



Developing an Interlanguage Pragmatic Competence Test on Routines in a Chinese EFL Context

Lan Xu

Suranaree University of Technology, Thailand
xulanyu80@sina.com

Anchalee Wannaruk

Suranaree University of Technology, Thailand
wannaruk@sut.ac.th

Abstract

Performing routines in interlanguage is vitally important for EFL learners since it can cause embarrassment between speakers from different cultures. The present study aims to 1) investigate the reliability and validity of an interlanguage pragmatic competence test on routines in a Chinese EFL context with multiple choice discourse completion task (MDCT), 2) examine the statistical characteristics of the test, and 3) explore the test takers' strategies used in the process of taking the test. Altogether 390 students from China took part in the study. The data were collected with the interlanguage pragmatic competence test on routines and think aloud protocol. Data analyses methods included the split-half method, item analysis theory, descriptive statistics, an independent T-test and content analysis. The results indicate that 1) the interlanguage pragmatic competence test on routines showed strong reliability and validity in the Chinese EFL context, 2) interlanguage pragmatic routines were not easy for the Chinese EFL learners, and situational routines were easier than functional routines, and 3) three strategies: repeated reading, interpreting or translating, and weighing options or justifying responses, were used in the process of taking the test by the Chinese EFL learners.

Key words: Pragmatic Routines; MDCT; Reliability; Validity; Chinese EFL Learners

Introduction

Interlanguage pragmatics investigates L2 learners' developing knowledge and ability in the use of pragmatic rules and conventions as well as practices of the target language (Kasper, 1998). Interlanguage pragmatic routines, as a component of interlanguage pragmatic knowledge, are social conventions used by second or foreign language learners with the target language in particular speech communities. In the present study, interlanguage pragmatic routines refer to the customized use of English by Chinese EFL learners in conversations. Based on whether they are tightly or loosely related to situations, pragmatic routines are known as situational routines or functional routines. In situational routines, many context factors are fixed. The occurrence of situational routines should satisfy the context requirements (Rover, 2014). For example, giving responses to congratulations and serving a customer in a shop belong to situational routines. In contrast, the occurrence of functional routines allows flexible variations in context conditions. Functional routines serve as a part of speech acts. In the present study, the speech acts of inquiry, greeting, comfort, and

refusal were included in the items of functional routines. In an EFL context, interlanguage pragmatic routines constitute a substantial part of interlanguage pragmatic knowledge and they are vitally important in the verbal handling of everyday life. EFL learners need to acquire a sizable repertoire of routines to efficiently cope with recurrent and expanding social situations and discourse requirements in the target language (Coulmas, 1981).

Interlanguage pragmatic routines are an indispensable part of EFL learners' language proficiency (Bachman, 1990). They have been included in EFL teaching syllabi for both middle schools and universities in China (Liu, 2012). However, the corresponding tests on English still focus on linguistic competence aspects, such as vocabulary, grammar, syntax, and cohesion (Liu, 2012). The gap between the teaching syllabi and the tests is mainly due to a lack of generally accepted tests to assess EFL learners' L2 pragmatic competence (Liu, 2012). Thus, the development of reliable and valid tests has become an urgent issue in measuring the ability to perform English pragmatic routines for EFL learners.

An interlanguage pragmatic competence test is a procedure or task that leads second or foreign language learners to process sequences of elements in the target language, with the consideration of the normal contextual constraints (Oller, 1979). Interlanguage pragmatic routines, as a substantial part of interlanguage pragmatic competence, can be tested by various methods, ranging from controlled written tasks to free production. Up to now, six testing methods in interlanguage pragmatic competence have been found, which are: written discourse completion task, multiple choice discourse completion task, oral discourse completion task, role play, self-assessment and role play self-assessment.

The present study focuses on multiple choice discourse completion task (MDCT). MDCT is an instrument which requires learners to read a written description of a situation and select what would be best to say in that situation. MDCT is easy to conduct with a large sample and it is objectively scored. MDCT has been used as an instrument for interlanguage pragmatic competence for the past 20 years (Nemati, Rezaee & Mahdi, 2014; Rose, 1994b; Rose & Ono, 1995; Rover, 2006, 2014; Tabatabaei, & Farnia, 2015; Yamashita, 1996a, 1996b). In China, the multiple-choice (MC) format is the main testing method applied to tests for English, including the Test for English majors Band 4 and Band 8, College English Test Band 4 and Band 6. In these tests, the number of participants is larger than eight millions (Liu, 2012), so Chinese EFL learners are familiar with this testing format. In addition, compared with other fields of interlanguage pragmatic knowledge, routines are less flexible and allow less individual production. For all the above reasons, MDCT should be a suitable method to test pragmatic routines in the Chinese EFL context. Although MDCT is less complicated in scoring, the design of it is highly complex. When MDCT is applied in interlanguage pragmatics, the distractors are not usually incorrect. The selection of the answer is judged by the degree of appropriateness in specific situations for the target language's culture. All the options in the MDCT except the most appropriate one must be clearly inappropriate in the target language culture beyond reasonable doubt (Bardovi-Harlig, 2015).

Although MDCT has been used as an interlanguage pragmatic instrument, the use of it for interlanguage pragmatic routines has been very limited. To ensure the results of a test, reliability and validity are the most important factors to take into consideration. In previous studies, the reliability of MDCT on interlanguage pragmatic routines has been analyzed by Rover (2014). Rover's test was web-based.



