

Assessing Pragmatics through MDCTs: A Case of Iranian EFL Learners

Majid Nemati

Associate Professor
University of Tehran
nematim@ut.ac.ir

Abbas Ali Rezaee

Associate Professor
University of Tehran
aarezaee@ut.ac.ir

Mohammad Mahdi Hajmalek

Assistant Professor
Khatam University
m.hajmalek@khatam.ac.ir

Abstract

Testing pragmatic competence has always posed daunting challenges to researchers and practitioners. As a step to address the gap in pragmatic testing, this paper delineates the procedural stages of developing and validating a context-sensitive Multiple Choice Discourse Completion Test (MDCT). Following a unitary view of validity, tenable argument and empirical evidence was accumulated to support the construct validity of the test. While 136 advanced learners of English took the developed test, it exhibited reasonable internal consistency ($\alpha = .72$). Furthermore, the results of correlational studies revealed acceptable association between the scores of this MDCT with both a written version of the same test and a previously validated test of pragmatics from the literature endorsing its concurrent validity. Also, the MDCT proved efficient in differentiating between native speakers and EFL learners as suggested by the results of an independent samples t-test. Finally, a pretest-posttest experimental study with 26 intermediate EFL learners was designed to check the sensitivity of the test towards developments in learners' interlanguage after which a t-test analysis corroborated the construct validity of the test. This array of evidence denotes that the suggested MDCT can be reliably used in EFL contexts as a valid measure of pragmatic competence.

Keywords: Pragmatics Assessment, Pragmatic Comprehension, DCT, MDCT

Received: January 2014; Accepted: September 2014

1. Introduction

Once it was believed that learning a second language is defined by and confined to mastering the formal aspects of the new language including its grammar and lexicon. However, the debut of an alluring new concept, namely communicative competence, in the 1970s (Hymes, 1972, 1974) and its later descendent, pragmatic competence (Bachman, 1990) changed our view of second language acquisition. Since the introduction of interlanguage pragmatics to language teaching profession by a series of pioneering studies including Brown and Levinson (1987), Kasper and Blum-Kulka (1993), and Kasper and Dahl (1991), to name a few, language testing has been influenced in line with this new trend of research in order to accommodate the new findings on the significant role of pragmatic competence in successful communication. Therefore, researchers and practitioners soon started to feel the need to develop new language testing techniques, tools, and procedures to assess the pragmatic competence of EFL or ESL learners. The Cross-Cultural Speech Act Realization Project (CCSARP) (Blum-Kulka, House, & Kasper, 1989) is a good example of early endeavors to operationalize the assessment of pragmatics. As one of the pioneering efforts to conceptualize and assess the pragmatic competence of language learners, this was followed by the development or adaptation of several assessment tools. However, after more than two decades of research, testing pragmatics is still considered a nascent field of study in need of nurturing (Roever, 2011). Devising a simultaneously reliable, valid, and practical method to assess pragmatics especially in large scales and in standardized testing situations continues to pose serious challenges to ILP researchers. To address this gap, the present study argues for the application of Multiple Choice Discourse Completion Tests (MDCTs) as a practical and efficient measurement tool, and reports on the process of developing and

validating a context sensitive multiple choice test of interlanguage pragmatics in Iran. Hence, following Messick (1989) and Bachman's (1990) framework of validity as a unitary concept and Bachman and Palmer's (2010) Assessment Use Argument (AUA), the present study asked whether it is possible to confidently employ multiple-choice testing in pragmatics. First, the procedures of test development will be explained in detail and, consequently, the study will report on the correlational and experimental evidence accumulated to establish an argument to support the use of the offered MDCT in pragmatic testing despite the criticisms alleged against multiple-choice question test formats.

2. Review of the Related Literature

A variety of diverse and multidimensional data collection or, interchangeably, assessment methods are frequently employed in ILP research (Ellis, 2008). However, devising an appropriate pragmatic test (PT) which collects relatively naturalistic data and at the same time accommodates researcher control has proved to be less than easy in ILP research (Nguyen, 2014). While these assessment methods have been applied to a number of ILP dimensions, they would generally conform to either of the two traditions in ILP research, namely sociopragmatics and pragmalinguistics as famously dichotomized by Leech (1983). As Roever (2007) reports, one of the largest projects venturing upon developing a test of pragmatics in the former tradition is Hudson, Detmer, and Brown (1995) who investigated the validity and compared the effects of different types of assessment tools. In the other side of the continuum, Roever (2006), who developed a web-based test of pragmatics, is considered the quintessence of testing pragmalinguistics.

