigm constitutes a fitting environment for assessing the complex construct of form-function-context mapping. Considerable advances have been made in recent years to the extent that a number of assessments already feature TBPrA, that is, they explicitly rate the task performance for pragmatic abilities needed for a particular task. However, some fundamental aspects, primarily regarding the interaction between task characteristics, test-takers, and task performance – that are crucial for reliable, valid, and generalizable inferences – still need to be explored further in order to advance TBPrA across different, operational, and potentially high-stakes assessment contexts.

References

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 453–476. doi:10.1191/0265532202lt240oa

Brindley, G. (1994). Task-centred assessment in language learning programs: The promise and the challenge. In N. Bird, P. Falvey, A. Tsui, D. Allison, & A. McNeill (Eds.), *Language and learning* (pp. 73–94). Hong Kong: Institute of Language in Education.

Brindley, G. (2009). Task-centered language assessment in language learning. The promise and the challenge. In K. Van den Branden, M. Bygate, & J. M. Norris (Eds.), *Task-based language teaching. A reader* (pp. 435–454). Amsterdamphia: John Benjamins.

Brown, J. D., Hudson, T., Norris, J. M., & Bonk, W. (2002). *An investigation of second language task-based performance assessments* (Technical Report #24). Honolulu, HI: University of Hawai'i, Second Language Teaching & Curriculum Center.

Byrnes, H. (2002). The role of task and task-based assessment in a content-oriented collegiate foreign language curriculum. *Language Testing*, 19(4), 419–437.
doi:10.1191/0265532202lt2380a

Byrnes, H., Maxim, H., & Norris, J. M. (2010). Realizing advanced FL writing development in collegiate education: Curricular design, pedagogy, assessment. *Modern Language Journal, Monograph*. Cambridge: Wiley-Blackwell.

Cho, Y., & Choi, I. (2017). *Writing from sources: Does audience matter?* Manuscript submitted for publication.

Clyne, M. (1994). *Inter-cultural communication at work: Cultural values in discourse*. Cambridge: Cambridge University Press.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Crystal, D. (Ed.). (1997). *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.

Curriculum Development Institute (2005). *Task-based assessment for English language learning at secondary level*. Retrieved from <http://cd1.edb.hkedcity.net/cd/eng/TBA_Eng_Sec/index.html>.

Elder, C. (2001). Assessing the language proficiency of teachers: Are there any border controls? *Language Testing*, 18(2), 149–170. doi:10.1177/026553220101800203

Elder, C. (2016). Exploring the limits of authenticity in LSP testing: The case of a specific-purpose language test for health professionals. *Language Testing*, 33(2), 147–152.

Elder, C., McNamara, T., Woodward-Kron, R., Manias, E., McColl, G., Webb, G., & Pill, J. (2013). *Towards improved healthcare communication. Development and validation of the language proficiency standards for non-native English speaking health professionals*. (Final report for the occupational English test centre). Melbourne, AUS: The University of Melbourne. Retrieved from <https://www.occupationalenglishtest.org/resources/uploads/2015/07/Elder-C-et-al-2013-Towards-Improved-Healthcare.pdf?x59645>.

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

Gilabert, R., & Barón, J. (2013). The impact of increasing task complexity on L2 pragmatic moves. In A. Mackey & K. McDonough (Eds.), *Second language interaction in diverse educational contexts* (pp. 45–69). Amsterdam: John Benjamins. doi:10.1075/lllt.34.06ch3

Grabowski, K. (2009). *Investigating the construct validity of a test designed to measure grammatical and pragmatic knowledge in the context of speaking*. Unpublished PhD dissertation, Teachers College, Columbia University, New York, NY.

Grabowski, K. (2013). Investigating the construct validity of a role-play test designed to measure grammatical and pragmatic knowledge at multiple proficiency levels. In S. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 149–171). Houndmills: Palgrave MacMillan. doi:10.1057/9781137003522.0013

Holmes, J. (2000). Doing collegiality and keeping control at work: Small talk in government departments. In J. Coupland (Ed.), *Small talk* (pp. 32–61). London: Continuum.

