

Developing a Pragmatics Test for Arabic ESL Learners

Sawako Matsugu

Liberal and General Education Center, Utsunomiya University,
Tochigi, Japan

Abstract

The present study replicated Liu's (2007) study on validating a multiple-choice discourse completion test (MCDCT) for measuring second language (L2) pragmatics. In order to investigate whether his validation methods can be applied to a different speech act with learners from a different linguistic and cultural background, this study was conducted for native speakers of Arabic learning English as a Second Language (ESL) to assess the speech act of refusal. First, an exemplar generation and situation likelihood questionnaire was administered to 15 Arabic participants to elicit refusal situations to be used for constructing a MCDCT. Following this, metapragmatic assessment was carried out in order to determine social variables of status/power, distance, and severity in each of the elicited situations. Then, a situation pilot questionnaire was given to 14 Arabic-speakers and 11 English-speakers to collect responses to be used for answer keys and distracters for the MCDCT. Two native speakers of English rated responses by Arabic-speakers and those marked inappropriate were used as distracters for the instrument, while English-speakers' responses were included as keys. Due to a small sample size and low inter-rater agreement, the resultant item number for the MCDCT became six.

Key words: Arabic learners of English, MCDCT, pragmatics, refusals, tests,

Introduction

Bachman (1990) has suggested that communicative competence consists of two components: organizational competence which refers to grammatical knowledge, and pragmatic competence which refers to the “capacity for implementing, or executing [organizational] competence in appropriate, contextualized communicative language use” (p. 84). This pragmatic competence has been identified as one of the indispensable communicative competences by many researchers (e.g., Bachman, 1990; Canale & Swain, 1980). When this pragmatic competence fails in communication, that is, when the speaker’s utterance is perceived differently by a hearer than how the speaker intended it to be perceived, serious miscommunication or misunderstanding could arise. However, pragmatics has not received vigorous attention in the language classroom compared to other skill areas. In addition, little attention has been given to the assessment of second language (L2) pragmatics (Liu, 2006), or more specifically, assessment of L2 pragmatics in classroom contexts (Ishihara, 2009). Reflecting this, there are no established and widely used pragmatics instruments available (Bachman, 1990; Roever, 2004). Difficulty in the assessment of pragmatics has also been pointed out due to its contextualized nature; there is a tension between the construction of authentic assessment tasks and practicality (McNamara & Roever, 2006; Roever, 2004).

Literature Review

Previous Studies on Assessment of L2 Pragmatics

In order to better assess language learners’ pragmatics competence, researchers have designed at least six instruments described below (Enochs & Yoshitake-Strain, 1996, 1999; Hudson, Detmer, & Brown, 1992; Hudson et al., 1995; Roever, 2005; Yamashita, 1996; and Yoshitake-Strain, 1997). The measures that Hudson et al. (1992; 1995) developed are as follows: (1) self-assessment test (SA), (2) listening laboratory production test (LL), (3) open discourse completion test (OPDCT), (4) multiple choice discourse completion test (MCDCT), (5) role-play self-assessment test (RPSA), and (6) role-play test (RP). These instruments for Japanese learners studying English as a second language (ESL) varied in type, method, and settings for each study, measuring speech acts of requests, refusals, and apologies along with the different socio-cultural variables of power, distance, and imposition.

In developing their MCDCT, the following three points were considered: (a) strategy use, (b) sociopragmatic misjudgments, and (c) phrasing/expressions. Keys were created based on native speakers’ responses, while distracters were from non-native speakers’ responses. Some issues associated with their MCDCT are that no systematic protocol study was conducted, that no agreement data on metapragmatic assessment between native speakers and non-native speakers of English was provided, and that no statistical analyses associated with reliability and validity were carried out (Liu, 2006).

These instruments were statistically analyzed by Yamashita (1996) with translated versions of Hudson et al.’s tests (1992, 1995) with some modifications. Her participants were learners of Japanese studying in a Japanese as a second language (JSL) context. Results showed that although high reliability and validity were found in the five assessment instruments (e.g, SA, LL, OPDCT, and RPSA), the MCDCT had many issues such as low reliability. It was also found that the translated instruments were appropriate to be used for English speakers studying JSL. One limitation with Yamashita’s study (1996) is that a metapragmatic assessment was not conducted to investigate social variables (Liu, 2006).