He assessed English pragmatic competence in performing routines with 267 EFL learners, including native speakers of American English and non-native speakers from 16 different countries. Rover (2014) obtained an overall Cronbach alpha reliability of 0.73. Rover (1996) also found that situational routines were more difficult than functional routines after a test with 117 EFL learners in Germany. The test takers in Rover's (1996) study considered that functional routines were more flexible, thus, they could perform them with their own experience or even imagination. With regard to the validity of MDCT, some researchers found that MDCT is a valid measure in testing interlanguage pragmatic competence, such as Liu (2007). However, Rose & Ono (1995) found that DCT was not valid in an interlanguage pragmatic competence test with Japanese EFL learners. In Rose and Ono's (1995) study, MC (now MDCT) was excluded from the general DCT.

In addition, the think aloud protocol was applied to explore the strategies the test takers used in the process of taking the test. It is a method used to gather data in productive design and development. In the present study, strategies refer to any mental action applied consciously or unconsciously with the intention of facilitating the process of completing the test in the interlanguage pragmatic competence test on routines. No previous studies have been found to use this method in exploring the strategies that the test takers used in the process of finishing an interlanguage pragmatic competence test.

Up to the present, studies with interlanguage competence testing on pragmatic routines are very limited (Li & Zou, 2015). No study of pragmatic competence test on routines has been found in a Chinese EFL context. The present study aims to develop a reliable and valid test in this field with Chinese EFL learners and to explore the strategies used by the EFL learners in the process of completing the test. It is hoped that the test will help EFL teachers to evaluate learners' ability to perform L2 pragmatic routines. It is also hoped that the test will help EFL learners' to find effective ways to improve their ability to perform L2 pragmatic routines. The following research questions were formulated.

- 1) What are the reliability and validity of MDCT in testing EFL learners' ability to perform interlanguage pragmatic routines?
- 2) What are the statistical characteristics of the test?
- 3) What strategies do the test takers use when taking the test?

Research Methodology

Participants

Three hundred and ninety English majors in China participated in the interlanguage pragmatic competence test on routines of the present study. The number of participants was decided by both the calculation using the formula "Necessary Sample Size = $(Z\text{-score})^2 * SD * (1 - SD) / (\text{margin of error})$ " (Smith, 2013) and the present situation in the Guizhou Province. In this formula, the Z-score relates to the confidence level. In the present study, a 95% confidence interval was used with the corresponding Z-score of 1.96. For the standard deviation, the safe decision is to use .5 which is the most acceptable number to ensure that the sample will be large enough (Smith, 2013). Margin of error determines how much higher or lower than the population mean that researchers want to let the sample mean fall (Smith, 2013). The present researchers used 5% for margin of error, which is the most acceptable percentage. After calculation with this formula, the sample size of the study should be around 385. As the present study was a part of a larger one, taking all the factors into consideration, the English majors from the second academic year were judged to be suitable for the research.

Four universities were selected according to the convenient sampling method. As the total number of English majors in the second academic year from the four universities in the Guizhou Province was 390, which was very close to the results of the above calculation, all of them were included. Four Chinese university teachers of English and three American university teachers of English were also invited to design the MDCT. In addition, 24 from the 390 participants were selected randomly for the think aloud protocol. The 24 students were selected because this number of students had made the data saturating, so no more students were invited to participate for this data collection method.

Research Instrument

The research instruments in the present study included the interlanguage pragmatic competence test on routines and the think aloud protocol. The test was used to examine the reliability and validity of the MDCT. The difficulty level of the test was also investigated in a Chinese EFL context. The think aloud protocol was used to discover the test takers' strategy use in the process of taking the test. The items in the MDCT covered both situational and functional routines. The MDCT items were developed through two methods: items adapted from previous studies (Jiang, 2009; Ma, 2010; Rover, 2014) and items developed by the researchers. The development of the items underwent the following five stages: adaptation from previous studies, development of new situations, development of options for the new situations, a content validity check, and a pilot study.

Stage 1. Adaptation from previous studies

Some MDCT items were adapted from previous research. Altogether 31 items were collected. In Jiang's (2009) and Ma's (2010) study, the items were designed for the Chinese EFL context. However, in Rover's (2014) study, the items were designed for a multi-cultural EFL context. Of the 31 items, only the items adapted from Rover (2014) were examined for reliability and validity, but the other items were not checked for reliability and validity. Thus, it was necessary to check the reliability and validity for the adapted items for the context of the present study.

Stage 2. Development of new situations

The development of new items followed two steps: 1) Find situations from textbooks (He, 2011; Jack & David, 2012; Jack, Jonathan & Susan, 2005; Michael, Jeanne & Helen, 2005). 2) Check the situations with both Chinese teachers and American teachers. The present study focused on interlanguage pragmatic routines, so the situations should occur both in the Chinese culture and American culture. To avoid the slight differences in appropriate use of English by native speakers from different cultures, the use according to the American culture was taken as the norm in the present study.

Stage 3. Development of options for the new situations

To develop the options for the new situations, the following steps were covered: 1) The responses in the English textbooks of the selected routine situations were set as the key for each item. 2) Based on the experience, both the Chinese teachers and the American teachers decided the distractors according to the Chinese students' typical inappropriate language use in performing routines with English. If there were any disagreement between the Chinese teachers and American teachers



with respect to their cultures, the American teachers' decision on the inappropriateness of distractors was considered as more authoritative since they were native speakers. 3) The degree of inappropriateness of the distractors was checked with a questionnaire. In this questionnaire, all the items and the distractors were listed with a five-point rating scale, ranging from the most inappropriate (five points) to the least inappropriate (one point). Thirty-three American students from a natural class in the Confucius College of Guizhou University were invited to do the questionnaire. As distractors in the MDCT, the options should be clearly inappropriate. The distractors with a mean score of over three points (including three points) were considered as an acceptable degree of inappropriateness. After calculation, all the distractors were found to have a mean score higher than three points, except for five distractors. These five distractors were revised or rewritten before the 33 native speakers were invited again to decide their degree of inappropriateness. Until the degree of inappropriateness of all the distractors was accepted, the revising or rewriting stopped. The students' responses to the routine situations were not considered as the means to collect distractors, because when they were used as distractors, they were not clearly inappropriate (Hudson et al. 1995). As a result, 12 items were developed, six for situational routines and six for functional routines.