Hudson, T., Detmer, E., & Brown, J. D. (1995). *Developing prototypic measures f cross-cultural pragmatics*. Honolulu, HI: University of Hawaiʻi, Second Language Teaching and Curriculum Center.

Kasper, G. (2006). Beyond repair: Conversation analysis as an approach to SLA. *AILA Review*, 19, 83–99. doi:10.1075/aila.19.07kas

Kasper, G., & Ross, S. (2013). Assessing second language pragmatics: An overview and introductions. In S. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp.1–40). Houndmills: Palgrave Macmillan. doi:10.1057/9781137003522.0007

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom contexts: The role of task complexity. *Modern Language Journal*, 99(4), 656–677. doi:10.1111/modl.12273

Lockwood, J. E. (2015). Language for specific purpose (LSP) performance assessment in Asian call centres: Strong and weak definitions. *Language Testing in Asia*, 5(3). doi:10.1186/s40468-014-0009-6.

McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.

McNamara, T. F., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell.

Messick, S. J. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* 23, 13–23. doi:10.3102/0013189X023002013

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, 19(4), 477–496. doi:10.1191/0265532202lt241oa

National Standards in Foreign Language Education Project. (1996). *Standards for foreign language learning: Preparing for the 21st century*. Yonkers, NY: Author.

Norris, J. M. (2000). Purposeful language assessment: Selecting the right alternative test. *English Teaching Forum*, 38(1), 18–23.

Norris, J. M. (2001). Identifying rating criteria for task-based EAP assessment. In T. Hudson & J. D. Brown, (Eds.), *A focus on language test development: Expanding the language proficiency construct across a variety of tests* (Technical Report #21, pp. 163–204). Honolulu, HI: University of Hawaiʻi, Second Language Teaching and Curriculum Center.

Norris, J. M. (2009). Task-based teaching and testing. In M. Long & C. Doughty, (Eds.), *The handbook of language teaching* (pp. 578–594). Malden, MA: Blackwell-Wiley. doi:10.1002/9781444315783.ch30

Norris, J. M. (2016). Current uses for task-based language assessment. *Annual Review of Applied Linguistics*, 36, 230–244.

Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments* (Technical Report #18). Honolulu, HI: University of Hawaiʻi, Second Language Teaching and Curriculum Center.

Okada, Y., & Greer, T. (2013). Pursuing a relevant response in oral proficiency interview role plays. In S. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 288–310). Houndmills: Palgrave Macmillan. doi:10.1057/9781137003522.0019

Quellmalz, E. (1991). Developing criteria for performance assessments: The missing link. *Applied Measurement in Education*, 4, 319–331. doi:10.1207/s15324818ame0404_5

Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing*, 23(2), 229–256. doi: 10.1191/0265532206lt3290a

Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463–481. doi: 10.1177/0265532210394633

Roever, C., Fraser, C., & Elder, C. (2014). *Testing ESL sociopragmatics: Development and validation of a web-based test battery*. Frankfurt: Peter Lang. doi: 10.3726/978-3-653-04598-7

Roever, C., & McNamara, T. F. (2006). Preview article. Language testing: The social dimension. *International Journal of Applied Linguistics*, 16(2), 242–258. doi: 10.1111/j.1473-4192.2006.00117.x

Taguchi, N. (2007). Task difficulty in oral speech act production. *Applied Linguistics*, 28, 113–135. doi: 10.1093/applin/amlo51

Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford: Oxford University Press.

Timpe, V. (2013a). "The difficulty with difficulty:" The issue of determining task difficulty in TBLA. *Journal of Linguistics and Language Teaching*, 4(1), 13–28.

Timpe, V. (2013b). *Assessing intercultural communicative competence. The dependence of receptive sociopragmatic competence and discourse competence on learning opportunities and input*. Frankfurt: Peter Lang.

Timpe-Laughlin, V., Wain, J., & Schmidgall, J. (2015). *Defining and operationalizing the construct of pragmatic competence: Review and recommendations*. (Research Report No. RR-15-06). Princeton, NJ: Educational Testing Service.