Enochs and Yoshitake-Strain (1996, 1999) conducted a validation study of Hudson et al. (1992,

1995) using Japanese EFL learners. Without modification, Hudson et al.'s (1992, 1995) six instruments were administered to investigate their reliability and validity. Results revealed that, similar to Yamashita (1996), reliability and validity were confirmed with the instruments except for OPDCT and MCDCT. Similar results were also found in Yoshitake-Strain's study (1997) in which she conducted a qualitative analysis on the four instruments (i.e., OPDCT, MCDCT, LL, and RP) used in Hudson et al.'s study (1992, 1995). Her participants were Japanese learners of English and analyses involved examining pragmatic features and the strategies used by them.

These previous studies have contributed to validating L2 pragmatics instruments. However, according to Hudson (2001), variability of speaker behavior in discourse needs to be considered. This means variability related to the social nature of the speech event and the speaker's strategic and linguistic choices to achieve communicative goals. Yamashita (1996) and Yoshitake-Strain (1997) lack such variability among the informants.

Roever (2005) developed web-based tests of ESL pragmalinguistics using MCDCTs, to investigate recognition of situational routine formulas, comprehension of implicature, and knowledge of speech act strategies. The instruments were pilot-tested several times, including concurrent verbal protocols. Results showed that the degree of imposition and proficiency caused difficulty in the speech act section, similarly, proficiency influenced knowledge of ESL implicature. Exposure to the target culture accounted for knowledge of routines. In addition, strong evidence for the validity of the MCDCT was found.

As shown earlier, previous studies except for Roever (2005) have revealed MCDCT's low reliability in assessing L2 learners' pragmatic competence. In addition, there are other weaknesses associated with related instruments such as that DCTs do not neatly replicate actual speech; some items used in DCTs are not used in real-world conversations (e.g., Golato, 2003). However, the primary advantage of MCDCTs is their high practicality in administration and the ease of scoring (Liu, 2006) unlike other measures such as RP or LL. In various classroom settings where instructors have a limited time in administering and scoring tests, there might be cases when they do not appropriately measure their students' learning outcomes due to impracticality of other pragmatic instruments. Since "the consequences of *not* administering the test" (Roever, 2004) should be taken into consideration, MCDCTs could still be instrumental in assessing L2 learners' pragmatic abilities if developed appropriately. In fact, Liu (2007) was able to design a MCDCT which achieved high reliability and validity to assess Chinese English as a Foreign Language (EFL) learners' knowledge of the speech act of apology by carefully following several steps. First, students in China were asked to identify the situations which require apologies, and to report how likely such situations would occur in their daily lives. Second, metapragmatic assessment that involved both Chinese and American students was carried out to analyze the social variables to be applied in each scenario, followed by validation of the scenarios and designing multiple-choice options for each scenario. Finally, a pilot test was conducted on Chinese university students.

Accordingly, since there is a call for more research as stated by Liu (2006, 2007) to investigate if his method can be applied to other speech acts in different linguistic and cultural contexts, the present study was conducted. The purpose of this study is to validate the instrument to assess the speech act of refusal demonstrated by native speakers of Arabic studying in an intensive English program (IEP) in a university in North America. This study can be considered as a replication of Liu's (2007) study described above with a different native language group of students in an ESL context with a different speech act (i.e., refusal).

The Speech Act of Refusals

The speech act of refusal happens in response to invitations or suggestions (Al-Eryani, 2007). Refusals are often influenced by gender, age, education level, power/status, and social distance (Brown & Levinson, 1987; Mills, 2003). It should be noted that different languages and cultures refuse in a different manner (Al-Eryani, 2007). Because refusal is a face-threatening act to the listener or requester, with aforementioned various factors that need to be considered when appropriately making a refusal, refusals are difficult speech acts (Know, 2004).

In the present study, refusals were selected as the target speech act because there was a call for research with different speech acts (Liu, 2007). Moreover, compared to other speech acts such as apologies or requests, refusals have not been widely studied in the area of language assessment.

Method

Method in Liu's (2007) Study

In his original study, there were five steps to develop the MCDCT, followed by its pilot test. The first five stages consisted of exemplar generation, situation likelihood investigation, metapragmatic assessment, situation pilot study, and development of multiple-choice options. For exemplar generation, topics for the scenarios which required apologies were obtained by means of an exemplar generation questionnaire. This was particularly important as authenticity is of special importance in language testing; in addition, its potential influence on test-takers' performance has been reported (Bachman & Palmer, 1996).

In situation likelihood investigation, likelihood of each of the 46 situations elicited in the exemplar generation questionnaire was investigated on a scale of one to five with five being the most likely and one being the least likely in their daily life. The 30 situations that earned the highest means were chosen for further investigation.