Stage 4. Content validity check

All the teachers were invited again to check the content validity of the MDCT. The purposes of doing this were 1) to delete the items which did not test knowledge of pragmatic routines, 2) to delete the items which were repetitive or similar, 3) to delete the items which were ambiguous for the readers, 4) to delete the items for which the keys were argumentative, and 5) to examine and correct the grammar and spelling of the items. After discussion, 20 items were selected for the pilot study. Among them, eight items were adapted and 12 were developed by the teachers, and then ten were used for situational routines and ten for functional routines.

Stage 5. Pilot Study

Sixty participants who were excluded from the 390 in the main study took part in the pilot study. In the pilot study, the items which were of an unacceptable difficulty level or discrimination power were rewritten or replaced. In addition, all the distractors were examined, and distractors which were found to have unacceptable discrimination power were rewritten or replaced. In the pilot study, seven items were discarded and replaced because they had too low difficulty level and discrimination power. Two distractors were unacceptable, so those distractors were replaced. After all the revisions, the MDCT was tried again, and all the items and their distractors were found to be acceptable. The reliability for the MDCT was .81 (>.70).

As a result, 20 items were kept for the main study. Among them, nine were for situational routines and 11 were for functional routines. Six items were adapted from previous studies and 14 items were newly developed for the present study.

Data Collection

The participants were given 30 minutes to finish the MDCT in the classroom. No discussion was allowed. The most appropriate option for each MDCT item was supposed to be selected. If the students chose the most appropriate option, they would get one point for each item, and if not, they would get "0". Immediately after the test,

the think aloud protocol was conducted to collect the qualitative data, i.e., to explore the strategies the students used in the process of taking the test.

Data Analysis

In order to answer the first research question, the data collected from the MDCT were put into SPSS (21.0) and calculated using the split-half method for reliability. The validity of the MDCT was calculated with item analysis software and Delta Sigma software. To answer the second research question, descriptive statistics and the independent T-test were applied. Research question three was answered with qualitative data collected from the think aloud protocol, using content analysis.

Results

Research Question 1: What are the reliability and validity of MDCT in testing EFL learners' ability to perform interlanguage pragmatic routines?

In the present study, the reliability of the MDCT was .89. The 'third technique' proposed by Madsen (1983) was employed to carry out the item analysis. The participants were divided into three groups based on their scores in the interlanguage pragmatic competence test on routines. The top 27% were in the high-level group, the middle 36% were in the medium-level group, and the bottom 27% were in the low-level group. The difficulty level and discrimination power of each item, together with the discrimination power of each distractor, were calculated using the Delta Sigma software. Table 1 presents the difficulty level and discrimination power of each item and each distractor.

Table 1: Results of Item Analysis for MDCT

Item	Choice	High	Medium	Low	P	r	Conclusion (Classical test theory)	Remark
1	A	14	28	27	.195	.124	D=low	Acceptable
	B	10	29	29	.186	.181	D=low	
	*C	73	80	36	.485	.352	---	
	D	12	22	30	.200	.171	D=low	
2	A	8	18	22	.143	.133	D=low	Acceptable
	B	10	23	17	.129	.067	D=low	
	C	5	14	17	.105	.114	D=low	
	*D	100	156	0	.656	.952	---	
3	A	5	27	33	.181	.267	D=fair	Acceptable
	B	10	22	28	.181	.171	D=low	
	C	8	12	13	.100	.048	D=low	
	*D	97	128	6	.592	.867	---	
4	A	10	44	47	.271	.352	D=good	Acceptable
	B	20	26	32	.248	.114	D=low	
	*C	45	39	23	.274	.209	---	
	D	18	39	45	.300	.257	D=fair	
5	A	10	20	18	.133	.076	D=low	Acceptable
	*B	105	153	1	.664	.990	---	
	C	8	17	22	.143	.133	D=low	



Item	Choice	High	Medium	Low	P	r	Conclusion (Classical test theory)	Remark
	D	7	16	13	.095	.057	D=low	
6	*A	100	147	1	.636	.943	---	Acceptable
	B	7	19	21	.133	.133	D=low	
	C	11	21	24	.167	.124	D=low	
	D	6	22	30	.171	.229	D=fair	
7	A	9	36	44	.252	.333	D=good	Acceptable
	B	15	29	30	.214	.143	D=low	
	C	13	39	31	.210	.171	D=low	
	*D	54	73	17	.369	.352	---	
8	A	8	15	17	.119	.086	D=low	Acceptable
	*B	102	155	2	.664	.952	---	
	C	10	19	17	.129	.067	D=low	
	D	6	19	20	.124	.133	D=low	
9	A	8	22	27	.167	.181	D=low	Acceptable
	B	9	28	36	.214	.257	D=fair	
	C	5	10	14	.090	.086	D=low	
	*D	94	126	11	.592	.790	---	
10	A	8	27	36	.210	.267	D=fair	Acceptable
	B	11	22	27	.181	.152	D=low	
	C	12	30	27	.186	.143	D=low	
	*D	80	81	29	.487	.486	---	
11	A	9	23	26	.167	.162	D=low	Acceptable
	*B	84	85	24	.495	.571	---	
	C	17	27	26	.205	.086	D=low	
	D	12	27	30	.200	.171	D=low	
12	A	5	30	42	.224	.352	D=good	Acceptable
	*B	58	52	31	.362	.257	---	
	C	17	28	32	.233	.143	D=low	
	D	11	41	43	.257	.305	D=good	
13	A	10	27	26	.171	.152	D=low	Acceptable
	*B	77	86	32	.500	.429	---	
	C	9	31	29	.181	.190	D=low	
	D	13	27	23	.171	.095	D=low	
14	A	17	30	33	.238	.152	D=low	Acceptable
	B	11	35	32	.205	.200	D=fair	
	*C	62	61	36	.408	.248	---	
	D	14	30	29	.205	.143	D=low	
15	*A	81	111	19	.541	.590	---	Acceptable
	B	8	26	25	.157	.162	D=low	
	C	11	31	37	.229	.248	D=fair	
	D	9	13	19	.133	.095	D=low	
16	A	13	21	24	.176	.105	D=low	Acceptable
	*B	88	96	17	.515	.676	---	