Timpe-Laughlin, V., & Park, I. (2017). *"Are you into beer pong?"– Exploring question-answer sequences in an L2 oral performance assessment*. Unpublished manuscript.

Washburn, G. (2001). Using situation comedies for pragmatic language teaching and learning. *TESOL Journal*, 10(4), 21–26.

Youn, S. J. (2008). *Developing a task-based assessment of EAP pragmatics*. Retrieved from <https://scholarspace.manoa.hawaii.edu/bitstream/10125/20180/1/Youn2008.pdf>.

Youn, S. J. (2013). *Validating task-based assessment of L2 pragmatics in interaction using mixed methods*. Unpublished doctoral dissertation, Graduate Division of the University of Hawaiʻi at Mānoa, Honolulu, HI.

Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, 32, 199–225. doi: 10.1177/0265532214557113

Bio notes

Eva Alcón Soler is Full Professor of English Language and Linguistics at the University Jaume I (Castelló). Her research interests are the acquisition of L2 pragmatics and the role of interaction in L2 learning. Her publications include *Learning how to Request in an Instructed Language Learning Context* (Peter Lang, 2008), *Intercultural Language Use and Language Learning* (Springer, 2007), and *Investigating Pragmatics in Foreign language Learning, Teaching and Testing* (Multilingual Matters 2008). She has also co-edited two special issues on the topic of interaction and language learning in a classroom context (*International Journal of Educational* Research, 2002; *International Review of Applied Linguistics*, 2009), and two special issues on the topic of pragmatic learning (*System,* 2005; 2015).

Júlia Barón is currently a full-time lecturer at the Institute for Multilingualism at the International University of Catalonia and a lecturer at the University of Barcelona. She is a member of the GRAL (Language Acquisition Research Group), which investigates different aspects of second language acquisition. Her research focuses on pragmatics and more specifically on interlanguage pragmatics. The objectives of her studies are to examine how learners of a foreign language acquire and develop the pragmatics of the target language. Other fields of study related to pragmatics that she is interested in are pragmatic transfer, the relationship between task complexity and interaction, and the pragmatics of interaction.

Monika Ekiert is Associate Professor in the Department of Education and Language Acquisition at LaGuardia CC, City University of New York, where she teaches undergraduate courses in bilingualism, linguistics, and academic writing. Her research interests lie at the intersection of second language acquisition and pedagogy. Currently, she is involved in research projects focusing on task based language teaching (TBLT), crosslinguistic influence (CLI) in second language acquisition, and written corrective feedback.

Roger Gilabert is an associate professor and researcher at the University of Barcelona, and he has conducted research and published extensively in areas of second and foreign language production and acquisition, task design, task complexity and task sequencing, the effects of oral and written modes on CAF, as well as individual differences in L2 production and acquisition. Within the Language Acquisition Research Group (GRAL) Roger Gilabert is a co-PI in a Spanish Ministry project on subtitling/captioning and language learning. He is also a co-PI in a Horizon 2020 project on reading and gamification led by UCL (UK).

Marta González-Lloret is a Professor at the University of Hawai'i, Manoa. Her research interests lay at the intersections of technology and TBLT (Task-based Language Teaching) and technology and L2 pragmatics. Her books include *A practical guide to integrating technology into task-based language teaching* (2016*) and Technology-mediated TBLT* (2014). She served as the Pragmatics Volume Editor in the *Wiley Encyclopedia of Applied Linguistics* (2012, Carol Chapelle General Editor) and she is currently Editor of the *Pragmatics & Language Learning* book series (NFLRC) as well as co-editor of *System* Journal (Elsevier).

Minkyung Kim is a PhD candidate in the Department of Applied Linguistics at Georgia State University. Her primary research interests are second language reading and writing, language assessment, lexical development, and teaching Korean as a foreign language. Her work has been published in *Modern Language Journal*.