The third step, metapragmatic assessment, was aimed at examining social variables (i.e., social distance, relative power and status, and the severity of a specific apology situation) empirically. This assessment was conducted to capture potential culture-specificity of pragmatic expectations, assessments, and sociopragmatic elements (Blum-Kulka, 1991; Spencer-Oatey, 1993). In addition, a metapragmatic judgment is useful in establishing degrees of equivalence between multiple languages both at the sociocultural and pragmalinguistic levels (Olshtain & Blum-Kulka, 1984). Both Chinese native speakers and English native speakers were recruited for this assessment in which they were asked to identify the variables of power, distance, and severity on a scale of one to five. Since this numerical scale did not produce high agreement among participants, different judgment scales were implemented to achieve higher agreement, followed by statistical analyses to identify significant differences between Chinese and English native speakers.

Following this, a situation pilot study was conducted to obtain preliminary data for alternatives and keys in each MCDCT item. Both Chinese and English native speakers were asked to respond to the situations which were narrowed down in the previous stage. Two trained English-speaking raters evaluated their responses on a five-point scale to determine their appropriateness.

For the development of multiple-choice options, the data collected in the situation pilot study were used to develop alternatives as well as keys for each item. Responses by English-speakers were coded as the "key" and responses by Chinese-speakers categorized as inappropriate were coded as "distractor." Based on these keys and alternatives, a new multiple-choice questionnaire was created for native English speakers to identify appropriateness of each alternative for all items. The same procedure was repeated after refining alternatives and items in the previous

questionnaire.

Procedure in the Present Study

The present study employed the following four steps: an exemplar generation and situation likelihood-combined questionnaire, metapragmatic assessment, a situation pilot study, and development of multiple-choice options. Description and participants involved in each step is described below.

Exemplar generation and situation likelihood-combined questionnaire. For this questionnaire, 15 male native speakers of Arabic completed the instrument to elicit ten situations that they had actually either experienced, or seen, or heard, involving a refusal in their home country (e.g., Saudi Arabia, Kuwait). Due to limited time, situation likelihood investigation was also included in the questionnaire in which participants were asked to mark the likelihood of each situation that they described on a scale of one to five with one being least likely and five being very likely.

The participants were all enrolled in the IEP at a university in the southwestern US. They were recruited for this study in their own classrooms after their regular instruction. Their English proficiency varied; from approximately a 30 in the internet-based Test of English as a Foreign Language (TOEFL) to between 60 and 70. Their length of stay in the US also varied from a few months to several years. For the lowest proficiency group of participants, a bilingual of Arabic and English translated the researcher's instruction to elicit accurate information for the questionnaire. They were also allowed to provide their answers in Arabic; however, only a few did so. The same bilingual graduate student translated their responses into English.

After giving instructions for the questionnaire by providing examples, the questionnaire was distributed and participants were asked to bring it to the researcher at their earliest convenience. This instrument elicited a total of 66 refusal situations as most participants did not provide ten situations.

Metapragmatic assessment. In Liu's (2007) study, a group of Chinese-speakers and another group of English-speakers completed a metapragmatic assessment questionnaire to determine social variables such as relative power and status, social distance, and the severity of a specific apology situation. However, in the present study, the assessment was conducted by the researcher herself only.

In each situation, the people involved were first identified (e.g., a father and a son). Then, social variables of power and status, social distance, and the severity of each refusal situation were examined. The variables were coded with either a lack or an existence of each variable. For power and status, equal status was also identified. Following this, all situations were compared with Liu's situations and ten situations that had similar variable combinations (e.g., higher interlocutor's status, a lack of social distance between the speakers, and a lack of severity) with Liu's were selected. Table 1 summarizes the situation variables after the metapragmatic assessment.

Table 1

The Distribution of the Situation Variables after the Metapragmatic Assessment

Situation Variables	Cell number and attributes									
	1	2	3	4	5	6	7	8	9	10

Power (P)	-	-	-	=	=	=	=	+	+	=
Distance (D)	-	+	-	+	-	+	-	+	-	+
Severity (S)	-	-	+	+	-	-	+	+	-	-

Notes: The cell numbers correspond to item numbers in the situation pilot questionnaire.