No matter which of the above-mentioned traditions they represent, a great proportion of research in ILP has dealt with production and, to a lesser extent,

comprehension of speech acts (Roever, 2011). While Discourse Completion Tasks (DCTs) have been very popular in ILP research following the speech act paradigm (Roever, 2011), other possible alternatives have been similarly prevailing. On the whole, these assessment tools are classifiable along a continuum ranging from open-ended tasks including written DCTs and role plays to focused tasks such as C-tests and multiple choice DCTs (Bardovi-Harlig, Mossman, & Vellenga, 2014). *DCTs* were first introduced to the field of interlanguage pragmatics in the early 1980s. Some of the earliest examples of the appearance of such tests to assess the realization of speech acts in the literature are Blum-Kulka (1982), Blum-Kulka, House, and Kasper (1989), and Bardovi-Harlig and Hartford (1993). DCTs provide participants with written descriptions of a social situation, or *scenarios*, followed by an incomplete dialogic interaction requiring the examinee to complete the conversation with whatever they find themselves most likely to say in a similar situation in real life (Brown & Ahn, 2011). DCT items can possibly take other varieties depending on the medium employed. For instance, in an ODCT, or oral DCT, participants are asked to utter their responses instead of writing them down whereas in an MDCT or multiple-choice DCT one of the provided options is selected as the most suitable response (Brown, 2001, 2008). Alternatively, integration of audio or video recordings instead of written descriptors is also possible (Bardovi-Harlig & Dörnyei, 1998; Niezgodna & Röver, 2001).

However, DCTs have been criticized on several grounds. Comparisons between the data elicited through DCTs with naturally occurring data point to a divergence in length or a number of other linguistic elements including repetitions, inversions, and omissions (Golato, 2003; Hartford & Bardovi-Harlig, 1992; Yuan, 2001). On the other hand, a number of studies have delineated the use of similar words and phrases in both DCTs and natural

Assessing Pragmatics through MDCTs: A Case...

conversations (Bodman & Eisenstein, 1998; Einstein & Bodman, 1993). Bardovi-Harlig et al. (2014) also list some drawbacks of open-ended tasks including the probability of eliciting an undesirable form rather than the targeted one or lack of an obligatory context, uneven opportunities for all participants to contribute, and interdependence of the interlocutors' choices in pair and group work. Despite all these criticisms, DCTs remain to be a popular elicitation tool while collecting data from both native speakers and learners.

Along with DCTs and MDCTs, *role plays, role enactments, self-assessment tasks, acceptability judgment tasks, and verbal protocols* are other possible varieties of PT. Among these, MDCTs are proved as the most convenient in terms of practicality on both levels of administration and scoring (Roever, 2011). Furthermore, they are particularly favorable when it comes to the assessment of pragmatic awareness. A number of studies have supported the role of awareness, consciousness raising, and metapragmatic knowledge in successfully mastering and employing pragmatic norms in communication (see Bardovi-Harlig, 2014; Garcia, 2004; Ishihara, 2007; and McConachy, 2013 for some examples). However, there have been several criticisms leveled at the reliability of such tests including the ones alleged by Brown (2001, 2008). To address these criticisms researchers have attempted to develop reliable MDCTs among which Tada (2005) and Liu (2006) have demonstrated some success, to name a few. Tada utilized video prompts to evolve an ODCT as well as an MDCT and obtained reliabilities of around .75 for both modalities. In another study, Liu (2006) compared a WDCT and an MDCT in a similar context and acquired a very high reliability of around .90 for the MDCT. However, since the native speaker responses were exclusively used as the target option and the learners' responses as distracters, it is disputable that the idiomaticity inherent in the natives' answers resulted in some bias on the side of

the respondents (McNamara & Roever, 2006). In this regard, it has been asserted that manipulating a number of elements in MDCTs can largely influence the reliability of the measure. For instance, Roever (2008) obtained a reliability of .90 for his MDCT by adding rejoinders, i.e., responses by a hypothetical interlocutor.

Particularly in PT, to summarize, warnings as early as Hartford and Bardovi-Harlig (1992), Beebe and Cummings (1995), and Rose (1997) and as recent as Ross and Kasper (2013), Brown and Ahn (2011), Roever (2011), Cohen (2010), and Fujiwara (2007)

have already pointed to the shortcomings and possible caveats inherent in different data collection modalities in ILP research, especially all varieties of DCTs, and have called for judicious application of these tests and cautious interpretation of their results. To address this need, the present study ventures upon the development of a reliable, practical, and context appropriate pragmatics test and forms an argument in favor of its application in courses of English as a foreign language especially in cases of large-scale assessment in Iran. Therefore, the present study sought to address the aforementioned gap in PT and asked whether a reliable and valid MDCT can be developed and implemented in Iran.