YouJin Kim is Associate Professor in the Department of Applied Linguistics and ESL at Georgia State University. She specializes in second language acquisition (SLA), second language pedagogy, and task-based language teaching and assessment. Within SLA, she is particularly interested in examining how interactional features impact second language development involving two target languages: English and Korean. She is also interested in the benefits of study abroad programs in learning additional languages among heritage and non-heritage language learners. Her research articles can be found in journals such as *Studies in Second Language Acquisition*, *Language Learning*, *Modern Language Journal*, *Applied Linguistics*, and *TESOL Quarterly* among others.

Folkert Kuiken is professor of Dutch as a Second Language and Multilingualism at the University of Amsterdam. He is also Academic Director of the Institute for Dutch Language Education at that same university. His research interests include the effect of task complexity and interaction on SLA, Focus on Form, and the relationship between linguistic complexity and functional adequacy. His publications have appeared in various journals (a.o. Applied Linguistics, Bilingualism: Language and Cognition, Journal of Second Language Writing, Language Learning, Language Testing). He (co)authored and (co)edited various books, among which *The lexicon-syntax interface in second language acquisition* (Van Hout, Hulk, Kuiken & Towell, 2003) and *Dimensions of L2 performance and proficiency* (Housen, Kuiken & Vedder, 2012).

Sofia Lampropoulou is Senior Lecturer in English Language at the University of Liverpool. Her research interests are in sociolinguistics and discourse analysis, and particularly style, speech representation and identity construction in, among others, storytelling, interviews and media discourse. She has published several articles as well as the monograph *Direct Speech, Self-Presentation and Communities of Practice* (Bloomsbury, 2013).

Mayya Levkina holds a doctorate in Applied Linguistics from University of Barcelona. She is currently an assistant professor at the Faculty of Philology, University of Barcelona, Spain and also at the International University of Catalonia, Spain. Her research interests lie in the areas of task-based language teaching (TBLT) and second language acquisition (SLA). She is particularly interested in task complexity, task sequencing; individual differences, working memory capacity and attention control, L2 pronunciation and L2 pragmatics. The research project she is currently involved in is related to intentional and incidental learning of English through authentic input, the case of subtitles.

Hakyoon Lee (Ph.D., University of Hawai'i at Mānoa) is a lecturer in the Department of World Languages and Cultures at Georgia State University. She has been teaching at Georgia State University since the fall semester of 2013. Her research interests are at the intersection of sociolinguistics, bilingualism and multilingualism, and immigrant education. With an emphasis on discourse analysis, her research focuses on understanding language learners' social interactions and their use of linguistic and cultural resources in various social contexts. She has published her work in *Journal of Language, Education, and Identity* (2011) and *Applied Linguistics* (2015).

Maria Pia Gomez-Laich is a Post Doctoral Research Associate at Carnegie Mellon University in Qatar. She holds a Ph.D. in Second Language Acquisition from Carnegie Mellon University. Her research focuses on the predictions of the Cognition Hypothesis of task-based language learning. In particular, she has investigated whether task complexity affects ESL students' interaction while performing collaborative writing tasks and their subsequent ability to use genre conventions and linguistic resources for writing academic texts in English. At CMU-Q, she is working on a research project aimed at enhancing the communication skills of students in the Information Systems Program.

Lourdes Ortega is a Professor at Georgetown University. Her main area of research is in second language acquisition, particularly usage-based, bilingual, and educational dimensions in adult classroom settings. Her books include *Understanding Second Language Acquisition* (2009, Routledge; translated into Mandarin in 2016) and *Technology-mediated TBLT* (2014). Her most recent publications are *Usage-inspired L2 instruction*, 2017 (co-edited with Andrea Tyler and colleagues), which class for innovating the approach known as focus on form through the application of usage-based insights to grammar instruction, and *Complexity Theory and language development*, 2018 (co-edited with ZhaoHong Han), a Festschrift honoring Diane Larsen-Freeman's work.

Caroline Payant is an assistant professor in the Education Department at the Université du Québec à Montréal. Her research examines task type, task modality, and task

repetition effects on interaction and L2/L3 development. Her work can be found in the International Journal of Bilingual Education and Bilingualism, Foreign Language Annals, Studies in Second Language Acquisition.

