RECONSIDERING THE DEVELOPMENT OF THE DISCOURSE COMPLETION TEST IN INTERLANGUAGE PRAGMATICS[1]

Afef Labben

Abstract

A survey of the field of Interlanguage Pragmatics (ILP) shows that the Discourse Completion Test (DCT)[2], also referred to as a 'discourse completion task' or a 'production questionnaire', has been the most frequently used instrument to evaluate second/foreign language learners' ability to perform speech acts in a target language, despite the harsh criticism leveled against its low construct validity and its failure to represent the features of authentic discourse. Interestingly, focusing on the statement of objectives of a number of ILP studies using DCTs, one can notice that such studies rarely refer to the DCT as a language test. In addition, an overview of the DCT design process as described in several ILP studies shows that ever since its adaptation for the study of pragmatic ability (Blum-Kulka, 1982), there has been a tendency to use or adapt one of the existing DCT versions used in previous studies based on the argument of comparability of results. While a number of ILP researchers tried to improve the design of the DCT by the inclusion of rejoinders or by enhancing the prompt material (e.g. Billmyer and Varghese, 2000), few attempts have been made to reconsider the DCT development process. McNamara and Roever (2006: 253) urge for the need for "more research on testing of sociopragmatic knowledge and design of discourse completion tests for testing purposes."

The present paper starts with an overview of the literature about DCTs with special reference to the cognitive validity of the instrument and to previous studies dealing with DCT structure and content. Then, with reference to research in the fields of language testing and psychometrics, it shows that, whether used for research or instructional purposes, the DCT shares several qualities with language tests. As such, it is argued that the DCT should be treated as a language test and not as a questionnaire and should, thus, undergo a rigorous developmental process. Based on recent models of language test construction, the paper concludes with an overview of the stages of DCT development.

Keywords: Discourse Completion Test; Interlanguage Pragmatics; Test development; Cognitive validity; Construct validity.

1.The written DCT

In the context of ILP, written DCTs or production questionnaires were first adapted by Blum-Kulka (1982) to study speech act realization patterns. DCT items typically consist

---

[1] An earlier version of this paper was presented at the Language Testing Forum held at the University of Southampton, UK, 21-23 November 2014.

[2] The term DCT in the present study is used to refer to the traditional written discourse completion test used in interlanguage and cross-cultural pragmatics.

of descriptions of speech act situations followed by incomplete discourse sequences that the respondents are requested to complete. DCTs can be open-ended or structured by a prompt preceding the incomplete discourse sequence and/or a rejoinder in the form of an interlocutor's reply (Cohen 1996). Such a rejoinder can be positive or 'preferred', i.e. providing a positive response to the speech act performed by the speaker, or negative or 'dispreferred', i.e providing a negative response to the speech act performed by the speaker (Johnston et al. 1998). DCT content can also vary in terms of the amount of contextual information included in the prompt. It can thus be content poor or content enhanced (see Billmyer and Varghese 2000). An overview of DCTs used in previous ILP and cross-cultural pragmatics (CCP) studies shows the existence of different versions of DCT items. Figure 1, below, provides examples of different versions of DCT items used and/or reported in the literature. Each example is introduced in terms of type and speech act investigated. It is important to note, though, that in theory and in addition to these six examples, other versions can be devised and potentially used in the forthcoming years.

Written DCTs have been extensively used in CCP and ILP studies. The popularity of DCTs is due to the fact that they enable the researcher to collect and codify large amounts of data in relatively short time periods (Johnston et al. 1998). DCTs also allow the manipulation of the variables underlying speech act situations, which facilitates comparability of data across languages (Johnston et al. 1998; Kasper and Dahl 1991). However, DCTs have been criticized for the lack of authenticity of the data they yield. Being highly controlled, they have low construct validity as to the extent to which they represent the oral features of authentic discourse (Johnston et al. 1998). DCTs have also been criticized for under-representing the construct they are reported to measure (Grabowski 2008). Studies which have attempted to strengthen the design of DCTs in order to improve their construct validity will be described in section 2 below.

Another type of validity that has been recently evoked in the testing literature might equally need to be investigated in DCTs however. Stressing the importance of what is referred to as 'cognitive validity' in reading tests, Bax (2013: 3) urges that:

> (W)hen we prepare reading tests, it is important to ensure that our tests are valid, and part of a test's validity involves ensuring that the mental processes which test-takers use as they respond are similar to and representative of the mental processes they would use in the target situation in real life – what is known as cognitive validity (Glaser 1991; Field 2012).

In order to investigate such a type of validity, it is perhaps important to explore the cognitive demands that a DCT places on respondents, hence determine the type of knowledge retrieved when responding to DCTs. It might also be crucial to investigate the question of whether DCTs make use of the same type of pragmatic knowledge accessed by language speakers in real life contexts.

Example 1: unstructured, content poor (complaint)
       It is not the first time that rock music is heard from your neighbor's apartment quite late at night.
       You pick up the phone and say:…..
       (Olshtain and Weinbach, 1993: 121)

Example 2: unstructured, content enhanced (request)
       It is 10.30 pm. on a Wednesday night and you have a paper due the next day. You are trying to
       finish the paper and you can't concentrate because you hear loud music coming from another
       student's room down the hall. You decide to ask her to turn the music down. The music has been
       on at this volume for half an hour. You have occasionally seen the student, Lucy, Row, in the same
       dorm during the past six months. She is a student like you, but you have never spoken to her. You
       have heard other people in the dorm complain about the volume of her music on several occasions
       although you never have because you study in the library. However, today the library closed early.
       You are only halfway through and you know that the professor for this class is very strict and does
       not give extensions. What would you say?
       (Billmyer and Varghese, 2000: 523)

Example 3: structured, positive rejoinder, content poor (request)
       A professor wants a student to present a paper in class a week earlier than scheduled. What would
       the professor say?...
       Student: sure. I should be finished with it by then.
       (Rose, 1992: 61)

Example 4: structured, negative rejoinder, content poor (request)
       Next weekend you're giving a party, and one of your friends is a great cook. You would like him
       to do the food for the party.
       You:…………………………………………………….…………………………………………
       Your friend: I'm sorry, but I'm going away next weekend.
       (Johnston et al., 1998: 175)

Example 5: structured, prompt, content poor (apology)
       You promised to return a textbook to your classmate within a day or two, after photocopying a
       chapter. You kept it for almost two weeks.
       Classmate: I'm really upset about the book because I needed it to prepare for last week's class.
       You…………………………………………………………………………………………………
       (Cohen, 1996: 390)

Example 6: structured, prompt, multiple rejoinder, content poor (apology)
       You arranged to meet a friend in order to study together for an exam. You arrive half an hour late
       for the meeting.
       Friend (annoyed): I've been waiting at least half an hour for you!
       You………………………………………………………………………………………………….....
       Friend: well, I was standing here waiting. I could have been doing something else.
       You…………………………………………………………………………………………………
       Friend: still, it's pretty annoying. Try to come on time next time.
       (Cohen 1996: 390)