1. Method

Aside from the native and non-native speakers of English who collaborated in a series of pilot studies leading to the development of the MDCT, a sample of 136 male and female advanced EFL learners took the final version of the test in the development phase of the study and the data from this group was used in the correlational analysis of test validity. This cohort, coming from nine classes, was selected based on their results on Oxford Placement Test (Allan, 1992)

Assessing Pragmatics through MDCTs: A Case...

from a population of male and female students aged between 18 and 26 majoring in English literature at University of Tehran or in medicine at Medical School of Shahid Beheshti University and learners in Advanced levels at Iran Language Institute (The ILI). Another group of 26 male intermediate learners, aged between 16 and 21, forming an intact class at the ILI participated in the experimental validation study.

Among a myriad of possibilities, the present study narrowed its scope to five speech acts of *request, apology, refusal, suggestion, and gratitude*. These speech acts, as the most widely probed ones under the speech act paradigm in ILP research (Roever, 2011), are opted for due to their higher frequency of appearance in everyday conversations including academic context interactions as the focus of the present study. In order to fulfill the goals set by the present study, different steps of test development will be discussed in the following sections. Furthermore, the results of both non-experimental (mainly correlational) and experimental studies and the consequent data analyses will be presented to constitute the necessary statistical evidence required to argue for the validity of the prepared test.

3.1. Step 1: Development of DCT

In order to generate scenarios corresponding to each speech act, several sources from the literature and corpus data were consulted. Initially, hints on the suitability, authenticity, and frequency of situations were taken from the Michigan Corpus of Academic Spoken English (MICASE) (Simpson, Briggs, Ovens, & Swales, 2002), a rich corpus of approximately 1.8 million words of academic interaction comprising nearly 200 hours of audio recorded during classroom discussions, laboratory sessions, conferences, lectures, and advising sessions. Eventually, the scenarios for these selected situation were adopted

and/or adapted from Bardovi-Harlig and Dörnyei (1998), Cohen and Olshtain (1993), Eisenstein and Bodman (1986), Eslami-Rasekh (2005, 2010), Jiang (2006), Kasper and Blum-Kulka (1993), Kondo (2010), Malamed (2010), Martinez-Flor (2010), Martinez-Flor and Alcon-Soler (2007), Safont-Jorda (2004), Schauer and Adolphs (2006), and Uso-Juan (2010). In order to run a “likelihood investigation” (Jianda, 2006), 10 scenarios were listed under each speech act and were presented to a class of 24 university students to rate their probability of occurrence in an Iranian campus context on a scale from 1 to 5. Finally, 4 of the most probable situations incorporating each speech act as rated by the participants were selected for inclusion in the final questionnaire. It was believed that a higher number of situations would significantly affect the participants’ sincere collaboration and jeopardize the practicality of both administration and scoring processes by excessively elongating the DCT. The selected situations and scenarios were converted into discourse completion items by removing the line encompassing the target speech act and randomly distributing items throughout the questionnaire. As an early pilot, the test was presented to 10 experts including Ph.D. holders and graduate students of TEFL and native speakers of English as well as 13 intermediate learners of English in order to check the comprehensibility and eligibility of the descriptors and conversations. Also, they were asked to name the speech act elicited by each scenario to ensure the applicability of the item. Consequently, based on the feedback obtained from the responses, necessary modifications were applied to eliminate any ambiguity or ambivalence. These modifications for the most part included paraphrasing, adding extra comments, adding a line by the first interlocutor, overtly elucidating the relationship between interlocutors including power relations, and explicitly naming the speech act that the item

was trying to elicit. At this stage, the final version of the WDCT was ready for administration.