Item	Choice	High	Medium	Low	P	r	Conclusion (Classical test theory)	Remark
	C	8	23	28	.171	.190	D=low	
	D	10	27	35	.214	.238	D=fair	
17	A	8	19	17	.119	.086	D=low	Acceptable
	*B	105	153	1	.664	.990	---	
	C	11	21	19	.143	.076	D=low	
	D	6	14	16	.105	.095	D=low	
18	*A	31	45	6	.210	.238	---	Acceptable
	B	24	47	44	.324	.190	D=low	
	C	21	32	43	.305	.210	D=fair	
	D	27	30	40	.319	.124	D=low	
19	*A	82	84	19	.474	.600	---	Acceptable
	B	13	27	25	.181	.114	D=low	
	C	8	19	30	.181	.210	D=fair	
	D	20	33	30	.238	.095	D=low	
20	A	7	17	18	.119	.105	D=low	Acceptable
	B	11	20	18	.138	.067	D=low	
	C	10	15	16	.124	.057	D=low	
	*D	102	156	0	.662	.971	---	

Note: The best choice is marked with an “*”.

“P” is the difficulty level index. “r” is the discrimination power index. “D” is the discrimination power.

In the above table, all the items were of an acceptable difficulty level, and the difficulty level of the items ranged from .210 to .664. The higher the index, the more difficult the items were. The discrimination power for all the items was acceptable, and the indices ranged from .209 to .990. The higher the index, the greater the discrimination power the items had. According to item analysis theory, distractors are acceptable when they are either good, fair or low in discrimination power, but not acceptable when they are with no discrimination power. All the distractors had an acceptable level of discrimination power. Among them, four distractors had good discrimination power, ten had fair discrimination power, and 46 had low discrimination power.

After calculating the difficulty level and discrimination power of each item and distractor, the distribution of the MDCT items according to the difficulty level and discrimination power was presented in Table 2.

Table 2. Distribution of Items in MDCT

Items that should be kept	Items in Routines	Total
1. P=easy D=good	2, 5, 6, 8,17,20	6
2. P=medium D=good	1,3,9,10,11,13,15,16,19	9
3. P=medium D=fair	14	1



Items that should be kept	Items in Routines	Total
4. P=difficult D=good	7	1
5. P=difficult D=fair	4,12,18	3

Note: "P" refers to the difficulty level. "D" refers to the discrimination power.

From Table 2, it can be seen that ten items had medium difficulty level. Six items were easy and four items were difficult. Among them, four items had fair discrimination power and 16 items had good discrimination power.

Research Question 2: What are the statistical characteristics of the test?

The mean score of the 20 items was .51 with a standard deviation of .25. Rather than describing the general picture of the items in the interlanguage pragmatic routines, the mean score and the standard deviation on each item are provided in Table 3.

Table 3: Item Characteristics

Item	Interlanguage competence test on pragmatic routines	
	Mean Score	Std. Deviation
Item 2	.66	.48
Item 5	.66	.47
Item 8	.66	.48
Item 17	.66	.47
Item 20	.66	.47
Item 6	.64	.48
Item 3	.59	.49
Item 9	.59	.50
Item 15	.54	.50
Item 16	.52	.50
Item 11	.50	.50
Item 13	.50	.50
Item 10	.49	.50
Item 1	.48	.50
Item 19	.47	.50
Item 14	.41	.49
Item 7	.37	.48
Item 12	.36	.48
Item 4	.28	.45
Item 18	.21	.41

From the above table, it can be seen that the highest mean score for the items was .66 and the lowest mean score was .21. Among all the items, Item 1, Item 2, Item 3, Item 5, Item 6, Item 9, Item 11, Item 16, and Item 17 are situational routines, and the remaining items are functional routines. In Table 3, the items are arranged from the highest mean score to the lowest mean score. To provide a clearer picture of how Chinese EFL learners performed in English pragmatic routines, Item 5, on which the largest number of the test takers obtained the most appropriate option, is taken as an example.

Betty is a foreign student in China. She has met Peng.

Betty: “I was told that you won the 100-meter race in a national sports meeting last week. Congratulations!”

Peng would probably say: “_____”

- a) Just lucky.
- b) Thank you.
- c) I can't say I did well.
- d) Oh, no.