Figure 1. Examples of DCT items

1.1. *DCT cognitive demands*

Golato (2003) maintains that DCTs generate metapragmatic data as they require respondents to report what they believe they would say in imaginary contexts. Hence, they can be criticized for testing the respondents' pragmatic knowledge as a permanent type of knowledge rather than as an on-line phenomenon retrieved in context. In fact, metapragmatic knowledge might be sensitive to the context of the task and the subjectivity of the respondents (Kasper and Dahl 1991). Thus, from a psycholinguistic perspective, it seems appropriate to ask whether this type of knowledge lies above the level of consciousness, i.e. whether the respondents to DCT are able to report what they would say in similar real life situations in an artificial situation through instantly accessing their pragmatic competence. Following this line of thought, whether respondents answer DCTs without making use of the same type of knowledge called upon in natural real-world situations is of crucial importance. Golato (2003: 110) argues that "a DCT is not an *on-line* task in which a person uses language spontaneously and without consciously focusing on linguistic output, but is instead an *off-line* task in which a person has time for introspection" [italics original]. In the same vein, Cohen and Olshtain (1994: 148) claim that because the DCT is "a projective measure of speaking," the cognitive process that language users follow to produce speech acts when responding to this method may not "truly reflect" natural speech act production. Similarly, based on empirical evidence resulting from comparing DCT data with data collected by means of a conversation analysis-informed test, Walters (2013: 193) concludes that "inferring online L2 pragmatic competence from a DCT is of doubtful validity."

In fact, learners may very likely rely on different processing abilities to perform task imposed speech acts. The cognitive mechanisms needed to respond to a DCT may differ from those employed in real life contexts. Bialystok (1993: 47) asserts that "(D)ifferent uses of language involve different processing abilities of language learners…" Referring to how language learners' use of language is affected by a specific language use situation, she further argues that language proficiency can be viewed in relation to the processing abilities of the learner vis-à-vis the task demands that a specific language use situation requires: "(W)here the two are congruent, learners will perform well; where the task demands are excessive relative to the learner's ability, learners will struggle." The cognitive demands that a DCT imposes on language speakers basically relate to the ability to understand and report through writing a pragmalinguistically and sociopragmatically 'appropriate' speech act in an artificial situation. Let us compare the abilities which might be required to answer a DCT to those probably used to perform a speech act in a real life context (Table 1).

On the face of it, it can be presumed from the abilities mentioned in Table 1 that responding to a DCT differs from performing a speech act in a real life context mainly in relation to the reading and writing skills required to respond to such a method. At a preliminary stage, then, we can hypothesize that to be able to understand the speech act situation described, the respondent should have the necessary reading skills involving understanding the cultural inferences and the vocabulary items used in the situation description, etc. At a later stage, the respondent should be able to report his/her spoken

response in the form of a written mode. Responding to DCTs, thus, seems to be more demanding[3], cognitively speaking, than producing speech acts in real-life contexts.

Table 1. Abilities required to produce a speech act[4]

Abilities required to respond to a DCT	Abilities used to perform a speech act in a real life context.
• Read and understand the situation description in terms of grammar, vocabulary items and syntactic structure used to describe the speech act situation. • Imagine the situation as a real life situation. • Grasp the contextual factors likely to affect the response: age, gender, power relationship, social distance, language, cultural background of the interlocutor, type of speech act needed, level of formality, level of politeness, etc. • Understand the cultural inferences involved in the situation. • Issue a sociopragmatic evaluation of the situation taking into account features of the context. • Choose from a variety of possibilities to perform the speech act in question the appropriate sociopragmatic strategies.[5] • Map the strategies into the target language by choosing the appropriate pragmalinguistic form to realize the speech act. • Accurately report through writing what would be said in that specific situation.	• Grasp the contextual factors likely to affect the response: age, gender, power relationship, social distance, language, cultural background of the interlocutor, type of speech act needed, level of formality, level of politeness, etc. • Understand the cultural inferences involved in the situation. • Issue a sociopragmatic evaluation of the situation taking into account features of the context. • Choose from a variety of possibilities to perform the speech act in question the appropriate sociopragmatic strategies. • Map the strategies into the target language by choosing the appropriate pragmalinguistic form to realize the speech act. • Perform the speech act in the appropriate mode (mainly oral).

---

[3] It is important to note that the content enriched DCT versions may be more cognitively demanding than the content poor ones. I am grateful to one of the reviewers who commented that "the more elaborate the content enhanced version, the more the DCT will be confounded with reading comprehension ability."

[4] The use of the abilities listed in this table may not necessarily occur in the order in which they are listed. The abilities may be used simultaneously or may follow another order. It is beyond the scope of this paper to deal with this issue. I also do not claim that the list is exhaustive. It is just an attempt to outline the main abilities I believe are used in speech act production based on models of language use including assessment, planning, and execution components (e.g. Bachman 1990) and the established distinction in pragmatics between pragmalinguistics and sociopragmatics (Leech 1983; Thomas 1983).

[5] One of the reviewers commented that DCT tasks by their very nature impose sociopragmatic choices and that what is left for the respondent may just involve pragmalinguistics. I agree with this observation but I also think that the amount of sociopragamtic information varies across DCT versions (see figure 1). Thus, while the content enriched DCT imposes sociopragmatic choices by detailing contextual features, the content poor DCT may leave it for the respondent to imagine such features.

The issue of the cognitive demands that DCTs make on respondents is surely far more complex than what has been mentioned and needs further research that relies on appropriate psycholinguistic methods of investigation to better account for the cognitive demands of DCTs, which is beyond the scope of this paper. To examine whether respondents on DCTs make use of the same mental processes they follow in responding to similar authentic situations, it is also important to explore the effects of the time of planning required in responding to DCTs and the type of knowledge targeted by such instruments.

1.1.1. Time of planning

Investigating the effects of planning and task type on second language performance, Foster and Skehan (1996) report considerable effects of planning on fluency and complexity. It is important to note, however, that although their study focused on only narrative, decision-making, and personal information exchange tasks, the researchers showed that planning interacted differentially with task type. Although few studies focused on whether the amount of time and planning respondents have to respond to DCTs affects their performance on these tasks, it seems safe to hypothesize, in the light of findings revealed by studies such as Foster and Skehan's, that the amount of time and planning devoted to responding to a DCT may affect respondents' performance. In fact, in a real life situation, some speech acts are performed under time pressure, which means that the speaker might realize the speech act in question 'unconsciously' using implicit knowledge (see section 1.1.2 below for discussion of type of knowledge accessed by DCTs). Still, some speech acts like apologies and requests offer the speaker some time to prepare what to say. In many apology situations, the offender offers apologies for offenses committed some time before meeting the offendee. Similarly in various request situations the requester prepares what to say in advance - that is, before meeting the requestee. Within the same line of thought, Schmidt (1993: 23) observes that:

> (P)ragmatic and discoursal language is not always used automatically and unreflectively. Conversations vary a great deal in terms of spontaneity and planning (Ochs 1979). Some people preplan telephone conversations and writing involves a great deal of conscious deliberation and choices in discourse organization.

