

PASAA

Volume 52

July - December 2016

**Testing University Learners' Interlanguage
Pragmatic Competence in a Chinese EFL Context**

Lan Xu

Guizhou University and Suranaree University of Technology

Anchalee Wannaruk

Suranaree University of Technology

Abstract

Speech acts are the major concern of interlanguage pragmatists. The present study aimed to 1) examine the reliability and validity of an interlanguage pragmatic (ILP) competence test on speech acts in a Chinese EFL context, and 2) investigate EFL learners' variations of ILP competence by language proficiency. Altogether 390 students participated in the present study. The students were divided into three groups based on their language proficiency. The data were collected with an ILP competence test and semi-structured interviews. The ILP competence test was in the form of a written discourse completion task (WDCT), including ten speech acts and 30 situations. Data analysis methods included the Many Facets Rasch Model (MFRM), one-way ANOVA, post-hoc Scheffe test and content analysis. The results indicated that the ILP competence test was with high reliability and validity, and variations existed in four aspects of conducting speech acts: 1) use of correct speech act, 2) typical expressions, 3) amount of speech and

information, and 4) degrees of formality, directness and politeness, according to the level of language proficiency. Overall, the students with higher language proficiency performed better than the ones with lower language proficiency.

Keywords: ILP competence test, speech acts, reliability, validity, language proficiency

Introduction

Interlanguage pragmatics is an interdisciplinary subject of second language acquisition and pragmatics. Interlanguage pragmatic (ILP) competence concerns foreign language learners' ability to comprehend and develop pragmatic knowledge (Kasper & Blum-Kulka, 1993). As an indispensable component of general language knowledge, interlanguage pragmatics investigates how language learners use their linguistic resources appropriately in particular contexts (Kasper & Rose, 1999).

After the idea of "interlanguage pragmatics" was introduced into language education (Cohen & Olshtain, 1981), more and more attention has been paid to it. Many researchers turned their interest to the relationship between ILP competence and language proficiency. Ellis (2008) states that language proficiency is vitally important for the acquisition of L2 pragmatics. Language proficiency is referred to as the learners' knowledge of L2 vocabulary and grammar, and their ability to use language skills (Bachman & Palmer, 1996). The common sense assumption is that the development of language competence is accompanied by the development of pragmatic competence (Arghamiri & Sadighi, 2013). However, some researchers do not agree with this.

Hoffman-Hicks's (1992) study represents the starting point of ILP competence. He found a positive relationship between ILP competence and language proficiency with 14 Indian French learners. Garcia (2004) conducted a study with 35 EFL learners from 12 different countries. He investigated four speech acts with 48 multiple

choice discourse completion task (MDCT) items. He found a positive relationship between ILP competence and language proficiency. Xu, Case and Wang (2008) investigated four speech acts with 126 EFL learners from 20 countries. By using a questionnaire with 20 scenarios, they found that the development of ILP competence and grammatical ability were positively related. However, Liu's (2004) study was conducted with 200 participants in a Chinese EFL context, and he did not find any relationship between ILP competence and language proficiency with a test of two speech acts. In 2014, Li and Jiang's study followed with a focus on ILP competence in the form of a written discourse completion task (WDCT) for four speech acts, but they did not find any relationship between ILP competence and language proficiency in a Chinese EFL context with a sample of 103 students.

Limitations can be found in previous studies on the relationship between ILP competence and language proficiency. Some studies were conducted with a too small sample size (Hoffman-Hicks, 1992; Garcia, 2004), and the results might not be representative enough. Some researchers collected data with MDCT or true/false questions (Garcia, 2004; Xu, et. al., 2009), so no qualitative data could be collected, and test takers might have achieved scores by chance. In addition, all previous studies covered limited types of speech acts, and no more than four speech acts have been found.

In order to understand EFL learners' ILP competence, an ILP competence test is needed. The present study applied WDCT as the testing tool for ILP competence, because WDCT is easy to administer with a large sample, and with both quantitative and qualitative data included, it could help deepen the understanding of language learners' ILP competence. In developing an ILP competence test, reliability and validity are the most important factors which should be taken into consideration. In previous studies, WDCT has been proved to be a reliable instrument in testing EFL learners' ILP competence on speech acts by most researchers (Yamashita, 1996a, 1996b; Hudson, 2001; Liu, 2004; Rover, 2014; Liu, 2015). For the validity of WDCT, researchers could not reach an agreement. Hudson, Detmer and

Brown (1995), Yamashita (1996a, 1996b), Ahn (2005) and Rover (2014) revealed that WDCT was valid in testing EFL learners' ability to conduct speech acts, while Rose (1994) and Rose and Ono (1995) drew an opposite conclusion.