3.2. Step 2: Development of the MDCT

Once the WDCT or the open-ended version of the questionnaire was ready, it was administered to a sample of 136 advanced learners of English as described earlier. Simultaneously, eight native speakers of English were asked to complete an online version of the same test uploaded to Murvey.com. An analysis of the replies to these written tasks launched the next step, that is, the development of the multiple choice version of the questionnaire (MDCT). As the first step to architecture the multiple choice test, the responses to each item by the NS (native speaker) collaborators and 20 randomly selected learners were analyzed based on the criteria introduced by Hudson, Detemer, and Brown (1995) who based their assessment on three context factors identified by Brown and Levinson (1987): *Power*, *Social Distance*, and *Imposition*. In this framework, the relative *Power* of the interlocutors refers to the respective social status of them and the authoritative position of one over the other, such as that of a teacher and student. The second criterion, degree of *Social Distance*, pertains to the membership of the participants in a similar social group while the last measure, the degree of *Imposition*, concerns the costs, damages, or consequences proceeding from the request, the mistake, the refusal, the suggestion or any other action embedded in the speech act. Based on this three-dimensional framework, the most frequently occurring responses of the NS respondents or combinations of their answers were selected along with a couple of the learners' responses assuming to violate the abovementioned criteria. Any grammatical infelicities in NNSs' (non-native

speakers) answers were eliminated in order to ensure the linguistic precision of each option and restrict the decision makings to pragmatic appropriateness.

Consequently, each scenario was accompanied with six to eight choices. A group of 10 NSs of English who had not encountered the questionnaire previously were asked to rank the responses listed for each scenario based on the perceived appropriateness. Next, the most frequently suggested response along with three of the least favored ones were employed to construct a primitive version of the MDCT. As a pilot run, the test was taken by 32 teachers of English in Iran and final fine-tunings and modifications were made to the test based on the statistical analysis of the options selected by these teachers as well as the occasional feedback offered by them. Particularly, any distractors which had attracted a competitively high attention were eliminated or revisited. Subsequently, an online copy of the test was presented to a web-based community of English native speakers to corroborate the choices as keys and ensure the robustness of the test. After 14 days of survey, 61 NSs took the test voluntarily. At this stage, it was ensured that all choices and particularly the distractors were functioning appropriately and very minor revisions were applied in a couple of cases. The outcome was the final product of the study as a multiple choice pragmatic test.

3.3. Step 3: Coding and Ratings

The MDCT was consequently administered to the same sample in the same nine classes of advanced EFL learners as previously described. The WDCTs from these 136 learners were scored by two raters based on scoring criteria in form of a grid. This grid was prepared based on three considerations: a) the three-dimensional rubric elaborated on earlier in this text (i.e., *Power*, *Social Distance*, and *Imposition*, Blum-Kulka et al., 1989); b) the similarity of the

Assessing Pragmatics through MDCTs: A Case...

responses to the keys of the MDCT in terms of the length of the utterances; c) speech act-specific criteria adapted from the literature as described in the rest of this paragraph. Cohen and Olshtain's (1981) taxonomy of apologies was applied to compare respondents' answers to apology situations with those of native speakers. This taxonomy comprises the 5 following categories: a) an expression of apology (including an expression of regret, an offer of apology, and a request for forgiveness), b) an offer of repair or redress, c) an explanation of an account, d) acknowledging responsibility for the offense, and e) a promise of forbearance. Further, Cheng's (2005) categorization of thanking strategies encompassing the following eight categories was applied to gratitude items: a) Thanking, b) Appreciation, c) Positive feelings, d) Apology, e) Recognition of imposition, f) Repayment, g) Other strategies, and h) Attention getter. In case of requests (items 2, 4, 6, and 14), directness and indirectness of strategies (Trosborg, 1995) were investigated as well as the use of mitigating devices (Schauer, 2004, 2007). Jiang (2006) identified nine distinct structural categories for suggestions against which the corresponding items in the present study were gauged. These categories include a) Let's... b) Modals, c) WH-questions, d) Conditionals, e) Performatives, f) Pseudo clefts, g) extraposed to-clauses, h) Yes-No questions, i) Imperatives. Finally, in order to operationalize an assessment of refusals, Beebe, Takahashi, and Uliss-Weltz's (1990) widely known taxonomy was applied. Beebe et al. (1990) identified three main categories of refusal strategies: a) Direct refusals (including performatives and non-informatives), b) Indirect refusals (including ten subcategories of regret, wish, excuse/reason, statement of alternatives, setting conditions, promise of future acceptance, statement of principle, statement of philosophy, attempt to dissuade interlocutor, mitigated refusal, and avoidance), and c) Adjunct refusals (comprising positive opinion/feeling or disagreement,

statement of empathy, gratitude/appreciation, and pause fillers). A five-level Likert scale was used to evaluate each criterion. Consequently the MDCTs were scored based on the key and all scores were submitted to IBM SPSS software version 22.