Consequently, speech act type may be another variable to consider when addressing the question of cognitive demands DCTs make and their resemblance to/difference from those employed in authentic speech act production. Another very important psycholinguistic factor to consider when addressing this question is the issue of consciousness involved in performing the speech act. How conscious is the evaluation of the speech act situation in real life contexts? Do language speakers attend to the pragmalinguistic and/or sociopragmatic strategies they use? Is pragmatic knowledge required to perform speech acts explicit? If yes, how can it be measured effectively? Such questions also relate to the cognitive side of responding to DCTs, specifically to the type of knowledge called upon to respond to DCTs.

1.1.2. Type of knowledge

Whether pragmatic knowledge accessed when responding to DCTs is implicit or explicit is also an issue to consider. A number of researchers have dealt with the issue of explicit knowledge. Ellis (2004: 229) defines explicit knowledge as "knowledge of language about which users are consciously aware." According to Ellis, then, explicit knowledge is "knowledge *about* language and *about* the uses to which language can be put." As far as pragmatic features are concerned, Ellis (2004: 243-244) points out that "(J)ust as learners can have explicit knowledge of linguistic aspects of the L2, so too they can develop a conscious understanding of pragmatic aspects." Ellis takes the example of parents who explicitly teach their children some standard formulas associated with expressive speech acts, as well as second language learners who develop such knowledge through instruction. He further argues that although researchers do not necessarily set out to examine explicit knowledge with regard to pragmatic features, the widely used DCT is "arguably more likely to tap explicit than implicit knowledge." Still for Ellis, explicit knowledge is "typically accessed through controlled processing when L2 learners experience some kind of linguistic difficulty in the use of the L2."

In relation to pragmatic knowledge, Wolfson (1989: 37) maintains that "rules of speaking and, more generally, norms of interaction are…largely unconscious." Building on some evidence, Schmidt (1993: 23) counters Wolfson, postulating that "pragmatic knowledge…seems to be partly conscious and partly accessible to consciousness, although it cannot be the case that all pragmatic knowledge is accessible to consciousness." Similarly, in their distinction between declarative and procedural knowledge, Faerch and Kasper (1984: 215) posited that context and sociocultural knowledge are part of what they call declarative knowledge or "knowledge that" which is "non-automatized and conscious." As a result, pragmatic knowledge might be accessible to consciousness.

Investigating the issue of type of knowledge accessed when responding to DCTs, Bardovi- Harlig (2013: 74) asserts that written DCTs are likely to tap explicit knowledge. She explains that "(M)ost DCTs are given as untimed tasks, further increasing the likelihood that a respondent might draw on explicit knowledge." But she equally states that "(T)ime pressure does not guarantee use of implicit knowledge and even lack of time pressure does not guarantee use of explicit knowledge." Roever (2011: 470) differentiates between real-life language use and elicited knowledge stipulates that:

> language use in interactions requires online processing and allows conclusions as to ability for use (Hymes 1972; Widdowson 1989), whereas the commonly used DCT involves offline processing and only allows conclusions as to knowledge.

Cognitive validity is a relatively new concept in language testing, hence little research, to my knowledge, has been conducted about the cognitive validity of DCTs. Taking into account existing findings about the psycholinguistic processes involved in responding to DCTs, and using methods that tap these processes, such as verbal reports,

future research might shed more light on this concept. I would like to turn now to another type of validity which has triggered a number of studies in the ILP literature, namely, construct validity. Often driven by the impetus to strengthen DCT design in order to improve its construct validity, a number of studies have investigated changes at the level of both DCT structure and content and their effects on DCT responses.

2. Previous research about DCT design

Most studies which have dealt with the issue of data collection methodology in ILP have focused on how DCT data differ from data collected by means of other methods (e.g. Hudson et al. 1995). A number of studies have compared data collected by means of role plays with those collected via DCTs (e.g. Sazaki 1998), Rose and Ono (1995) and Hinkle (1997) compared data collected by DCTs with multiple choice questionnaire data. Beebe and Cummings (1985) compared DCT data with ethnographic data. Such studies have generally established that, though practical, DCTs have low construct validity when compared with less controlled methods and with authentic data. However, such studies have also neglected the different DCT types and have treated the DCT as an instrument with one version. Few studies focused on the effects of DCT item design per se on responses (e.g. Billmyer and Varghese 2000; Johnston et al. 1998). The following section overviews ILP studies focusing on method induced effects of different DCTs.

2.1. *Effects of rejoinders*

Rose (1992) investigated the effects of hearer response on native speakers' production of requests in English. He used two questionnaires which were identical in all respects except in the inclusion of hearer response. He concluded that hearer response did not have a significant effect on the performance of requests by American speakers of English, but pointed out that such a study was culturally biased as it did not investigate the effect of hearer response in other languages, especially those "characterized by hearer-based interaction" (p. 60). As it may be the case that different cultures have different expectations about the hearer's role in a conversation. In a later study, comparing DCTs with multiple choice questionnaires, Rose (1994) investigated the cultural appropriateness of these two methods of data collection and concluded that level of directness in request production was affected by the type of method used. While both Americans and Japanese informants preferred more indirect strategies when responding to multiple-choice questionnaires, the Japanese used more direct strategies when responding to the DCTs. Doubting the cultural appropriateness of DCT situations in Rose's (1994) study, Rose and Ono (1995) further validated the cultural appropriateness of DCT situations. Their findings supported those of Rose (1994), namely, that the Japanese respondents were more direct when using the DCT than when responding to the multiple-choice questionnaire.

Another study focusing on the effect of structuring DCTs on the production of speech acts was conducted by Bardovi-Harlig and Hartford (1993). The study compared an 'open questionnaire' to a 'dialogue completion task' including a prompt, i.e., an

interlocutor's initiation of the conversation. The two methods affected differentially the responses of natives (NS) and non-natives (NNS), in that the respondents (NNS more than NS) produced more talk in dialogue completion tasks.

Johnston et al. (1998) investigated the effects of the different types of rejoinders in DCTs on NS and NNS production of complaints, requests, and apologies. The researchers concluded that the choice of particular strategies to realize the speech acts investigated was affected by the type of rejoinder. They further asserted that, based on such results, findings from studies using different types of DCTs cannot be compared and stressed the need for further validation studies.

It is worth noting at this level that whether the use of rejoinders itself contributes to making the DCT answers closer to real-life interactions by including the negotiation aspect characterizing authentic discourse, or whether it adds to the artificiality of the answers is a point to consider. Golato (2003: 93), for example, doubts whether rejoinders are beneficial, in making the situations more realistic, or detrimental in affecting the informants' responses. After all, most often when we interact with people we do not know how they would react to our speech and what they would say exactly. In the same vein, Roever (2004: 296) states that the main criticism leveled against the use of rejoinders has been related to "their effect on the authenticity of test taker responses, since in real-world communication one can only guess but not know for sure how the interlocutor will respond."