Generally speaking, the research on ILP competence testing is still at the beginning stage, and China is no exception (Yue, 2015). The present researchers have not found any research which covers a wide range of speech acts in an ILP competence test. In addition, up to the present, researchers have not found any research conducted in the Guizhou Province, China, to investigate the relationship between ILP competence and language proficiency. The Guizhou Province has the second largest ethnic minority population of the country with 49 different minority groups which accounts for 38.9% of the province's total population. Thus, it is quite interesting to conduct a study to explore the relationship between ILP competence and language proficiency in the Guizhou Province. It is hoped that the study will enrich ILP competence testing literature. It is also hoped that the study will be helpful for EFL teachers and learners in teaching and learning English pragmatics. Two research questions were to be answered in the present study:

- 1) What are the reliability and validity of WDCT in testing EFL learners' ILP competence on speech acts?
- 2) Do the EFL learners with different language proficiencies perform differently in the ILP competence test?

Research Methodology

Participants

The participants in the present study were 390 second-year English major students from four universities in the Guizhou Province of China who had just completed their Test for English Majors Band 4 (TEM-4).

The students were divided into three groups based on their TEM-4 scores with the trichotomy method. There were an equal number of students in each group. TEM-4 is a test that English majors

in China have to take in their second academic year, and it is considered as a tool to evaluate the learners' language abilities. In addition, 24 of the 390 students were selected for the semi-structured interviews, eight students in each language proficiency group.

Research instrument

The research instruments in the present study included both WDCT and semi-structured interviews. The WDCT was developed for the Chinese EFL context by the present researchers, including ten speech acts with 30 situations. In order to develop the WDCT, 100 Chinese university students excluded from the 390 participants in Guizhou province and 33 native English speakers in the Confucius School of Guizhou University were invited. The development of the WDCT experienced the following four steps: 1) selecting speech acts to be tested, 2) generating situations, 3) investigating likelihood, and 4) checking for content validity.

Step 1: Selecting speech acts to be tested

All the speech acts in Searle (1969) and the speech acts that appeared in the previous studies were listed in a questionnaire. The 100 students were required to select the most frequently used ten speech acts in their daily life. With the exception of three students who did not select the number of speech acts as required, all the rest fulfilled the requirement. The three students' questionnaires were discarded. The top ten speech acts (advice, gratitude, greeting, congratulation, apology, request, compliment, inquiry, refusal and compliment response) which were most frequently chosen were kept.

Step 2: Generating situations

In this step, a questionnaire was designed with an example situation written both in English and Chinese for each speech act. The 100 students were required to write one situation, either in English or Chinese, for each speech act. Altogether 173 speech acts were obtained. The numbers of situations collected for each speech act were not equal, ranging from 11 to 23.

Step 3: Investigating likelihood

All the situations collected in Step 2 were organized under each speech act and the Chinese situations were translated into English. A questionnaire was designed to explore the possibility of occurrence for each situation. This questionnaire was designed with a five point rating scale, ranging from “1”not possible at all to “5”the most appropriate. The 33 native English speakers were invited to select the possibility of the situation happening in their own culture. By calculating the mean scores, the top 3 situations for each speech act were kept. In total, 30 situations were collected.

Step 4: Checking for content validity

The 30 situations collected from Step 3 were reorganized without changing the original meaning. Two American teachers in the School of Foreign Languages, Suranaree University of Technology, Thailand, were invited to check the content validity of the situations. After that, the inappropriate expressions were revised, the situations which could not elicit the expected speech act were rewritten, and the situations which were not typical in both America and China were replaced.

Data Collection

The WDCT was administered to 390 university students in classroom circumstances. The time given for the WDCT was 90 minutes. Immediately after that, the semi-structured interviews were conducted. The language used in the interview was Chinese, and all the interviews were recorded. The time length for each interviewee was around 20 minutes.

Data Analysis

The data in the present study were analyzed both quantitatively and qualitatively. In order to answer the first research question, the reliability and validity of the WDCT were calculated under the Many

Facets Rasch Model (MFRM) with FACETS (3.71.4) software. For the second research question, the data were analyzed quantitatively with a one-way ANOVA with post-hoc scheffe test, and qualitatively with content analysis to investigate the variations of ILP competence with the different language proficiency groups. The interview data were analyzed qualitatively, using content analysis for understanding the EFL learners' opinions and experiences on their acquisition of English pragmatics.