Situation pilot study. In order to collect responses to be used as either keys or alternatives for the MCDCT, a questionnaire was developed based on the situations narrowed down in the metapragmatic assessment. Ten situations were listed with the directions and an example (see Appendix A). Participants for this situation pilot study consisted of 14 Arabic-speaking students most of whom also completed the first exemplar generation and situation likelihood questionnaire (N = 14; 13 male, 1 female). In addition, 11 native speakers of English who were all graduate students of Teaching English as a Second Language (TESL) or applied linguistics answered the same questionnaire. The latter group of participants were contacted by e-mail for recruitment and those who agreed to participate in the study returned the questionnaire to the researcher. The procedure and directions were given in the e-mail. Both groups of participants were asked to respond in writing as if they encountered the ten situations. Although most Arabic participants responded to all ten situations, nine responses were not provided out of a possible 140 responses. On the other hand, most English speaking participants provided their answers in all ten situations; however, one participant was unable to respond to one situation.

Following this, two native English speakers who did not complete the questionnaire rated Arabic participants' responses. Before the rating session, training was provided based on Hudson et al.'s (1995) training manual, which was also used in Liu (2007). They were asked to rate their answers on a scale of one to five, one being least appropriate and five being most appropriate. They were reminded to ignore grammatical errors as long as the responses were comprehensible. Moreover, since both raters had extensive experiences in teaching ESL, it was assumed that they had more tolerance for pragmatic errors than average native English speakers without such a background. Therefore, they were instructed not to be overly tolerant when evaluating their responses.

After their grading, inter-rater reliability was computed. The percentage of exact agreement was 32.4 %. This increased to 65.7 % when the percentage of time raters' agreed within one point. Cohen's Kappa was considerably low, at .13.

Development of multiple-choice options. After reviewing the two raters' evaluation of Arabic participants' responses, two to three responses marked inappropriate (i.e., scores of one or two) with exact or close agreement between the raters (i.e., within one point difference) were included as alternatives. Following this, English-speakers' responses were reviewed and similar responses to each item were eliminated. Then, two of their responses were included per item as keys. Grammatical errors in Arabic participants' responses were corrected as much as possible.

Results and Discussion

Exemplar Generation and Situation Likelihood Questionnaire

The exemplar generation and situation likelihood questionnaire which was administered to Arabic participants generated a total of 66 situations that involved a refusal. Table 2 below summarizes the situations elicited by the instrument.

Most frequently occurred refusal situations were between friends, reported 23 times (see

Appendix B for some examples of various refusal situations). The likelihood of the situations involving friends varied, but higher frequency (i.e., frequency of 4 and 5) was indicated more than lower frequency (i.e., frequency of 1 and 2). Refusal situations with family members such as “parents and son,” “father and son,” “mother and son,” “sister and brother,” and “between brothers” were often described, producing a total of 28 situations. Particularly, situations between “mother and son” were reported most with a higher frequency of occurrences with 4s and 5s. Among non-family member relationships, except for the relationship “between friends,” situations involving “girlfriend and boyfriend,” “strangers,” and “classmates” were reported with a few occurrences. Finally, the category “others” included situations involving “government official and ordinary person,” “shop clerk and customer,” “teacher and student,” and “employer and employee.” For these situations, both relatively high and low frequency of occurrences was reported.

Table 2***Results of Exemplar Generation and Situation Likelihood Questionnaire***

Relationship	Occurrences	Likelihood					N/A
		1	2	3	4	5	
Parents-son	2	1	0	0	1	0	0
Father-son	6	1	1	0	2	2	0
Mother-son	10	0	2	0	4	2	2
Sister-brother	3	0	0	0	3	0	0
Between brothers	6	0	2	0	4	0	0
Grandmother-grandson	1	1	0	0	0	0	0
Girlfriend-boyfriend	4	0	0	0	1	1	2
Between strangers	1	1	0	0	0	0	0
Between friends	23	1	7	2	9	5	0
Between classmates	2	0	0	1	0	0	1
Others	8	0	3	0	4	0	2
Total	66	5	15	3	28	10	7

Notes: The numbers below “likelihood” shows the frequency scale. Numbers in each relationship show raw counts of each frequency in each situation.

Overall, the questionnaire seemed to be able to elicit refusal situations. However, some participants with low English proficiency did not understand the directions and provided unrelated responses. Having expected this problem, the aforementioned bilingual graduate student translated the researcher’s instruction into Arabic for this questionnaire for low English proficiency learners. However, this translation service was not provided to other students who had higher English proficiency. How much of this translation influenced refusal situation elicitation is not known; however, it would have been better to provide translation to all participants or not provide translation at all to minimize any extraneous influences on the instrument.

Another issue worth noting is that even though participants were asked to describe refusal situations that they had experienced, seen, or heard in their home country, it is not known whether the situations they described actually happened or not. They could be something they made up by themselves. Also, even if participants had actually experienced those situations, it is

not clear whether they actually happened in their home country or not. Since all participants had lived in the US when this study was conducted, it is possible that they were unintentionally thinking about refusal situations that they had experienced in the US.