2.2. *Effects of content enrichment*

While almost all ILP studies have focused on the effects of contextual variables on the production of speech acts, few studies have considered the question of how much context description should be provided in DCT scenarios. Billmyer and Varghese (2000) investigated the effects of enriched contextual information on NS and NNS production of requests. In addition to situational factors considered in ILP studies so far, namely Power, Distance, and Rank of imposition (PDR), other contextual details built on Hymes' (1972) constituents of speech, were added to the situations. The researchers concluded that enriching the situational prompt or 'enhancing' the situation description produces "more robust external modification and elaboration than do the archetypal content-poor prompts which most DCT studies to date have used" (p.543). The researchers further pointed out that the few contextual variables considered in most speech act studies, namely, PDR, are not sufficient to elicit closer types of data to natural discourse. Similarly, citing Douglas (2000) who urges that prompt description should be detailed, Cohen (2004: 314) highlights the importance of the contextual features identified by Hymes.

Still in relation to PDR, one recent criticism has been leveled against the underrepresentation of context in terms of these three 'fixed' variables in pragmatics tests. Recent claims stress the negotiable aspects of these variables emphasizing that they are context dependent and that they are co-constructed by interactants while conversing (Grabowski 2008; Kasper and Ross 2013). Such a view approaches context not as a fixed, pre-defined, or static concept, but rather as a variable, dynamic, and

flexible notion built by discourse participants while interacting; hence the call to use interactive tasks in the assessment of pragmatic ability to allow for the negotiation aspect characterizing authentic situations to emerge in the testing situation. Reporting on Korsko's (2004) study, Grabowski (2008: 133) notes that in interactive DCTs, unlike in limited production DCTs, an interlocutor's response is affected by the following interlocutor's turn which is in turn affected by the following turn, hence their resemblance to natural turn-taking behavior and their advantage over single response tasks. While the interactive DCT may yield more authentic data than the traditional DCT, however, it may be more difficult to evaluate due to the complexity and interdependence of the conversational moves that it may generate.[6]

However, it is crucial to query whether thick description leads to improved output, i.e. to closer performance to real-life interactions on the part of the respondents. It may be the case that including several details in the situation description tires the informants and affects negatively their performance instead of helping them. Enhanced DCT versions, namely those relying on a detailed description of DCT situations, are, by definition, longer than versions including broad descriptions of scenarios. As such, they require more time to complete. For example, Roever (2004) argues that long prompts are cognitively demanding and learners might find them difficult. He asserts that, in any production test, "(T)he longer completion times negatively impact validity by limiting the number of items that can be administered and thereby narrowing the content domain and lowering test reliability" (p.296). He argues, however, in favor of limiting the range of possible responses and facilitating rating work by the inclusion of a rejoinder. In his later works, Roever (2011) calls for the development of tasks for the assessment of interactional competence of language learners. Multiple-rejoinder DCTs might thus receive more attention in the future. Cohen (2008: 221), for example, asserts that multiple-rejoinders are attempts to "make DCTs more reflective of the conversational turn-taking of natural speech."

To sum up what has been dealt with in the previous section, DCTs have been the most extensively used instruments in ILP but unfortunately little care has been paid to their construction despite the revealing findings of the few studies which have dealt with DCT structure and content. Whether the different DCT versions mentioned in Figure 1 are comparable and whether they measure the same construct raises questions of validity and reliability. In recent ILP literature, the DCT has still been treated as a 'one version instrument'. The following section argues for treating DCTs as language tests to minimize the method induced effects on the data they generate.

3. DCTs: Tests or questionnaires?

A survey of ILP studies conducted so far shows that the three terms 'discourse completion tests', 'discourse completion tasks', and 'production questionnaires' are used interchangeably to refer to the same method of data collection. Terms such as 'questionnaire' and 'test', may be misleading, however, in that the use of one or the other in its complete psychometric sense entails several implications for the study and the type of data generated. A questionnaire is different in a number of ways from a test.

---

[6] I am grateful to one of the reviewers for this observation.

Dörnyei (2003: 6-7) posits that the basic difference between tests and questionnaires lies in the fact that:

> (A) 'test' takes a sample of the respondent's behavior/knowledge and, on the basis of this sample, inferences are made about the degree of the development of the individual's more general underlying competence/abilities/skills (e.g., overall L2 proficiency). Thus, a test measures how *well* someone can do something. In contrast, questionnaires do not have good or bad answers; they ask for information about the respondents (or 'informants') in a non-evaluative manner, without gauging their performance against a set of criteria or against the performance of a norm group. (Italics original)

To apply such a definition in relation to DCTs as used in the field of ILP, it can be argued that the sample of respondents' behavior collected via a DCT is their performance of a given speech act in a particular situation. The inferences made about the informants' underlying abilities relate basically to their ability to perform the speech act in question in a target second (L2) or foreign (FL) language. The answers of the respondents are 'gauged', to use Dörnyei's term, against the performance of native speakers of the target language – since almost all ILP studies using DCTs have used a control group ('norm group') of native speakers of the language investigated to evaluate the performance of the learners/NNS. In fact, Dörnyei (2003: 7-8) states explicitly that DCTs *are not* questionnaires: "(T)hey are written, structured language elicitation instruments and, as such, they sample the respondent's competence in performing certain tasks, which makes them similar to language tests."

It is worth noting, however, that focusing on the stated objectives of the majority of ILP studies using DCTs, one notices that the DCT is rarely referred to as a language test. Let us recall the definition of ILP and then look at some examples of statements of objectives as formulated by some ILP studies. With reference to the focus of ILP studies Blum-Kulka et al. (1989: 10) point out that "in interlanguage pragmatics attention has been focused on *learners'* inappropriate speech act realizations in order *to uncover their pragmatic knowledge at a given time in their learning process*" (emphasis mine). Although in such a definition no terms related to the field of language testing are used, it can be safely deduced that the purpose of ILP studies, as stated in the above definition, is *to test learners' skills to perform speech acts in a given L2/FL at a specific proficiency level*. Interestingly, in her definition of ILP, Kasper (1997: 116) argues that, in this field:

> learners' and native speakers' comprehension and production of a particular speech act are compared, with the native speaker data serving as a baseline (cf. Kasper & Blum-Kulka 1993a). This approach has been useful in identifying where learners' and native speakers' linguistic action patterns are the same or different, and it tempting to conclude that the identified differences are the "stuff" that should go into the pragmatic component of a language teaching curriculum.