This item is a situational routine, and the test takers were required to give a response to “congratulations”. Sixty-six percent of the test takers chose the most appropriate answer. It was because the test takers had learnt this English pragmatic routine in class or from textbooks. However, the number of test takers who chose the three distractors was almost equally balanced. The characteristic of these distractors is that they show the influence of Chinese culture. Influenced by Taoism, Chinese people normally respond to “congratulations” in a modest way. Hence, all the distractors were successfully designed. Chinese culture resulted in the test takers showing negative pragmatic transfer in conducting this routine in English.

To compare which type of pragmatic routines was more difficult, the mean scores of the two types of routines were compared with the independent T-test. After calculation, it was found that the functional routines were more difficult than the situational routines and a significant difference existed in the two groups at a level of $p < .01$ significance. The following table provides the mean scores, the standard deviations and the significance level of the two types of routines.

Table 4: Comparison of Situational Routines and Functional Routines

	Mean	S.D.	Sig. Level	Variation Pattern
Situational Routines (9 items)	.58	.32	P<.01	Situational routines> Functional routines
Functional Routines (11items)	.45	.21		

Research Question 3: What strategies do the test takers use when taking the test?

In the think aloud protocol which was conducted immediately after the test, the test takers reported their task processing strategies. The think aloud protocol also made its contribution to the validation process of the interlanguage pragmatic competence test on routines. The data collected from the think aloud protocol were recorded verbally and then analyzed by a content analysis. Three Chinese teachers of English in Guizhou University were invited to analyze the qualitative data collected from think aloud protocol with content analysis method. Generally speaking, three strategies were reported: 1) read an item or parts of an item repeatedly, 2) interpret or translate an item, and 3) weigh options and justify responses.

Repeated Reading

Among the 24 test takers who did the think aloud protocol, 19 of them reported that they read an item or parts of an item repeatedly. The items which were reread the most often were Item 1, Item 3, Item 4, Item 10, Item 12 and Item 13. The students also revealed two reasons for rereading, one was that the options were too distracting, such as item 1, item 4 and item 10, and the other was that the descriptions for the



items were too long and with compound structures, such as Item 3, Item 12 and Item 13. For example, S3 (student 3) mentioned “*sometimes I have to read the options several times to find the most familiar one*”. S5 said “*I think my English is not good, I have to read the description of some items repeatedly before I understand the meaning*”.

Interpretation or Translation

The students used the strategy of interpretation and translation frequently. Some of them (S2, S6, S10, S18, S24) reported that they used this strategy for every item. S2 said “*I interpret every item, including both the description of items and the options*”. S11 mentioned “*For some items, the description is very long, so I always paraphrase it to make the meaning clear before I make a choice*”. For example, for item 13, S22 described her process to interpret. “*...I barely passed. I made a hopeless mess of it. I don't know why I did so badly...*”, and S22 interpreted it as “*I didn't pass, I did it badly and I don't know why*”. S8, S13 and S16 mentioned “*I translate everything I read into Chinese before I made a choice*”. The students revealed their limited vocabulary and low comprehension of understanding complex sentences as their main reasons for using the strategy of interpretation and translation when taking the test.

Weighing Options or Justifying Responses

Lacking NSs' intuitions, the test takers did not have a spontaneous response for they read the options. Weighing each option or justifying responses was a necessary strategy for them. Nearly all of them reported the use of this strategy. They usually read all the options before they made a choice. S1, S3, S9, S12, S15, S6, S23 and S21 used their past experience in communicating in English and their sociopragmatic knowledge to justify some of the responses to the items. For example, “*when people go to a shop, the assistant should say 'can I help you'? Yes, we practice it for many times in class*” (S9, Item 6). “*Let me recall what my foreign teacher always says when he meets me, 'Nice to meet you' or 'good afternoon', 'good morning'? Oh, no, he always says 'Hi', but there is no such choice. Ok, I will make a guess*” (S21, Item 7).

Discussion

Reliability and validity are complementary aspects of the validation process (Bachman, 1990). The present research is proved to have a high reliability and validity in the MDCT when students were tested on interlanguage pragmatic routines in the Chinese EFL context. The results were in line with some previous studies (Liu, 2004, 2007; Rover, 2014), but different from some others (Rose, 1994b; Rose and Ono, 1995; Yamashita, 1996a, 1996b; Yoshitake, 1997; Duan, 2012).

Nevertheless, it is noteworthy that the reliability and validity of the MDCT in this study were high. The reliability was calculated by the split half method. The validity was decided by the difficulty level and discrimination power of the items as well as the discrimination power of the distractors. All the items had an acceptable difficulty level and discrimination power. All the distractors were as functional as intended. The three strategies used in the process of taking the test also confirmed the high validity of the test, while the obstacles to the students' understanding of English revealed in the first two strategies were obvious, and especially for the last strategy, weighing options or justifying responses, which was used the most frequently by the test takers, and the frequent use of this strategy indicates that Chinese culture really plays a role in confusing EFL learners in selecting the most appropriate option for the

items of English pragmatic routines. This might be due to the procedures involved in the development of the test. The situations and options of the MDCT, except for a small part adapted from the established tests, were independently developed in several stages based on the Chinese EFL context. According to Rover (2014), the basic concern for testing interlanguage pragmatic knowledge is that the items should represent the real-world language use instead of being based on the intuitions of the designers, which may or may not be accurate enough to reflect reality.

For the MDCT development, all the situations were both closely related to the participants' real life and English native speakers' life. All the distractors were generated by the English teachers in China based on the experience of frequent inappropriate performances on English pragmatic routines of the EFL learners. The MDCT was tested several times in the pilot study. Revisions were conducted after each time for both the items and their distractors. Finally, both the Chinese teachers and the American teachers were invited to check the content validity in order to ensure that the situations could occur in both countries. In a word, the procedures in developing the test were effective in enhancing reliability and validity. The different stages involved in generating the situations, including the most appropriate option generation, distractor generation, the most appropriate option confirmation, distractor revision, and the relevant statistical calculations, contributed to the quality of each situation in the pragmatic competence test on routines.