Furthermore, the participants were all native speakers of Arabic, but their home country slightly varied; a majority from Saudi Arabia, but a few from Kuwait, Egypt, and Yemen. Since cultural differences would be assumed even among countries that share the same language, the effects of these cultural differences on their responses are possible. Yet, a few native speakers of English who responded to the subsequent questionnaire indicated that some of the situations listed on the instrument were somewhat unfamiliar to them. This suggests that the situations elicited in the first questionnaire should contain at least a few authentic situations unique to Arabic-speaking cultures.

Metapragmatic Assessment

In order to identify situational variables, namely, power/status relationship, distance between the speakers, and severity of each situation, metapragmatic assessment was conducted. Table 3 shows its results. A + in power indicates that the speaker had a higher status than the listener and a – indicates that the speaker had a lower status than the listener. A = means an equal status. For the variable of distance, a + shows the speaker and the listener were familiar with each other, whereas a – shows they were strangers. For the variable of severity, a + means the situation was severe, while a – means the situation was not severe.

Table 3
Summary of Situational Variables

Relationship	Power	Distance	Severity
Parents-son	-	+	-
Father-son	-	+	-
Mother-son	-	+	-
Sister-brother	+/-	+	-
Between brothers	+/-	+	-
Grandmother-grandson	-	+	-
Girlfriend-boyfriend	+	+	-
Between strangers	=	-	-
Between friends	=	+	-
Between classmates	=	+	-
Others	varies	-	+/-

For the power variable, situations involving parents were all given a – as well as a situation with “grandmother and grandson.” Situations with “sister and brother” or “between brothers” were given both a + and a – as it was not known if the sibling was older or younger. Situations with “boyfriend and girlfriend” were given a plus as it seems that men generally have a higher social status than women in Arabic-speaking countries. “Classmates,” “friends,” and “strangers” were all given a = since they are in equal status in most cases.

In regards to distance, since most situations involved people that participants knew well, such as family members, friends, classmates, a + was given to all situations except for “others” and “strangers.”

Finally, most situations were involved with trivial requests which were refused subsequently, such as refusing to go to the market with parents, or to go to lunch with friends. Therefore, a – was given to all situations except for “others.”

Relationships named “others” was identified as “varies” for power and a - for distance as it included a customer and a shop clerk, or a student and his professor, or a government official and an ordinary citizen. For severity, most situations were more severe than other situations with family members or friends. For example, there was a case reported that a government official asked for a bribe to complete a paperwork requested by a citizen who actually refused to give a bribe. Because not providing a bribe could result in not having the paperwork done, this case was considered severe. As mentioned previously, after identifying social variables, their combinations of variables were compared with Liu’s (2007), and ten situations that carried similar variable combinations with Liu’s study were chosen for a situation pilot study (see Table 1 given earlier for more reference).

Metapragmatic assessment is important in developing pragmatics instrument as it is possible that people from different cultures perceive situations differently (Liu, 2006, 2007). Therefore, Liu had two groups of participants for this assessment: a group of native speakers of English and another group of Chinese speakers to examine the degree of equivalence between two languages both at the sociocultural and pragmalinguistic levels (Olshetain & Blum-Kulka, 1984). In the present study, this assessment was conducted by the researcher herself. Since she is not extremely familiar with Arabic-speaking culture, such as how age and gender influence a power relationship, the results of this metapragmatic assessment could be problematic. As Liu did, it would have been better if native speakers of Arabic and English actually identified social variables, especially because the researcher’s native language is not English, and because the major purpose of this assessment was to investigate the degree of equivalence between two languages from sociocultural and pragmalinguistic perspectives. Yet, this metapragmatic assessment had to be conducted to select situations for the following situation pilot study.

Situation Pilot Study

A questionnaire that contained the ten situations chosen as a result of metapragmatic assessment was administered to both native speakers of Arabic and English (see Appendix A for the sample questionnaire). The purpose of this pilot study was to elicit responses to the situations, which would be used as keys and alternatives for the MCDCT. A total of 131 responses were obtained from Arabic participants and 109 responses from American participants. Among Arabic participants, there seemed to be some learners who did not understand the directions probably due to their low English proficiency. Accordingly, some responses were either incomprehensible or unrelated to the situations. The resulting responses from Arabic participants were 102. These responses were evaluated by two native speakers of English who did not provide responses to this pilot study questionnaire.