3.4. Validation Studies

As explicably discussed by Messick (1989), Bachman (1990), and Bachman and Palmer (2010), validity is a unitary concept relying on a series of theoretical rationales and empirical evidence supporting the appropriateness of an assessment use for a particular purpose. Following this holistic view of validity, a number of studies and analyses were implemented to form an argument for the construct validity of the suggested MDCT. First of all, correlational evidence was collected by comparing the results of the MDCT taken by the 136 participants of the present study and the learners' performance on a written form of the same test. Also, by applying a partial correlation analysis, the performance of the 26 learners in the second group of the study was compared on the present MDCT and a test of pragmatic comprehension utilizing acceptability judgment tasks as its modality, whose validity was previously established in the literature by Bardovi-Harlig and Dörnyei (1998). The partial correlation was used to control for the effects of proficiency based on the students' scores on the placement test in order to make sure that any possible association between the sets of scores is not merely a natural function of their general proficiency level. Furthermore, the performance of the 136 participants was compared to that of the 61 native speakers who voluntarily collaborated in the pilot study as a reference population. These three arguments were used to establish the concurrent criterion relatedness or concurrent validity of the test.

Assessing Pragmatics through MDCTs: A Case...

In a second experimental phase of the validation study, interventionist approach was employed to check the construct validity of the study. In this round of data collection, 26 intermediate learners of English at the ILI were exposed to eight weeks of treatment including pragmatic instruction in the form of role plays, pair works, and meta-pragmatic feedback. The MDCT was administered to all the participants once prior to and once after the treatment as pre-test and post-test. A repeated measures t-test was used to check whether the test was efficient in capturing the anticipated improvement in the students' pragmatic competence. The results from all these studies will be presented in the next section along with a discussion of the analyses.

2. Results

Reliability estimates were calculated for both versions of the test based on their administration to 136 EFL learners. Both tests demonstrated reasonable internal consistency with Cronbach alpha coefficients of .80 and .72 for the DCT and MDCT versions respectively. Furthermore, in order to examine the agreement between the two raters of the DCTs, the inter-rater reliability for 25 random questionnaires scored by both raters based on the criteria explained earlier in the text was estimated. The results showed that the ratings of the two raters fairly agreed with a reliability estimate of $\alpha = .83$.

As the next step, the participants' performance on the MDCT was compared to their pragmatic production results on the earlier DCT version of the test to establish its concurrent validity and justify the use of multiple-choice modality for a test of pragmatics. The results showed that the students' performance on the MDCT was moderately associated with their performance on the DCT with a significant Pearson correlation of $r = .48, n = 136, p < .001$ and a coefficient of determination equal to $r^2 = .23$. Although this correlation

cannot be claimed to be eminently strong, it can be argued that the results demonstrated reasonable correlation between Iranian EFL learners' performance on the MDCT and the same test of pragmatics but with a different modality. Hence, taking into account the practicality issues associated with administration and scoring of DCTs especially with large samples, these results might provide a rationale to employ MDCTs instead of DCTs. As further criterion-related analysis, the performance of these 136 learners of English on this MDCT was compared with that of the 61 native speakers' as a reference group based on an independent samples t-test. The results are presented in Table 1 below.

Table 1. Independent Samples T-Test between the Performances of Natives and Non-Natives

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2- tailed)	Mean Difference	Std. Error Difference
MDCT	Equal var assumed	32.90	.000	18.95	195	.000	37.637	1.985
	Equal var not assumed			25.10	193.68	.000	37.637	1.500

As it is evident in Table 1 above, native speakers outperformed the non-native group in taking the MDCT and performed significantly more successfully than the other group as it was anticipated $t(195)=32.90, p<.001$. Therefore, it can be argued that the test was efficient in discriminating between two groups who are expected to differ significantly in the construct targeted by the test, namely pragmatic competence. This forms another argument in confirming the concurrent validity of the test. As a third measure of the concurrent validity of the test, the performance of the participants on the newly developed MDCT was compared to their performance on a previously validated test of the same construct in the literature by Bardovi-Harlig and

Assessing Pragmatics through MDCTs: A Case...

Dörnyei (1998). For this purpose, both tests were administered to the 26 intermediate learners with a one-month time interval in between and their results were compared using a partial correlation controlling for proficiency. Proficiency was controlled for in order to make sure whether the two tests were associated regardless of the students' proficiency level. The results are presented in Table 2 below.