Such a definition seems to be closely related to Dörnyei's definition of language tests above, since according to Kasper, the performance of speech acts of language

learners is 'gauged' against that of native speakers in order to identify areas of difficulty (areas of difference) in using the target language and compensate for, or improve learners' performance in such areas in language teaching. In the same vein, discussing ways for teaching speech acts, Cohen (1996: 413-414) states that DCTs may be used as diagnostic tests "to assess the ability to produce the speech acts." Results yielded by means of "such assessment measures" then could be used to "plan teaching goals and procedures." Kasper and Blum-Kulka (1993: 12) maintain that:

> (I)n its canonical form, ILP research, following received methodology in interlanguage studies (Selinker, 1972) by comparing learners' IL production and comprehension with parallel L1 and L2 data, provides the methodological tool to determine where and how learners' pragmatic performance differs from L2, and to establish where IL specific behaviors appear to be influenced by learners' L1 knowledge.

In fact, having a deeper look at the statement of objectives of a number of ILP studies, one can notice that most of these studies refer to terms which are closely related to the field of language testing such as 'proficiency' 'ability', 'problems in using pragmatic knowledge', 'lack of competence', 'transfer', 'proficiency effects', and 'L2 pragmatic knowledge'. Despite the use of such terms, however, one can also notice that verbs such as to 'test', 'assess', 'measure', or 'evaluate' speech act performance/pragmatic ability/pragmatic competence are rarely used in the statement of objectives of ILP studies. Instead, general terms such as 'investigate', 'uncover', 'compare', 'examine', 'study', and sometimes simpler and more general terms are used to formulate research objectives. For example, referring to the ILP dimension of the Cross-Cultural Speech Act Realization Project (CCSARP)[7], Rintell and Mitchell (1989: 249) state that the project was formed, in part, "to *look at* the production of these speech acts [requests and apologies] by learners of a number of languages" (emphasis mine). Using an oral DCT to investigate Dutch English requests and how they relate to learners' assessment of sociopragmatic factors, Hendricks (2008: 337) states that the study aims at "*taking a detailed look* at how Dutch learners of English use request strategies and request modification in English…"(emphasis mine). Likewise, using a written DCT, Samavarchi and Allami (2012: 73) state that their study "intends *to investigate* the pragmatic competence of Iranian learners of English"(emphasis mine).

It is also relevant to point out that although ILP researchers often point to comparisons they make between learners' production of speech acts and that of natives', they generally do not specify the purpose of such comparisons from an assessment perspective. Focus is rather made on 'similarities', 'differences', 'transfer', 'pragmatic failure', and other issues which carry an assessment dimension. Discussing the use of the DCT in ILP, Rintell and Mitchell (1989: 250) state that the "method has been especially effective….for the comparison of strategies used by native speakers and learners of the same language." The purpose of such a comparison, however, was not defined with reference to language assessment.

Such a choice may be explained by a variety of reasons. First, and probably due to the paucity of ILP studies especially in non-Western contexts, most ILP researchers

---

[7] The CCSARP Project was initiated in 1982 and involved researchers from different countries to study speech act data collected from 13 languages and varieties.

prefer to endow their studies with an exploratory nature. As such, they conduct a particular study *to explore* how a speech act is performed by learners of an L2/FL which has not been investigated so far. Interest in ILP studies in later years and the significance of their findings for language teaching has continuously triggered researchers to investigate speech acts in a variety of languages apart from English. The speech act production studies using DCTs conducted recently (e.g. Bella 2014; Lin 2014; Salehi 2014; etc.) testify that the impetus in finding how learners with different L1s produce speech acts with an already available method at hands seems to have undermined interest in the validity of the DCT. However, a word of caution here is in order, if a particular study has an exploratory nature, then the researcher, should perhaps adhere to a descriptive analysis of the data as is the case in cross-cultural studies rather than venture into assessing learners' pragmatic ability. A second reason might relate to the replicative nature of a good number of ILP studies, i.e. those using the same version of the DCT with the same situations on the grounds of facilitating comparability of results. The focus of such studies has been on the findings rather than on the formulation of research objectives, the data collection methods employed, or the validity and reliability of results. Third, it is also probable that some researchers are cautious when dealing with testing and evaluation issues given all the implications such terms generate in relation to reliability and validity of research instruments employed.

Another explanation might relate to the objectives of the study in which the DCT is used. Several researchers have differentiated between studies intended for research and those intended for measuring speech act ability in the classroom (e.g. Roever 2004; Cohen 2004). Distinguishing between research instruments and assessment instruments, Roever (2004: 299) asserts that the latter "have real-world consequences for the test takers." He explains that while research instruments tend to have as their target data intended to reflect authentic instances of language use, assessment measures "are intended to engage knowledge, and the tester's definition of the construct under investigation determines what exactly that knowledge should be" (p. 297). Thus, while DCTs used for instructional purposes generate ratings and scores, and have real-world consequences for respondents, DCTs used for research purposes generate descriptive data. Yet, it is important to note that both types draw inferences about respondents' pragmatic ability and have implications on the field of language teaching. Cohen (2004: 319) asserts that:

> while the emphasis in the literature has clearly been on data collected for research purposes, the characteristics of the task and of the given respondents are bound to influence the results from speech act measures used for language assessment purposes as well[…]A question that prevails with any language assessment measure, and all the more so with pragmatic measures that produce variable data under the best of circumstances, is the extent to which the data collected are to some extent an artifact of the task itself.

Finally, it is also relevant to indicate the area of research for which the DCT is used, specifically, whether it is employed to study speech acts from a CCP or from an ILP perspective. Although the difference is subtle, it has important implications with regard to choice of data collection method. While CCP adopts the view that people from

different speech communities interact according to their own cultural norms, which may result in miscommunication (Boxer 2002), ILP focuses on how learners produce pragmatic behavior in a target language. Thus, if the DCT may seem an appropriate method to use in CCP, in that it succeeds in shedding light on the major semantic formulas used to realize a particular speech act in different speech communities, it may be an invalid test to use in ILP where it is used as a language test and therefore should undergo rigorous developmental stages before its administration.

In fact, the overlap between the areas of ILP and CCP has led not only to the inappropriate use of data collection methods, DCTs in particular, but also to narrowness of ILP research foci. Criticizing the narrow focus of ILP researchers on the ability of nonnative speakers to produce pragmalinguistically and sociopragmatically appropriate speech acts in the target language, Callies (2013) posits that ILP should encompass other areas of pragmatic research such as the interplay between syntactic and pragmatic knowledge. In trying to explain why the field of ILP has been limited to the study of speech act production, he cites Kasper (2010: 141) who asserts that ILP studies were highly influenced by CCP to the extent that they adopted "its research topics, theories and methods."