Generally the questionnaire was able to elicit responses to refusal situations well from both groups of participants. However, some situations seemed to be universal; that is, pragmatic transfer was possible in which even Arabic participants with low English proficiency did not have trouble providing appropriate answers. Therefore, in situations like item number six where a stranger asked for a cigarette, most Arabic participants were able to supply appropriate answers. This prevented obtaining inappropriate responses to be used as alternatives. Furthermore, as indicated earlier, the agreement between the two raters was quite low; which made it difficult to obtain many inappropriate responses as distracters. Both raters were graduate

students of TESL or applied linguistics, but one was female in her late 20s and the other was male in his late 30s. These age and gender differences might have influenced their ratings. The effect of their ESL teaching experience did not seem to influence the results, however, because the male rater who had longer teaching experiences than the female rater was not as lenient as the female rater. Another issue related to the rating was that the training session was not conducted together, but they were given separately, although the training was provided in a very similar manner. If the session had been given together and if the raters had discussed their ratings, higher rater agreement could have been achieved.

Development of Multiple-Choice Options

To develop keys and distracters for the MCDCT, inappropriate responses by Arabic-speakers which had high rater agreement were chosen as alternatives for each item. Responses by English-speakers which had similar lengths as other distracters were selected as keys as much as possible, because longer responses often give a clue to test-takers (Miller, Lin, & Gronlund, 1995). Situations that were not successful in obtaining inappropriate responses were eliminated. The number of alternatives varies due to the distractor availability and the rater agreement rate (see Appendix C for the MCDCT).

Since four situations could not obtain inappropriate responses, remaining six situations were presented in the MCDCT. Most distracters seem to be clearly inappropriate to each given situation. Generally speaking, Arabic participants' responses were shorter than those by English-speaking participants. This made it difficult to control the length of answer keys. Responses marked inappropriate by Arabic learners often lacked explanations or reasons for the refusal and they seemed abrupt. These issues surely need to be addressed for them to gain better pragmatics skills. On the other hand, although there were some varieties of responses among native English speakers, most supplied reasons and an apology for refusing a request.

It is unfortunate that due to a limited number of obtained situations in the first questionnaire and low rater agreement, four situations had to be eliminated from this instrument. More varieties of situations need to be collected in order to obtain and assess a large enough sample of learners' pragmatics competence. This is particularly important to achieve validity (Miller et al., 1995). In addition, raters have to be trained better to achieve a higher agreement rate in order to collect more possible distracters.

Limitations

Several limitations need attention. First of all, the sample size for both instruments was very small. This, particularly for Arabic participants, negatively influenced the number of situations obtained for the first questionnaire, and responses for the second questionnaire. A larger number of participants will be needed for a future study. Their English proficiency level should have been controlled. In order to recruit as many participants as possible, Arabic learners with limited English skills were also included in this study. In order to address this potential problem, translation was provided for some subjects. However, as mentioned previously, since not all participants had access to this service, not only was it not fair to participants, but that lack of translation could have possibly influenced their responses.

Rating for the second instrument suggested a major problem. With the low rater agreement, most responses by Arabic speakers had to be discarded. Liu (2007) also had problems in achieving a high agreement rate; however, by identifying the sources of disagreement, and improving the training, higher agreement needs to be obtained.

Finally, for metapragmatic assessment, it would have been necessary to have groups of Arabic and English speakers to identify social variables and examine the extent to which their perceptions differed for the situations. Even without having a large group of participants of those two languages, at least it would have been more valid to have a few Arabic speakers to check the researcher's interpretation of the variables.

Conclusion

Even though pragmatics has still not gained as much popularity in ESL/EFL classrooms as other areas such as grammar and vocabulary, it is still an important area with which even advanced learners have trouble. Because instruction and assessment cannot be separated, language assessment researchers should continue to strive for developing better instruments to measure L2 pragmatics for better pragmatics instruction for language learners. Liu's (2007) study well reflects the need for constructing and validating such instruments.

Accordingly, further studies to develop instruments for L2 pragmatics should be carried out based on his method. First, it is proposed that native speakers of English, in addition to native speakers of Arabic, will be recruited for the first questionnaire of exemplar generation and situation likelihood. This way, it should help examine any differences in refusal situations between the two cultures. Second, metapragmatic assessment should be conducted more thoroughly by actually having both Arabic and English native speakers identify social variables. Next, with newly obtained data, multiple-choice options should be revised and the number of items should be increased. Following this, a pilot study of the new MCDCT should be carried out accompanied by statistical analyses to examine reliability and validity of the instrument. If a large group of Arabic learners participate in the pilot study, Rasch analyses should be conducted as well.