In addition, the results of the present study reveal that pragmatic routines are not easy for Chinese EFL learners since the highest mean score was only .66 for the items. This can be explained by cultural differences and university requirements.

Cultural differences are thought of as one of the major causes that lead to pragmatic failures (Hinkel, 1999). Cultural differences are the differences in rules and habits among people from different cultural backgrounds when communicating with each other (Yin, 2009). People grow up in certain communities where their customs of using language are formed. To accept different behaviors for achieving the same goal is not easy for them. It is not exaggerating to say that language is the reflection of culture and culture is the soul of language. The cultural differences between the English and the Chinese, which are caused by historical allusions, regional and natural environments, religious beliefs, and modes of cognitive thinking, can be seen in many aspects (He, 2011).

The requirements of the teaching syllabus for English in China also influence the test takers' acquisition of L2 pragmatic routines. Although pragmatic knowledge has been included in the teaching requirements of English since 2007, no enough attention has been paid to it and English pragmatic knowledge needs to be incorporated into tests. Chinese students are used to taking tests and considering passing tests as the motivation for their studies. However, up to now, no established tests of English pragmatics are available (Li & Zou, 2015). Although some studies (Rover, 2014) have examined the possibilities of tests on English pragmatic routines, the instruments have been applied for research purposes only, and no decisions have been made about using them for pedagogical purposes. If English pragmatic competence test on routines had been developed and adopted by schools and universities, EFL learners would be more motivated to improve their competence.

Furthermore, it was found that situational routines were easier than functional routines, and this result is the opposite to what Rover (1996) found. Classroom teaching and materials in textbooks could explain this. As mentioned in the introductory part of this article, many situations selected for textbooks may be too



idealized. Situational routines, as a matter of fact, allow little flexibility, so those idealized examples might be helpful for EFL learners to accumulate an amount of the standard use of English in pragmatic routines. In contrast, the contextual factors are quite flexible in functional routines, and it is nearly impossible for classroom teaching to reach the goal of inputting various situations into functional routines.

Conclusion

The present study developed an interlanguage pragmatic competence test on routines in the Chinese EFL context. The results show that MDCT is a reliable and valid measure through careful designing and planning. The mean scores of the MDCT items were not high and the situational routines were easier than the functional routines. In addition, three strategies were reported by the students in the process of taking the pragmatic competence test on routines.

Chinese teachers seldom teach English pragmatic routines in class. However, pragmatic routines have been proved to be teachable (Bardovi-Harlig, 2015). A corpus of English pragmatic routines based on authentic language use has been in construction (Bardovi-Harlig, Mossman & Vellenga, 2015). The teaching of English pragmatic routines could be conducted by making use of such corpus. Practice is also necessary for EFL learners. Different activities could be organized to practice English pragmatic routines, such as role plays, short dialogues and so on. Since cultural barriers slow down EFL learners' speed in acquiring English pragmatic routines, more knowledge about western culture should be introduced in class. As far as the present researchers have investigated, most universities in China have courses in British and American culture or Western culture. However, the teaching content in these courses is too general and nothing related to pragmatic routines is taught. In fact, English culture and learners' native culture need to be put into instruction so as to provide the learners with a very clear picture of the differences between the two cultures. It is hoped that EFL teachers will make a conscious effort to include English pragmatic routines into their teaching, and in this way they can help EFL learners develop their awareness of the importance of paying attention to English pragmatic routines.

Although the present research made great efforts in developing the test, there were some limitations. First, MDCT could not elicit qualitative data of the participants' performances. Qualitative data might be helpful to deepen understanding of EFL learners' ability in performing L2 routines. Second, only 20 situations of English pragmatic routines were included in the present study. However, more situations do exist, so the study did not cover all typical situations in English pragmatic routines. In future studies, it is suggested that more testing methods could be involved in an interlanguage pragmatic routines test, such as written discourse completion tasks, oral discourse completion tasks and role plays. In addition, it is also suggested that more situations could be adopted for use in such a test by making use of the established corpora on authentic language use.

Acknowledgements

The present study originated from doctoral research supported in part by a grant from Suranaree University of Technology, Thailand. This study would not have been completed without the generous help of Dr. Peerasak Siriyothin, Asst. Prof. Dr. Kornwipa Poonpon, Asst. Prof. Dr. Nawanmin Prachanant and Dr. Butsakorn Yodkamlue. We would also like to acknowledge the teachers and students who participated in the study from China and Thailand.

About the Authors

Lan Xu, Ph.D. is a lecturer at School of Foreign Languages, Guizhou University, China. Throughout her career, she has taught a wide range of courses, such as Academic writing, English grammar and comprehensive English. Her research interests include pragmatics and language testing.

Anchalee Wannaruk, Ph.D. is an Associate Professor from the School of Foreign Languages, Suranaree University of Technology, where she currently serves as a Chair. Throughout her career, she taught a wide range of courses, i.e. Business English, Technical English, Second Language Acquisition and Discourse Analysis. Her interests include discourse analysis, sociolinguistics, pragmatics and corpus linguistics.