What is worth noting, however, is that this overlap has led to 'confusing' definitions of both areas: ILP and CCP. Such overlap can be seen, for example, in Martínez-Flor and Usó-Juan's (2006: 24) definition of CCP in which they state that CCP "deals with the comparison of learners' pragmatic performance with that of native-speakers (NSs)[…]" Such a definition of CCP as concerned with the comparison of learners' performance with that of natives might be confusing for a number of reasons. First, using learners as subjects in CCP studies is often done for convenience as educational settings provide access to homogeneous groups of subjects and not because CCP researchers are interested in studying the pragmatic competence of these learners. Second, it is worth remembering that CCP is concerned with how the realization of pragmatic behaviors, including speech act production, varies across cultures, and not with gauging learners' pragmatic behaviors against those of natives, which is the concern of ILP par excellence. It seems appropriate to call then for delimiting the boundaries between ILP and CCP especially when the overlap between these two fields affects the research methods employed.

In sum, although not directly stated, the DCT has been used in the majority of ILP studies to test L2/FL learners' ability to perform speech acts as compared with the standard of native speakers' performance. As such, it shares several qualities with language tests although its use has most of the times been limited to research rather than to instructional purposes. Accordingly, it seems safe to assume that the stages needed to develop a DCT should be similar to those followed to develop a language test.

4. DCT development process

Bachman and Palmer (1996) identify three main stages in the test development process: Design, operationalization, and administration. The design stage relates to the identification of the purpose of the test, the tasks in the target language use domain to which the test pertains, the characteristics of test takers, the definition of the construct to be measured, the identification of test usefulness, and the management of resources. The

operationalization stage includes the development of test specifications which would generate the tasks to be included in the test. The third stage relates to test administration. Due to the practical nature of the last two steps, however, this section will only detail the design stage of the DCT and provide general definitions of the operationalization and administration stages.

4.1. *Design*

Identifying the purpose of the test: A DCT may be used to collect data about how a speech act is performed in a given language for anthropological or cross-cultural purposes, i.e. to establish a speech act set for the speech act under study and/or compare how a particular speech act is realized in two different linguistic and/or cultural communities. It may also be used, as in the majority of ILP studies, to measure L2/FL learners' ability to perform a given speech act in an L2/FL.

   Identifying tasks in the Target Language Use (TLU) domain: Information collected via DCTs serve as predictions about similar performances in a real-life domain. As such, the TLU domain to which the DCT pertains is that of speech act performance in real-life situations. By investigating how informants perform speech acts when responding to a DCT, it is hypothesized that we are predicting how they would *actually* perform the speech act under investigation in authentic contexts.

   Defining the characteristics of test takers: DCT respondents may be NS of the language whose performance is used as a 'model' of how speech acts are performed in that particular language, or NNS who might be second or foreign language learners whose performance is used to evaluate their pragmatic ability in a target language. Most studies employing DCTs endeavored to control subject variability by using homogenous samples in terms of age, gender, educational background, and a variety of other personal factors.

   Defining the construct to be measured: In the design stage, the construct to be measured is defined theoretically. Bachman and Palmer (1996: 89) hold that "(I)n language testing, our theoretical construct definitions can be derived from a theory of language ability, a syllabus specification, or both." If DCTs are used as language tests in educational settings, the theoretical definition of the construct might be syllabus-based but if however they are used to collect information about how informants to the study perform speech acts in 'similar' real-life situations for research purposes, then the theoretical definition of the construct DCTs are intended to measure should be developed from a theory of language ability. It is important to note at this level that one of the main criticisms leveled against tests purporting to measure pragmatic ability in general and DCTs in particular is the underrepresentation of the construct being measured. Grabowski (2007: 1) explains that "(A)lthough some researchers have attempted to measure aspects of pragmatic knowledge, the vast majority have not incorporated a clearly articulated pragmatic component into the test construct."

   Developing a plan for evaluating the qualities of usefulness: For Bachman and Palmer (1996: 17), "(T)he most important consideration in designing and developing a language test is the use for which it is intended, so that the most important quality of a

test is its usefulness." In their model of test usefulness they include six qualities: Reliability, construct validity, authenticity, interactiveness, impact, and practicality. In what follows these qualities will be first defined according to the researchers' model then applied to the DCT. It is worth noting, however, that according to Bachman (2004: 5), test usefulness cannot be determined theoretically without reference to the performance of test takers as "the evaluation of test usefulness must include the empirical investigation of test performance." He states that such investigation must include "the processes or strategies test takers use in responding to specific tasks" and the responses and scores obtained in responding to these tasks. This section, therefore, will only overview the qualities of usefulness as empirical investigation is beyond the scope of this paper.

a- Reliability: Bachman and Palmer (1996: 19) define reliability as "consistency of measurement." As such, a DCT may be said to be reliable if the respondents' performance to one version of the instrument parallels their performance to another version of the same instrument. The reliability of different DCT versions has thus to be researched.

b- Construct validity: for Bachman and Palmer (1996: 21) a construct is "the specific definition of an ability that provides the basis for a given test or test task and for interpreting scores derived from this task." Construct validity within such a model of test development refers to "the extent to which we can interpret a given test score as an indicator of the ability(ies), or construct(s), we want to measure" (italics original). Stressing the importance of construct validity, Roever (2006: 234-235 cited in Rylander et al., 2013: 67) explains that "the more the test reflects the construct, the more likely it is that scores obtained on it will be an accurate representation of the strength of attributes associated with that construct." The construct validity of different DCT versions is an issue to consider in ILP. While the general construct that DCTs are reported to measure is 'the ability to produce speech acts in a target language', different DCT versions might provide different measures of this ability.

c- Authenticity: Bachman and Palmer (1996: 23) explain that authenticity relates to "the degree of correspondence of the characteristics of a given language test task to the features of a TLU task." They hold that for a test to be useful, performance on that test should correspond "to language use in specific domains other than the language test itself." In relation to DCTs, authenticity relates to the degree of realism involved in both speech act situations included in the DCT and subjects' performance when responding to these situations. Thus, it is hypothesized that the situations used *are* real-life situations and that what informants report they would say in these situations is what they would *really* say in authentic language performance contexts. Different DCT versions again may have different degrees of authenticity.

d- Interactiveness: Bachman and Palmer (1996: 25) explain that interactiveness relates to "the extent and type of involvement of the test taker's individual characteristics in accomplishing a test task." In relation to DCTs, the individual factors that are likely to affect test takers' performance on the DCT might include, but are not limited to, age, gender, linguistic and cultural background, level of proficiency, and length of residence in the target language community.

e- Impact: this relates to the "impact on society and educational systems and upon the individuals within those systems" (Bachman and Palmer, 1996: 29). As far as DCTs are concerned, the impact of the test may be instructional if the test is used to measure learners' pragmatic ability and build on the findings to design or improve pragmatics teaching materials, or it may be anthropological if it is used to investigate how a speech act is realized in a particular linguistic/cultural community.

f- Practicality: this concerns "the relationship between the resources that will be required in the design, development, and use of the test and the resources that will be available for these activities" (Bachman and Palmer, 1996: 35). Despite the various disadvantages that DCTs have, most studies employing some version of a DCT opted for this instrument on the basis of its practicality since it enables researchers to collect and codify large amounts of data in relatively short time periods.