Many steps are to be implemented to design a valid and reliable L2 pragmatics instrument through the present study. It is hoped that with revision and expansion of Liu's methods, this study will provide a basis for future studies in not only the area of language assessment but also L2 pragmatics.

About the Author:

Sawako Matsugu, a Japanese native, has a PhD in applied linguistics from Northern Arizona University. Her research interest is language testing. She has taught EFL in Japan as well as ESL in an intensive English program in the USA. She is currently an instructor in Utsunomiya University.

References

- Al-Eryani, A. (2007). Refusal strategies by Yemeni EFL learners. *The Asian EFL Journal*, 9, 19-34
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Blum-Kulka, S. (1991). Interlanguage pragmatics: The case of requests. In R. Phillipson, E. Kellerman, L. Selinker, M. Smith, and M. Swain (Eds.), *Foreign/second language*

- pedagogy research: A commemorative volume for Claus Faerch* (pp. 255-272). Clevedon: Multilingual Matters.
- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language use*. Cambridge: Cambridge University Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Enochs, K., & Yoshitake-Strain, S. (1996). Self-assessment and role plays for evaluating appropriateness in speech act realizations. *ICU Language Research Bulletin*, 11, 57-76.
- Enochs, K., & Yoshitake-Strain, S. (1999). Evaluating six measures of EFL learners' pragmatic competence. *JALT Journal*, 21, 29-50.
- Golato, A. (2003). Studying compliment responses. A comparison of DCTs and naturally occurring talk. *Applied Linguistics*, 24, 90-121.
- Hudson, T. (2001). Indicators for pragmatic instruction: Some quantitative tools. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in Language Teaching*. Cambridge: Cambridge University Press.
- Hudson, T., Detmer, E., & Brown, J. D. (1992). *A framework for testing cross-cultural pragmatics*. Honolulu: University of Hawaii Press.
- Hudson, T., Detmer, E., & Brown, J. D. (1995). *Developing prototypic measures of cross-cultural pragmatics*. Honolulu: University of Hawaii Press.
- Ishihara, N. (2009). Teacher-based assessment for foreign language pragmatics. *TESOL Quarterly*, 43, 445-470.
- Know, J. (2004). Expressing refusals in Korean and in American English. *Multilingua*, 23, 339-364.
- Liu, J. (2006). *Measuring interlanguage pragmatic knowledge of EFL learners*. Frankfurt am Main: Peter Lang.
- Liu, J. (2007). Developing a pragmatics test for Chinese EFL learners. *Language Testing*, 24, 391-415.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell Publishing.
- Miller, D. M., Linn, R. L., & Gronlund, N. E. (1995). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Pearson Education.
- Mills, S. (2003). *Gender and politeness*. Cambridge: Cambridge University Press.
- Olshtain, E., & Blum-Kulka, S. (1984). Cross-linguistic speech act studies: Theoretical and empirical issues. In L. MacMathuna, D. Singleton & J. Svartuik (Eds.), *Language Across Culture* (pp. 235-248). Dublin: IRALL.
- Roever, C. (2004). Difficulty and practicality in tests of interlanguage pragmatics. In D. Boxer & A. D. Cohen (Eds.), *Studying speaking to inform second language learning* (pp. 283-301). Clevedon, England: Multilingual Matters.
- Roever, C. (2005). *Testing ESL pragmatics: Development and validation of a Web-based assessment battery*. Frankfurt: Peter Lang.
- Spencer-Oatey, H. (1993). Conceptions of social relations and pragmatics research. *Journal of Pragmatics*, 20, 27-47.
- Yamashita, S. (1996). *Comparing six cross-cultural pragmatics measures*. Unpublished doctoral dissertation, Temple University, Philadelphia.
- Yoshitake-Strain, S. (1997). *Interlanguage competence of Japanese students of English: A multi-test framework evaluation*. Unpublished doctoral dissertation, Columbia Pacific

University, California.

**Appendix A
Questionnaire 2**

Name: _____ Level: _____ Your country: _____

Directions: Please read the following example first. Then read the following situations carefully. Next, please respond to each situation in detail. That is, what would you say if you were in the following situations? Please pay special attention to each situation and who you are talking to (ex. Is he/she somebody you need to show respect or not?).

Situation: My mother asked me to clean the living room. But I was too busy and I didn't want to do it.

My mother: "*Sawako, can you clean the living room?*"

My response: *I have other things to do. So I can't.*

- 1) Situation: A shop clerk tries to sell his product very persistently, but you don't want to buy it. What would you tell the clerk?

Your response: _____

- 2) Situation: Your father asked you to get some groceries from the market, but you don't want to. What would you tell your father?