Table 2. *Partial Correlation between MDCTs & a Validated PT, Controlling for Proficiency*

Control Variables		Acceptability judgment test	
Prof.	Total	Correlation	.54
	MDCT	Significance (2-tailed)	.000
		df	26

As the results suggest, there is still acceptably significant correlation between the two measures of pragmatic competence even after controlling for proficiency; $r=.54$, $n=26$, $p<.001$. Therefore, it can be concluded that learners' performance on this new MDCT was comparable to their performance on a valid test of pragmatic comprehension even after eliminating the effects of proficiency. Comparing the results yielded by this MDCT, these three pieces of evidence seem to be reasonable supports for the concurrent validity of the test.

To accumulate further argument to support the construct validity of the test, an experimental study was designed as explained earlier in this text. The results of the repeated measures t-test are presented in Table 3 below.

Table 3. *One-Sample Test for the Pretest and Posttest of Pragmatic Instruction*

	t	df	Sig. (2-tailed)	Mean Difference
pretest	21.56	25	.000	58.269
posttest	28.52	25	.000	64.423

As the results of the repeated measures t-test in Table 3 indicate, the performance of the learners on this multiple-choice test of pragmatics

significantly improved after a period of pragmatic instruction; $t(25)=21.56$, $p<.001$. This suggests that the MDCT was sensitive to improvements in learners' interlanguage and was capable of capturing these developmental patterns. This can be resorted to as another argument supporting the construct validity of the test.

3. Discussion and Conclusion

The present study asked whether it was possible to reliably and validly evaluate Iranian English learners' pragmatic competence using a multiple-choice discourse completion test. To address this question, the procedures for developing an MDCT were described based on the five speech acts of *apology*, *refusal*, *gratitude*, *request*, and *suggestion*. The MDCT was developed, piloted, and administered and consequently correlational as well as experimental studies were designed to establish its validity as a test of pragmatic competence in Iran. First of all, fairly high reliability estimates of internal consistency were obtained for the test as an essential feature of the test and a prerequisite for its validity according to Bachman (1990). In order to ensure the content validity of the test, in addition to consulting experts in all stages, the procedures followed to develop the MDCT, as explained earlier in detail, were all inspired by successful examples from the literature including Tada (2005), Liu (2006) and Roever (2008).

As the next step, several pieces of correlational evidence were collected to support the concurrent validity of the test. More precisely, it was indicated that the results of the test correlated well with a written or open-ended version of the same DCT as well as the students' performance on a previously validated test of pragmatics with a different modality (acceptability judgment tasks) from the literature even after controlling for the proficiency level of the participants.

Assessing Pragmatics through MDCTs: A Case...

Furthermore, it was shown that the test was capable of discriminating between a sample of native speakers and English learners as they are expected to demonstrate significantly different levels of ability on the construct under investigation. Finally, an experimental study was designed to study whether the suggested test was capable of capturing the development in students' interlanguage over time. The results showed that the performance of the learners on this MDCT improved significantly after an anticipated increase in their pragmatic competence by going through 2 months of pragmatic instruction. This array of evidence suggests that the prepared MDC is reliable and can be used as a valid test of pragmatic competence and comprehension in Iranian EFL contexts despite the criticisms leveled at multiple-choice tests in general and MDCTs in particular.

On the whole, the paper attempted developing a multiple choice test of pragmatics and responded to the challenges posed to the use of multiple-choice items in academic and language testing in general and in assessing pragmatics in particular. Therefore, the paper challenged the criticisms against MDCTs including the ones alleged by Brown (2008), Brown and Ahn (2011), Burton (2005), Bush (2015), Cohen (2010), and Ross and Kasper (2013) among others by proposing a reliable, valid, and practical MDCT to be used especially in large scale pragmatic testing in higher education. The findings, on the other hand, supported the results of and built on a few studies which had managed to make a case for use of MDCTs including Liu (2006), Roever (2008, 2011), and Tada (2005). At the end, it is suggested that this line of research may continue its search for valid and practical tests of pragmatics focusing on other speech acts to be used in an Iranian context, especially applicable to large scale testing situations. Furthermore, it is suggested that this MDCT can be employed in

studies of pragmatics especially to investigate the effects of instruction or feedback which can in turn provide further inspection of its validity.