Identifying resources and developing a plan for their allocation and management: In terms of human resources, the researcher undertaking the study is usually the test developer, writer, rater, and administrator in studies conducted for research purposes. It is important to note, however, that in some studies in order to increase inter-rater reliability, some researchers have the data coded by other raters. In terms of material resources, administration of the DCT basically requires space (a room where informants may respond to the DCT), and photocopied material (where the informants will write what they would say as a response to a particular speech act situation). In terms of time, the DCT may be demanding in the design stage, especially if the researcher does not rely on previously developed versions of the DCT to collect data for his/her study. As such, developing specifications for the DCT and eventually writing culturally appropriate 'real-life' situations may be demanding on the researcher in terms of time. Time will also be needed to rate (if the purpose is instructional) or codify and analyze (if the purpose is research) the data collected.

4.2. *Operationalization*

Bachman and Palmer (1996: 90) define this stage in relation to the development of test specifications for the tasks which will be included in the test. In addition to the specifications, this stage involves how the tasks will be organized and sequenced in the DCT, writing the tasks, writing the instructions and the scoring procedures. They hold that "by specifying the conditions under which language use will be elicited and the method for scoring responses to these tasks, we are providing the *operational definition* of the construct" (italics original).

4.3. *Administration*

In this stage, the DCT is administered to a group of informants, data are collected and analyzed in relation to two purposes: "assessing the usefulness of the test," and "making the inferences or decisions for which the test is intended" (Bachman and Palmer 1996: 91).

5. Discussion and conclusion

The present paper has overviewed studies about the written DCT with a special focus on those dealing with DCT design. The paper has pointed out that such studies have been triggered in part by criticism against the low construct validity of DCTs and argued for researching their cognitive validity. The final section of the paper has argued that the written DCT should be treated as a language test, particularly in ILP studies, and has outlined the DCT development process.

Recent research in pragmatics testing has made use of technology to upgrade pragmatics tests. Sophisticated statistical tools such as FACETS, Rash Measurement, Generalizability Theory (e.g. Brown 2008; Roever 2008) and Differential Item Functioning (e.g. Roever 2007, 2010) have been used to investigate different pragmatics tests including DCTs. CALT, and both web-based and video-based instruments have been suggested as alternative tools for the assessment of pragmatic ability (e.g. Roever 2006; Rylander et al. 2013). It is worth pointing out, though, that the use of such tools might be restricted to researchers (if the purpose is research) and to teachers (if the purpose is instructional) who can have access to and afford the use of such tools in their research/instructional contexts. It would appear that there needs to be a shift from pragmatics assessment research involving 'closed laboratory instruments' that can only be accessed by a small group of researchers, to applying these measures in the broader second/foreign language contexts involving countries with limited logistics and an eminent need to assess the pragmatic competence of their learners for both instructional and research purposes. Another recent trend, partly triggered by the low validity of the DCT, has advocated the use of conversation analysis tools for the assessment of learners' interactional competence (see Walters 2013 for more on conversation analysis-informed tests). Although such an approach seems promising in eliciting more authentic instances of language use, it might not be very practical in terms of the load it makes when coding and scoring or analyzing the data collected. As such, it is expected that in many second/foreign language contexts, the traditional DCT will remain to be used.

To conclude, collecting speech act data with an available method at hands has been a temptation for a number of ILP researchers. Interest in studying different languages and cultures has undermined criticism of instruments used to collect such data. The DCT has been and will probably remain the most used instrument to collect pragmatics data. Bardovi-Harlig (2013: 71) asserts that controlled tasks, DCTs included, "continue to evolve as researchers investigate questions in acquisition that require contexts that can be finely manipulated." Some attempts at improving the instrument have been made but little attempts have been made to validate the different versions of the written DCT in different cultures and/or language groups or to develop it following the same steps required in developing any other language test. Future research might

focus on the development of DCT specifications taking into account the revealing findings of ILP studies conducted so far.

Acknowledgements

References

Bachman, L.F., and A.S. Palmer (1996) *Language testing in practice*. Oxford: Oxford University Press.

Bachman, L.F. (2004) *Statistical analyses for language assessment.* Cambridge: Cambridge University Press.

Bardovi-Harlig, K., and B.S. Hartford (1993) Learning the rules of academic talk: A longitudinal study of pragmatic change. *Studies in Second Language Acquisition* 15.3: 279-304.

Bardovi-Harlig, K. (2013) Developing L2 pragmatics. *Language Learning* 63.1: 68-86.

Bax, S. (2013). Readers' cognitive processes during IELTS reading tests: Evidence from eye tracking. *ELT Research Papers* 13-06.

Beebe, L.M., and M.C. Cummings. (1985) Speech act performance: A function of the data collection procedure? Paper presented at the TESOL convention, New York.

Bella, S. (2014) A contrastive study of apologies performed by Greek native speakers and English learners of Greek as a foreign language. *Pragmatics* 24.1: 679-713.

Bergman, M.L., and G. Kasper (1993) Perception and performance in native and non-native apology. In G. Kasper, and S. Blum-Kulka (eds.), *Interlanguage pragmatics*. Oxford: Oxford University Press, pp. 82-107.

Bialystok, E. (1993) Symbolic representation and attentional control. In G. Kaspe,r and S. Blum-Kulka (eds.), *Interlanguage pragmatics*. Oxford: Oxford University Press, pp. 43-57.

Billmyer, K., and M. Varghese (2000) Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. *Applied Linguistics* 21.4: 517-552.

Blum-Kulka, S. (1982) Learning to say what you mean in a second language: A study of the speech act performance of learners of Hebrew as a second language. *Applied Linguistics* 3.1: 29-59.

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

Boxer, D. (2002) Discourse issues in cross-cultural pragmatics. *Annual Review of Applied Linguistics* 22: 150-167.

Brown, J.D. (2008) Raters, functions, item types and the dependability of L2 pragmatics tests. In E.A. Soler, and A.M. Flor (eds.), *Investigating pragmatics in foreign language learning, teaching and testing* (Vol. 30). *Multilingual Matters* 224-48.

Callies, M. (2013) Advancing the research agenda of Interlanguage Pragmatics: The role of learner corpora. In *Yearbook of Corpus Linguistics and Pragmatics 2013*. New York**:** Springer, pp. 9-36.

Cohen, A.D. (1996) Developing the ability to perform speech acts. *Studies in Second Language Acquisition* 18.2: 253-269.

Cohen, A.D. (1996) Speech acts. In N. Hornberger, and S. McKay (eds.), *Sociolinguistics and language teaching*. Cambridge: Cambridge University Press, pp. 382-419.

Cohen, A.D. (2004) Assessing speech acts in a second language. In D. Boxer, and A.D. Cohen (eds.), *Studying speaking to inform second language learning* (Vol. 8). *Multilingual Matters* 302-327.

Cohen, A.D. (2008) Teaching and assessing L2 pragmatics: What can we expect from learners? *Language Teaching* 41.2: 213-235.