Your response: _____

- 3) Situation: In a university class, you were asked to leave the class because of your bad behavior (e.g., sleeping in class, chatting with your classmates), but you don't want to. What would you tell the professor?

Your response: _____

- 4) Situation: Your best friend asked you for a lot of money, but you don't want to. What would you tell him?

Your response: _____

- 5) Situation: A stranger asked you for a cigarette, but you don't want to give one. What would you tell the person?

Your response: _____

- 6) Situation: Your friend asked you to have lunch with him, but you have to take your father to the market. So you cannot go to lunch with your friend. What would you tell your friend?

Your response: _____

- 7) Situation: Your co-worker (= colleague) asked you to cover for him because he is sick. But you are not close to the person and you don't want to cover for him. What would you tell your co-worker?

Your response: _____

- 8) Situation: Your little brother asked you to buy an expensive toy for him, but you don't want to. What would you tell him?

Your response: _____

- 9) Situation: You work at a grocery store and your customer called you to deliver some food to his home, but you can't because you are too busy. What would you tell the customer?

Your response: _____

- 10) Situation: Your friend asked you to help with his homework, but you have an appointment with your dentist, so you can't help him. What would you tell him?

Your response: _____

Appendix B
Refusal Situations

Likelihood	Relationship	Situation
4	Between brothers	A brother asked another brother to do the work that he was supposed to do by himself. The task that was originally asked by their father.
2	Government officials-ordinary person	Somebody needed a paper work at a government office and was asked to give money as a bribe to a government official.
2	Student-professor	In a university class, a student was asked to leave because of his bad behavior, but the student refused and university guards had to take him out from the classroom.
4	Shop clerk-customer	A persistent clerk tries to sell products to his/her customer. But the customer refused.

3	Acquaintance-friend	A person needed to sell his car because he was in need of money. But his friends tried to take an advantage of the situation and offered less money than the car's actual value. The seller refused.
2	Father-son	A father asked his son to bring some groceries from the supermarket, but the son refused.
3	Between friends	A friend invited another friend to lunch, but he refused.
4	Sister-brother	A sister asked her brother to go with her to clothing shops, but he refused.
2	Mother-son	A mother asked her son to go with her to see his grandmother but he refused.
1	Grandmother-grandson	A grandmother asked her grandson to go with her to see his mother but he refused.
5	Mother-son	A mother asked her son to take her to the market.
?	Girlfriend-boyfriend	A girlfriend asked her boyfriend to have breakfast with her.

Appendix C
MCDCT

Note: Alternatives in italics are responses from native speakers of English.

1) Situation: A shop clerk tries to sell his product very persistently, but you don't want to buy it.

What would you tell the clerk?

- a) *No, thank you, I'm not interested.*
- b) I have something similar.
- c) I'm sorry, I don't want to buy anything. I am really sorry.
- d) Sorry, I don't have time. I have to go to my house.
- e) *No.. it looks great, but I'm not going to get anything today.*

2) Situation: Your father asked you to get some groceries from the market, but you don't want to.

What would you tell your father?

- a) You should go.
- b) *Can I do it later?*
- c) I don't have enough money. Sorry.
- d) *Ok, but I'm really busy today.*

3) Situation: Your best friend asked you for a lot of money, but you don't want to. What would you tell him?

- a) I don't have money right now.
- b) *I'd love to help you out, but unfortunately, I can't lend you the money.*
- c) *I really don't have it to spare.*
- d) Now, I can't give you.

4) Situation: Your friend asked you to have lunch with him, but you have to take your father to the market. So you cannot go to lunch with your friend. What would you tell your friend?

- a) *I wish I could but I have to help my dad.*
- b) *That would be fun, but I promised to take dad shopping. How about tomorrow?*
- c) What about next day?
- d) Sorry, I don't have time.

5) Situation: You work at a grocery store and your customer called you to deliver some food to his home, but you can't because you are too busy. What would you tell the customer?

- a) Sorry because we don't have a car.
- b) *I'm sorry. We are too busy right now, and there is a long delay on food delivery.*
- c) Sorry, I don't have time.
- d) *I'm sorry sir. We aren't able to bring groceries directly to our customers. We just don't have the staff for that kind of service.*

6) Situation: Your friend asked you to help with his homework, but you have an appointment with your dentist, so you can't help him. What would you tell him?

- a) Talk to somebody else. I have an appointment.
- b) Sorry, I don't have to do it.
- c) *I can't right now because I have to go to the dentist, but what about later today?*
- d) *I'm sorry, but I have a dentist appointment then. Maybe a different time?*