Results

Two American teachers for English in Guizhou University were recruited to rate the WDCT. The rating rubrics were adapted from Hudson et al. (1995), and four aspects in conducting speech acts were evaluated with a five point rating scale, ranging from "1" not appropriate at all to "5" completely appropriate. The total score for each item was 20 points. The four aspects in the rating rubrics were: 1) use of correct speech act, 2) typical expressions, 3) amount of speech and information, and 4) degrees of formality, directness and politeness. If the two raters were not self-consistent or consistent with each other in rating the WDCT, a third rater would be invited.

Research Question 1: Reliability and Validity of WDCT

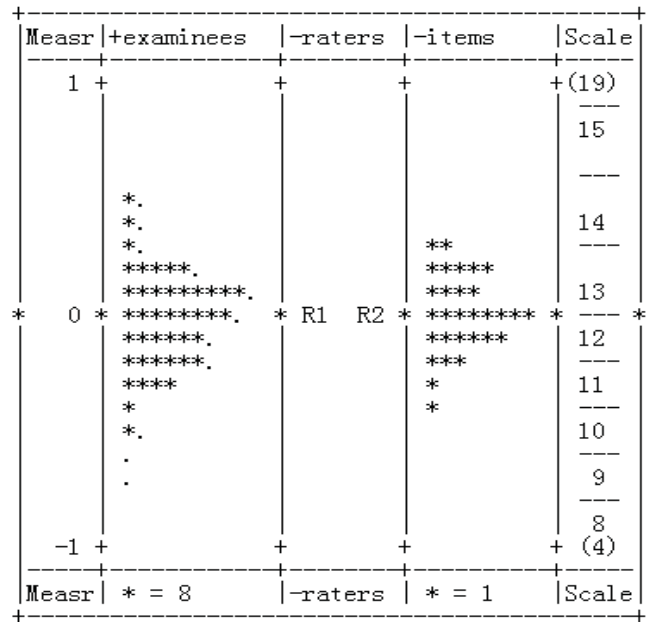
The reliability and validity of the WDCT were examined from four facets: 1) ability of the examinees, 2) leniency/severity of the raters, 3) difficulty of items, and 4) rating scales. The following model describes the relationship of the four aspects and the results for a test. Figure 1: WDCT evaluation model



The map in Figure 2 is a general description of the reliability and validity of the WDCT in the present study. In detail, the performance of the examinees, the leniency/severity of the raters, the difficulty of the items and the rating scales are shown in column 2, column 3, column 4 and column 5, respectively. The first column

provides the linear, equal-interval logits scale on which all facets of the WDCT are positioned. The examinees were ordered from higher performance to lower performance, ranging from +1.0 logits to -1.0 logits. Examinees with “+” were with higher abilities while examinees with “-” were with lower abilities. The two raters were similar on level of leniency/severity and their leniency/severity level was at around 0.0 logits. The items’ difficulties were ranged from +1.0 logits to -1.0 logits. Items with “+” logits were more difficult and items with “-” were less difficult. In the last column, it shows that the EFL learners achieved scores from 4 to 19 for the items.

Figure 2: Facet map for the WDCT



More specifically, for the examinees, their ability measures spanned +.53 logits to -.65 logits. The Infit MnSq (mean square) spanned 1.79 to .44. The Infit ZStd (Z Standard Score) spanned +3.5 to -3.7. Four examinees (S17, S15, S19 and S18) were misfit since their infit MnSq was higher than the maximum (mean + 2 deviations), and three examinees (S53, S40 and S34) were overfit because their infit MnSq was lower than the minimum (mean - 2 deviations) (Linacre, 2014). The percentage (1.8%) of the examinees who were misfit or

overfit was still acceptable (< 2.0%) (Pollitt and Huchinson, 1987). In addition, the separation index was 3.47 (>2.00) and the separation reliability was .92 (>.70), which indicates a significant difference existed among the examinees' ability. The fixed Chi-square was 5236.1 with a d.f. (degree of freedom) of 389 and the significance level was .00 (<.01), which further confirms a significant difference among the examinees. More details about the examinees are presented in the following table.