References

- Allan, D. (1992). *Oxford placement tests*. Oxford: Oxford University Press.
- Bachman, L. (1990). *Fundamental considerations in language testing*. New York: Oxford University Press.
- Bachman, L. & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Bardovi-Harlig, K. (2014). Awareness of meaning of conventional expressions in second-language pragmatics. *Language Awareness, 23*, 41–56.
- Bardovi-Harlig, K. & Dörnyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *TESOL Quarterly, 32*, 233-263.
- Bardovi-Harlig, K., & Hartford, B. S. (1993). Learning the rules of academic talk: Alongitudinal study of pragmatic development. *Studies in Second Language Acquisition, 15*, 279–304.
- Bardovi-Harlig, K., Mossman, S., & Vellenga, H. E. (2014). The effect of instruction on pragmatic routines in academic discussion. *Language Teaching Research, 18*, 1-27.
- Beebe, L. & Cummings, M. (1995). Natural speech act data versus written questionnaire data: How data collection method affects speech act performance. In S. Gass & J. Neu (Eds.), *Speech acts across cultures: Challenges to communication in a second language* (pp. 65-86). Berlin: Mouton de Gruyter.
- Beebe, L., Takahashi, T., & Uliss-Weltz, R. (1990). Pragmatic transfer in ESL refusals. In R. C. Scarcella, E. S. Andersen, & S. D. Krashen (Eds.). *Developing communicative competence in second language* (pp. 55-73). New York: Newbury House.
- Blum-Kulka, S. (1982). Learning how 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*, 29–59.
- Blum-Kulka, S., House, J., & Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies*. Norwood: Albex.

Assessing Pragmatics through MDCTs: A Case...

- Bodman, J., Eisenstein, M.(1988). May God increase your bounty: the expressions of gratitude in English by native and non-native speakers. *Cross Currents* 15 (1), 1–21.
- Brown, J. D. (2001). Six types of pragmatics tests in two different contexts. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 301–325). New York: Cambridge University Press.
- Brown, J. D. (2008). Raters, functions, item types and the dependability of L2 pragmatics tests. In E. Alcón Soler & A. Martínez-Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching and testing* (pp. 224–248). Clevedon: Multilingual Matters.
- Brown, J. D. & Ahn, R.C. (2011). Variables that affect the dependability of L2 pragmatics tests, *Journal of Pragmatics*, 43, 198–217.
- Brown, P., & Levinson, S. D. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Cheng, S. W. (2005). *An exploratory cross-sectional study of interlanguage pragmatic development of expressions of gratitude by Chinese learners of English*. Unpublished PhD dissertation; The University of Iowa.
- Cohen, A. D. (2010). Approaches to assessing pragmatic ability. In N. Ishihara & A. D. Cohen (Eds.), *Teaching and learning pragmatics: Where language and culture meet* (pp. 264-285). Harlow: Pearson Education Limited.
- Cohen, D. A. & Olshtain, E. (1993). The production of speech acts by EFL learner. *TESOL Quarterly*, 27, 33-56.
- Eisenstein, M. & Bodman, J.(1993). Expressing gratitude in American English. In: G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 64–81). New York: Oxford University Press.
- Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.
- Eslami-Rasekh, Z. (2005). Raising the pragmatic awareness of language learners. *ELT Journal*, 59, 199-208.
- Eslami-Rasekh, Z. (2010). Refusals: How to develop appropriate refusal strategies. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical, and methodological issues* (pp. 217-236). Amsterdam: John Benjamins Publishing Co.

- Fujiwara, Y. (2007) Critical language testing on pragmatics tests: Are pragmatics tests really appropriate? *Asian Englishes*, 10(1) 24-42.
- Garcia, P. (2004). Developmental differences in speech act recognition: A pragmatic awareness study. *Language Awareness*, 13, 2, 96-115.
- Golato, A.(2003). Studying compliment responses: a comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics* 21 (1), 90–121.
- Hartford, B. S. & Bardovi-Harlig, K.(1992). Experimental and observational data in the study of interlanguage pragmatics. In: L. F. Bouton & Y. Kachru (Eds.), *Pragmatics and Language Learning, vol. 3. Division of English as an International Language* (pp. 33–52). Urbana-Champaign: University of Illinois.
- Hudson, T., Detmer, E., & Brown, J. D. (1995). Developing prototypic measures of cross-cultural pragmatics (Technical Report 7). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Hymes, D. (1972). On communicative competence. In J. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin Books.
- Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press.
- Ishihara, N. (2007) Web-based curriculum for pragmatics instruction in Japanese as a foreign language: An explicit awareness-raising approach. *Language Awareness*, 16 (1), 21-40.
- Jianda, L. (2006). Assessing EFL Learners' Interlanguage Pragmatic Knowledge: Implications for Testers and Learners. *Reflections on English Language Teaching* 5(1), 1-22.
- Jiang, X. (2006). Suggestions: What should ESL students know? *System*, 34, 36-54.
- Kasper, G., & Blum-Kulka, S. (Eds.). (1993). *Interlanguage pragmatics*. New York: Oxford University Press.
- Kasper, G. & Dahl, M. (1991). Research methods in interlanguage pragmatics. *Studies in Second Language Acquisition*, 12, 215-247.
- Kondo, S. (2010). Apologies: Raising learners' cross-cultural awareness. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical, and methodological issues* (pp.145-162). Amsterdam: John Benjamins Publishing Co.
- Leech, G. (1983). *Principles of pragmatics*. London: Longman.