References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. New York: Oxford University Press.
- Bardovi-Harlig, K. (2015). Operationalizing conversation in studies of instructional effect in L2 pragmatics. *System*, 48, 21-34. doi: 10.1016/j.system.2014.09.002
- Bardovi-Harlig, K., Mossman, S., & Vellenga, H. E. (2015). Developing corpus-based materials to teach pragmatic routines. *TESOL Journal*, 6(3), 499-526. doi: 10.1002/tesj.177
- Coulmas, F. (Ed.). (1981). *Conversational routine: Explorations in standardized communication situations and prepatterned speech* (Vol. 96). Walter de Gruyter.
- Duan, L. L. (2012). The study of reliability and validity of pragmatic competence testing measures. *Foreign Languages Studies*, 4, 84-96.
- Ellis, R. (2008). *The study of second language acquisition*. New York: Oxford University Press.
- He, G. (2011). Influences of cultural differences between the Chinese and the Western on translation. *Journal of Language Teaching and Research*, 2(2), 483-485. doi: 10.4304/jltr.2.2.483-485
- Hinkel, E. (1997). Appropriateness of advice: DCT and multiple choice data. *Applied Linguistics*, 18(1), 1-26. doi: 10.1093/applin/18.1.1
- Hinkel, E. (1999). *Culture in second language teaching and learning*. Cambridge: Cambridge University Press.
- Jack, C. R. & David, B. (2005). *Four corners 3*. Cambridge: Cambridge University Press.
- Jack, C. R., Jonathan, H. & Susan, P. (2005). *Interlanguage 2*. Cambridge: Cambridge University Press.
- Jiang, Z. H. (2009). *A probe into the assessment of Chinese college students' pragmatic competence* (Doctoral dissertation). Shanghai Foreign Languages Studies.
- He, Z. R. (2003). *Notes on pragmatics*. Nanjing: Nanjing Normal University Press.
- Hudson, T., Detmer, E. & Brown, J. D. (1995). *Developing prototypic measures of cross-cultural pragmatics. Technical Report 7*. Honolulu, HI: University of Hawaii, Second Language Teaching and Curriculum Center.
- Kasper, G. (1997). Can pragmatic competence be taught? *NetWork*, 6, 105-119.



- Kasper, G. (1998). Interlanguage pragmatics. In Byrnes, H. (Ed.), *Learning foreign and second languages: Perspectives in research and scholarship* (pp. 232-256). New York: The Modern Language Association of America.
- Li, Q. H., & Zou, R. (2015). A review on the ILP competence testing. *Journal of PLA University of Foreign Languages*, 38(4), 43-50.
- Liu, J. D. (2004). Measuring interlanguage pragmatic knowledge of Chinese EFL learners. (Doctoral dissertation). City University of Hong Kong).
- Liu, J. D. (2007). Developing a pragmatics test for Chinese EFL learners. *Language Testing*, 24(3), 391-415. doi: 10.1177/0265532207077206
- Liu, J. D. (2012). Chinese EFL learners' English proficiency and their pragmatic competence development. *Foreign Languages in China*, 9(1), 64-70.
- Ma, T (2010). The interlanguage pragmatic competence of English majors and its development in ethnic universities---A case study of Beijing ethnic university. (Doctoral dissertation). Shanghai International Studies University.
- Madsen, H. S. (1983). *Techniques in testing*. New York: Oxford University Press.
- Michael, M., Jeanne, M. & Helen, S. (2005). *Touchstone 1*. Cambridge: Cambridge University Press.
- Nemati, M., Rezaee, A. A., & Mahdi H. M. (2014). Assessing pragmatics through MDCTs: A case of Iranian EFL learners. *Iranian Journal of Applied Language Studies*, 6(2), 43-66.
- Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- Rose, K. R. (1994a). On the validity of discourse completion tests in non-western contexts. *Applied Linguistics*, 15(1), 1-14. doi: 10.1093/applin/15.1.1
- Rose, K. R. (1994b). On the validity of discourse completion tests in non-western contexts. *Applied Linguistics*, 15(1), 1-14.
- Rose, K. R. & Ono, R. (1995). Eliciting speech act data in Japanese: The effect of questionnaire type. *Language Learning*, 45(2), 191-223. doi: 10.1111/j.1467-1770.1995.tb00438.x
- Rover, C. (1996). Linguistische routinen: Systemarische, psycholinguistische und fremdsprachendidaktische Überlegungen. *Fremdsprachen und Hochschule*, 46, 43-60.
- Rover, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing*, 23(2), 229-256. doi: 10.1191/0265532206lt329oa
- Rover, C. (2014). *Testing ESL pragmatics*. New York: Peter Lang. doi: 10.3726/978-3-653-04780-6
- Smith, S. (2013). Determining sample size: How to ensure you get the correct sample size. Retrieved from: <https://www.qualtrics.com/blog/determining-sample-size/>
- Tabatabaei, S., & Farnia, M. (2015). Learner's English proficiency and their pragmatic competence of refusal speech acts. *Beyond Words*, 3(1), 53-77.
- Yamashita, S.O. (1996a). *Comparing six cross-cultural pragmatics measures*. (Doctoral Dissertation). Temple University.
- Yamashita, S.O. (1996b). *Six measures of JSL pragmatics. Technical Report 14*. Honolulu, HI: University of Hawaii, Second Language Teaching and Curriculum Center.
- Yin, L. (2009). Cultural differences of politeness in English and Chinese. *Asian Social Science*, 5(6), 154-156. doi: 10.5539/ass.v5n6p154
- Yoshitake, S. S. (1997). *Measuring interlanguage pragmatic competence of Japanese students of English as a foreign language: a multi-test framework evaluation*. (Doctoral Dissertation). Columbia Pacific University.