Derek Reagan is a Ph.D. student in Applied Linguistics at Georgetown University. Previously, he graduated with a M.A. in Teaching English as a Second Language from the University of Idaho, and he completed a Fulbright English Teaching Assistantship in Natal, Brazil. His research interests include task-based language teaching, task modality, pragmatics, and second language pedagogy, pragmatics.

Andrea Révész is Associate Professor in applied linguistics at the UCL Institute of Education, University College London. Her main research interests lie in the areas of second language acquisition and instruction, in particular, the roles of tasks, input, interaction and individual differences in instructed second language development. Her current work also examines the cognitive processes underlying second language writing, speaking, and listening performance. She serves as associate editor of the journal Studies in *Second Language Acquisition* and is Vice-President of the *International Association for Task-based Language Teaching* (TBLT).

Naoko Taguchi is Professor in the Modern Languages Department at Carnegie Mellon University, where she teaches courses in second language acquisition and Japanese language and culture. Her research interests include pragmatics, intercultural competence, technology assisted learning, and English-medium education. She has authored/co-authored several books, including *Second Language Pragmatics: From Theory to Research* (2018, Routledge), *Second Language Pragmatics* (2017, Oxford University Press), *Developing Interactional Competence in a Japanese Study Abroad Context* (2015, Multilingual Matters), and Context, *Individual Differences, and Pragmatic Competence* (2012, Multilingual Matters). She is currently editing the *Routledge Handbook of SLA and Pragmatics*. She is the co-editor of a new journal, *Applied Pragmatics* (John Benjamins).

Veronika Timpe-Laughlin is a research scientist in the field of English Language Learning and Assessment at Educational Testing Service (ETS). Her research interests include L2 pragmatics, task-based language teaching and assessment, intercultural communication, and technology in L2 instruction and assessment. She has authored two books and has published in leading journals such as *Language Assessment Quarterly*. Before joining ETS, Veronika worked and taught in the English Department at TU Dortmund University, Germany.

Eivind Torgersen is Professor of English in the Department of Teacher Education at the Norwegian University of Science and Technology. Most of his work focuses

on language variation and change in London and Multicultural London English, in particular modelling of phonological change and the use of spoken corpora in sociolinguistic research. Other research interests are in experimental phonetics, second language acquisition and multilingualism in the English language classroom in Norway.

Soo Jung Youn (PhD, University of Hawaiʻi) is Assistant Professor of Applied Linguistics, English Department, Northern Arizona University, USA. Her academic interests are L2 pragmatic assessment, task-based language teaching, quantitative research methods, and conversation analysis. Her studies have recently been published in *Language Testing*, *System*, *Applied Linguistics Review*, *Papers in Language Testing and Assessment*, and book chapters.

Ineke Vedder is senior lecturer of Italian linguistics at the University of Amsterdam, and research member of the Amsterdam Center for Language and Communication (ACLC). Her research interests include instructed SLA, academic writing in L2/L1, L2 pragmatics, task-based language assessment. Her recent publications have appeared in various journals (e.g. Journal of Second Language Writing, Language Testing, International Review of Applied Linguistics in Language Teaching, EuroAmerican Journal of Applied Linguistics and Languages). She co-edited various books, among which *Dimensions of L2 performance and proficiency* (Housen, Kuiken & Vedder, 2012) and *Pragmatica e interculturalità in italiano lingua seconda* (Santoro & Vedder, 2016).

Index

This volume is the first book-length attempt to bring together the fields of task-based language teaching (TBLT) and second language pragmatics by exploring how the teaching and assessment of pragmatics can be integrated into TBLT. The TBLT-pragmatics connection is illustrated in a variety of constructs (e.g., speech acts, honorifics, genres, interactional features), methods (e.g., quantitative, quasi-experimental, conversation analysis), and topics (e.g., instructed SLA, heritage language learning, technology-enhanced teaching, assessment, and discursive pragmatics). Chapters in this volume collectively demonstrate how the two fields can together advance the current practice of teaching language for socially-situated, real-world communicative needs.

John Benjamins Publishing Company