Cohen, A.D., and E. Olshtain (1994) Researching the production of second-language speech acts. *Research methodology in second-language acquisition* 143-156.

Dornyei, Z. (2003) *Questionnaires in second Language Research: Construction, administration, and processing*. Mahwah, NJ: Lawrence Erlbraum.

Douglas, D. (2000) *Assessing languages for specific purposes*. Cambridge, UK: Cambridge University Press.

Ellis, R. (2004) The definition and measurement of L2 explicit knowledge. *Language Learning* 54 : 227–275.

Faerch. C., and G. Kasper (1984) Pragmatic knowledge: Rules and procedures. *Applied Linguistics* 5.3: 214-225.

Foster, P., and P. Skehan (1996). The influence of planning and task type on second language performance. *Studies in Second language acquisition* 18.3: 299-323.

Golato, A. (2003) Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied linguistics* 24.1: 90-121.

Grabowski, K.C. (2007) Reconsidering the measurement of pragmatic knowledge using a reciprocal written task format. *Columbia University Working Papers in TESOL & Applied Linguistics* 7.1: 1-48.

Grabowski, K.C. (2008) Investigating the construct validity of a performance test designed to measure grammatical and pragmatic knowledge. *Spaan Fellow Working Papers in Foreign Language Assessment* 6: 131-179.

Hinkel, E. (1997) Appropriateness of advice: DCT and multiple choice data. *Applied linguistics* 18.1: 1-26.

Hendriks, B. (2008) Dutch English requests: A study of request performance by Dutch learners of English. In M. Pütz, and J. Neff-van Aertselaer (eds.), *Developing contrastive pragmatics: Interlanguage and cross-cultural perspectives* (Vol. 31). Berlin: Walter de Gruyter, pp. 335-354.

Hudson, T., E. Detmer, and J.D. Brown (1995) *Developing prototypic measures of crosscultural pragmatics* (Tech. Rep. No 7). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.

Hymes, D. (1972) On communicative competence. *Sociolinguistics* 269-293.

Johnston, B., G. Kasper, and S. Ross (1998) Effect of rejoinders in production questionnaires. *Applied Linguistics* 19.2: 157-182.

Kasper, G. (1997) The role of pragmatics in language teacher education. In K. Bardovi-Harlig, and B. Hartford (eds.), *Beyond methods. Components of second language teacher education*. New York: McGraw Hill, pp. 113-141.

Kasper, G. (2010) Interlanguage pragmatics. In M. Fried, J.O. Östman, and J. Verschueren (eds.), *Variation and change: Pragmatic perspectives* (Vol. 6). Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 141-154.

Kasper, G., and M. Dahl (1991) Research methods in interlanguage pragmatics. *Studies in Second Language Acquisition* 13: 215-247.

Kasper, G., and S. Blum-Kulka (eds.) (1993) *Interlanguage pragmatics*. Oxford: Oxford University Press.

Kasper, G., and S. Ross (2013) Assessing second language pragmatics: An overview and introductions. In S. Ross, and G. Kasper (eds.), *Assessing Second Language Pragmatics*. Basingstoke, UK: Palgrave Macmillan, pp. 1-40.

Korsko, P. (2004) The narrative shape of two-party complaints in Portuguese: A discourse analytic study. Unpublished doctoral dissertation, Teachers College, Columbia University, New York City.

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

Lin, M.F. (2014) An interlanguage pragmatic study on Chinese EFL learners' refusal: Perception and performance. *Journal of Language Teaching and Research* 5.3: 642-653.

Martínez-Flor, A., and E. Usó-Juan (2006) Learners' use of request modifiers across two University ESP disciplines. *Ibérica* 12: 23-41.

McNamara, T.F., and C. Roever (2006) *Language testing: The social dimension*. Oxford: Blackwell.

Olshtain, E., and L. Weinbach (1993) Interlanguage features of the speech act of complaining. In G. Kasper, and S. Blum-Kulka (eds.), *Interlanguage pragmatics*. Oxford: Oxford University Press.

Rintell, E., and C.L. Mitchell (1989) Studying requests and apologies: An inquiry into methods. In S. Blum-Kulka., J. House, and G. Kasper (eds.), *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Albex, pp. 248-72.

Roever, C. (2004) Difficulty and practicality in tests of interlanguage pragmatics. In D. Boxer, and A.D. Cohen. (eds.), *Studying speaking to inform second language learning* (Vol. 8). *Multilingual Matters* 283-301.

Roever, C. (2006) Validation of a web-based test of ESL pragmalinguistics. *Language Testing* 23.2: 229-256.

Roever, C. (2007) DIF in the assessment of second language pragmatics. *Language Assessment Quarterly* 4.2: 165-189.

Roever, C. (2008) Rater, item and candidate effects in discourse completion tests: A FACETS approach. In E.A. Soler, and A.M. Flor (eds.), *Investigating pragmatics in foreign language learning, teaching and testing* (Vol. 30). *Multilingual Matters* 249-266.

Roever, C. (2010) Effects of cultural background in a test of ESL pragmalinguistics: A DIF approach. *Pragmatics and language learning 12*.

Roever, C. (2011) Testing of second language pragmatics: Past and future. *Language Testing* 28.4: 463-481.

Rose, K.R. (1992) Speech acts and questionnaires: The effect of hearer response. *Journal of pragmatics* 17.1: 49-62.

Rose, K.R. (1994) On the validity of discourse completion tests in non-Western contexts. *Applied Linguistics* 15.1: 1-14.

Rose, K.R., and R. Ono (1995) Eliciting speech act data in Japanese: The effect of questionnaire type. *Language learning* 45.2: 191-223.

Ross, S., and G. Kasper (eds.) (2013) *Assessing Second Language Pragmatics*. Basingstoke, UK: Palgrave Macmillan.

Rylander, J., P. Clark, and R. Derrah (2013) A video-based method of assessing pragmatic awareness. In S. Ross, and G. Kasper (eds.), *Assessing Second Language Pragmatics*. Basingstoke, UK: Palgrave Macmillan, pp. 65-97.

Salehi, R. (2014) A comparative analysis of apology strategy: Iranian EFL learners and native English speakers. *Procedia-Social and Behavioral Sciences* 98: 1658-1665.

Samavarchi, L., and H. Allami (2012) Giving condolences by Persian EFL learners: A contrastive sociopragmatic study. *International Journal of English Linguistics* 2.1: 71-78.

Sasaki, M. (1998) Investigating EFL students' production of speech acts: A comparison of production questionnaires and role plays. *Journal of Pragmatics* 30.4: 457-484.

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

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

Walters, F.S. (2013) Interfaces between a discourse completion test and a conversation analysis-informed test of L2 pragmatic competence. In S. Ross, and G. Kasper (eds.), *Assessing Second Language Pragmatics*. Basingstoke, UK: Palgrave Macmillan.

Wolfson, N. (1989) *Perspectives: Sociolinguistics and TESOL*. New York: Newsbury House.