Appendix: Interlanguage Pragmatic Competence Test

Name: _____ (in Chinese) Student No: _____

Instruction: Read each of the following situations and decide which one is the BEST choice in the situation. Then put your answers on the ANSWER SHEET by circling the corresponding letters.

1. One day, when a lady bumped into Jim,
The lady said: "I am terribly sorry."
Jim would probably say: " _____ "
a) No bother.
b) It's nothing.
c) That's all right.
d) Don't mention it.
2. At a bus stop
Man: "Excuse me, do you know which bus to catch for London Road, please?"
Woman: "Sorry, I've no idea."
The man would probably say: " _____ "
a) Oh!
b) Really?
c) Forget it.
d) Thank you.
3. In a factory, Rose, the guide, is interpreting for a group of foreign guests. When they have finished visiting one workshop, she wants the group to follow her to the next workshop.
Rose would probably say: " _____ "
a) Come here!
b) Follow me!
c) Move on!
d) This way, please.
4. Ted is inviting his friend to a little party which will be held at his house tomorrow night.
Ted: "I'm having a little party tomorrow night at my place."
How would Ted probably continue: " _____ "
a) Are you there?
b) Why aren't you showing?
c) Do you think you could come?
d) How would you think about it?
5. Betty is a foreign student in China. She has met Peng.
Betty: "I was told that you won the 100-meter race in a national sports meeting last week. Congratulations!"
Peng would probably say: " _____ "
e) Just lucky.
f) Thank you.
g) I can't say I did well.
h) I could have done better if it hadn't been so cold.



6. Suppose you're a shop assistant and a customer comes up to you.
You would probably say: “ _____ ”
- Can I help you?
 - I'd be glad to serve you.
 - You are welcome, what do you want?
 - Would you like to buy something?
7. On the way to the school cinema, Lisa saw Professor Blake walking to the cinema, too.
Lisa would probably say: “ _____ ”
- Where are you going?
 - Are you going to the film?
 - Nice to see you, Professor Blake.
 - Good afternoon, Professor Blake.
8. Claudia calls her college classmate Dennis, but his roommate answers the phone and tells her that Dennis isn't home. Claudia would like the roommate to tell Dennis something.
Claudia would probably say: “ _____ ”
- Can you take a note?
 - Can I leave a message?
 - Can you tell him something?
 - Can I give you information?
9. Jack phones Peter's office.
Jack: “Hello, I'd like to speak to Peter, please.”
Peter would probably say: “ _____ ”
- That is Peter.
 - This is me.
 - It's me here.
 - This is Peter speaking.
10. Mr. Green's secretary, Miss Kent, went to the airport to meet Mr. Barnes for her boss.
Miss Kent would probably say: “ _____ ”
- Are you Mr. Barnes?
 - You are Mr. Barnes, aren't you?
 - Excuse me, would you be Mr. Barnes?
 - Excuse me, would you please tell me if you are Mr. Barnes?
11. Tom is having dinner with his friends, and suddenly his phone rings.
Tom would probably say to his friends: “ _____ ”
- Sorry.
 - Excuse me.
 - Wait a minute.
 - You eat first, and I will answer the phone.

12. Lisa and Mike are classmates. One day, Lisa is sitting in the classroom, and Mike walks in and wants to sit beside Lisa.
Mike: "Excuse me, do you mind if I sit here?"
Lisa would probably say: "_____"
a) Yes, please.
b) Go ahead.
c) Up to you.
d) No, my friend will sit here.
13. Tom is talking to his school-mate, David.
Tom: "How did you do in the exam, David?"
David: "I barely passed. I made a hopeless mess of it. I don't know why I did so badly."
Tom would probably say: _____
a) It's not your fault.
b) Just try to forget about it.
c) Oh, I shouldn't ask you about it.
d) Hey, how about having a big dinner together?
14. Jane is at the beach and wants to know what time it is. She sees a man with a watch.
Jane would probably say: "_____"
a) Excuse me, how late is it?
b) Excuse me, can you say the time?
c) Excuse me, do you have the time?
d) Excuse me, what does your watch show?
15. Sam is having dinner at a friend's home. His friend offers him more food but he couldn't possibly eat another bite.
Sam would probably say: "_____"
a) No, thanks, I'm full.
b) No, thanks, I've eaten.
c) No, thanks, I've done.
d) No, thanks, I've finished.
16. Cathy works as a secretary in the international Computer Engineering Corporation. One day she worked very late. Her boss said to her: "Thanks a lot. That's a great help."
Cathy would probably say: _____
a) Forget it.
b) You're welcome.
c) Oh, it's nothing.
d) Don't mention it.
17. In a hotel dining room
A foreign guest tells the waitress what he wants for dinner, but the waitress does not hear him clearly.
The waitress would probably say: "_____"
a) Repeat?
b) Pardon?
c) What?
d) Again?



18. Tim is ordering food at a restaurant where you can sit down or take the food home with you.
The waitress would probably say: “ _____ ”
- a) Take away?
 - b) For home or here?
 - c) For here or to go?
 - d) For going or staying?
19. In London, you want to go to Heathrow Airport by taxi.
You would probably say to the taxi driver: “ _____ ”
- a) Heathrow airport. Please.
 - b) Would you please take me to Heathrow airport?
 - c) Excuse me, would you mind taking me to Heathrow airport?
 - d) Excuse me, could you possibly take me to Heathrow airport?
20. One of your friends gives you a gift
Your friend: Here is a little something for you. I hope you like it.
You would probably say: “ _____ ”
- a) Oh. I have got one already.
 - b) No, no, I really can't accept it.
 - c) Oh, you really shouldn't have gone to the expense.
 - d) That's very nice of you. Oh, that's just what I want. Thanks a lot.