Assessing Pragmatics through MDCTs: A Case...

- Liu, J. (2006). *Measuring interlanguage pragmatic knowledge of EFL learners*. Frankfurt: Peter Lang.
- Malamed, L. H. (2010). Disagreement: How to disagree agreeably. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical, and methodological issues* (pp. 199-216). Amsterdam: John Benjamins Publishing Co.
- Martínez-Flor, A. (2010). Suggestions: How social norms affect pragmatic behavior. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical, and methodological issues* (pp. 257-274). Amsterdam: John Benjamins Publishing Co.
- Martínez-Flor, A. & Alcon-Soler, A. (2007). Developing pragmatic awareness of suggestions in the EFL classroom: A focus on instructional effects. *CJAL*, 10 (1), 47-76.
- McConachy, T. (2013). Exploring the meta-pragmatic realm in English language teaching. *Language Awareness*, 22 (2), 100-110.
- McNamara, T. F., & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Basil Blackwell.
- Messick, S. A. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp. 13–103). Washington, DC: American Council on Education.
- Nguyen, M. T. T. (2014). Using conversation tasks and retrospective methodology to investigate L2 pragmatics development: the case of EFL criticisms and responses to criticisms. *The Language Learning Journal*, DOI: 10.1080/09571736.2014.908404.
- Niezgoda, K. & Röver, C. (2001). Pragmatic and grammatical awareness: A function of the learning environment? In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 63-80). Cambridge: Cambridge University Press.
- Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing*, 23, 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 A. Martínez-Flor, & E. Alcon (Eds.), *Investigating pragmatics in foreign language learning, teaching, and testing* (pp. 249–266). Clevedon: Multilingual Matters.

- Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language testing*, 28(4) 463-481.
- Rose, K. R. (1997). Pragmatics in teacher education for nonnative-speaking teachers: A consciousness-raising approach, *Language, Culture and Curriculum*, 10(2), 125-138.
- Ross, S. J. & Kasper, G. (Eds.) (2013). *Assessing second language pragmatics*, Hampshire: Palgrave Macmillan.
- Safont-Jordà, M. P. (2004). An analysis on EAP learners' pragmatic production: A focus on request forms. *IBÉRICA*, 8, 23-39.
- Schauer, G. (2004). May you speak louder maybe? Interlanguage pragmatic development in requests. In S. Foster-Cohen & M. Sharwood-Schmith (Eds.), *EUROSLA Yearbook* (Vol. 4, pp. 253-273). Amsterdam: John Benjamins.
- Schauer, G. (2007). Finding the right words in the study abroad context: The development of German learners' use of external modifiers in English. *Intercultural Pragmatics*, 4, 193-220.
- Schauer, G. & Adolphs, S. (2006). Expressions of gratitude in corpus and DCT data: Vocabulary, formulaic sequences, and pedagogy. *System*, 34, 119-134.
- Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan corpus of academic spoken English*. Ann Arbor, MI: The Regents of the University of Michigan. Retrieved from quod.lib.umich.edu/m/micase/
- Tada, M. (2005). *Assessment of EFL pragmatic production and perception using video prompts*. Unpublished doctoral dissertation, Temple University.
- Trosborg, A. (1995). *Interlanguage pragmatics: Requests, complaints, apologies*. *Studies in Anthropological Linguistics* (Vol. 7). Berlin: Mouton de Gruyter.
- Usó-Juan, E. (2010). Requests: A sociopragmatic approach. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical, and methodological issues* (pp. 237-256). Amsterdam: John Benjamins Publishing Co.
- Yuan, Y. (2001). An inquiry into empirical pragmatics data-gathering methods: written DCTs, oral DCTs, field notes, and natural conversation. *Journal of Pragmatics*, 33(2), 271-292.