Table 1: Facets result in WDCT for examinees' ability (Arranged by fN)

Examinee	Measure	SE	Fit	
			Infit MnSq	Infit ZStd
S17	.20	.07	1.79	3.5
S15	.20	.07	1.53	2.5
S19	-.34	.06	1.52	2.5
S18	.01	.06	1.53	2.5
S246	-.23	.06	1.37	1.8
...
S181	-.65	.06	.92	-4
S241	-.65	.06	.92	-4
S301	-.65	.06	.92	-4
...
S195	.53	.07	.82	-9
S255	.53	.07	.82	-9
S315	.53	.07	.82	-9
S375	.53	.07	.82	-9
...
S65	-.24	.06	.67	-1.9
S149	-.20	.06	.65	-2
S53	-.32	.06	.61	-2.4
S40	-.05	.06	.60	-2.4
S34	.33	.07	.44	-3.7
Mean	-.02	.06	1.00	.0
SD	.23	.00	.20	1.1

Model, Sample: Separation 3.47 Reliability .92.

Model, Fixed (all same) chi-square: 5236.1 d.f.: 389 Significance (probability): .00

Note: The examinees are arranged from the most capable to the least capable.

"..." means examinees in the middle are omitted.

For the raters, Rater 1 was more severe than Rater 2 and their difference of leniency/severity was .02 logits. No rater was identified as misfitting or overfitting since each Infit MnSq was within the mean \pm 2 deviations and each Infit ZStd was within \pm 2.0. Both raters were self-consistent. The separation index was 1.47 (<2.00) and the reliability of separation was .68 (<.70). The chi-square was 3.2 with a d.f. of 1, and

the chi-square significance was .08 (>.05), which indicates that the leniency/severity of the two raters were not significantly different. The follow table provides more information of the raters.

Table 2: Facets result in WDCT for the raters' leniency/severity (Arranged by fN)

Rater	Measure	SE	Fit	
			Infit MnSq	Infit ZStd
R1	.01	.00	1.06	1.8
R2	-.01	.00	.94	-2.0
Mean	.00	.00	1.00	-.1
SD	.01	.00	.08	2.8

Model, Sample: Separation 1.47 Reliability .68

Model, Fixed (all same) chi-square: 3.2 d.f.: 1 significance (probability): .08

Note: The raters are arranged from severe to lenient.

For the items, Table 3 presents the range of item difficulty spanned from .28 to -.45 logits. No items were misfitting or overfitting since their Infit MNSq was within mean \pm 2 deviations and Infit Zstd was within mean \pm 2.0. The separation index was 9.03 (>2.00) and the reliability of separation was .90 (>.70), which indicates that the items' difficulty level was significantly different. The chi-square 2332.1 with a d.f. of 29 and the chi-square significance .00 (< .01) further confirms this.

Table 3: Facets result in WDCT for item difficulty(Arranged by fN)

Item	Measure	SE	Fit	
			Infit MnSq	Infit ZStd
17	.15	.02	1.11	2
19	.28	.02	1.10	1.9
14	-.02	.02	1.09	1.6
114	-.08	.02	1.10	1.8
13	-.06	.02	1.10	1.8
...
123	-.18	.02	.90	-1.9
12	-.07	.02	.90	-1.9
117	-.08	.02	.90	-2
15	-.04	.02	.91	-1.7
125	-.45	.02	.92	-1.5
Mean	.00	.02	1.00	.0
SD	.16	.00	.17	1.4

Model, Sample: Separation 9.03 Reliability .90

Model, Fixed (all same) chi-square: 2332.1 d.f.: .29 significance (probability): .00

Note: The items are arranged from the most difficult to the least difficult.

The rating scale statistics show the construct validity of the WDCT. For the rating scale, the logit values of the average measures ranged from $-.75$ to $.38$, and they were monotonically increasing. The outfit MnSq was near the expected value of 1.0. No one was greater than 2.0, which indicates that the rating scales were functioning as intended. For the step calibration, the measures were monotonically increasing and the distance for each of the two rating scales was not larger than 4.0 logits, and it suggests that there was no central tendency in the rating. The following table provides more details of the rating scale statistics.

Table 4: Rating scale statistics

Category score	Data			Avge Meas	Fit		Step Calibration	
	Counts Used	%	Cum. %		Exp. Meas	Outfit MnSq	Measure	S.E.
4	25	0	0	-.75	-.52	.6		
5	88	0	0	-.70	-.47	.6	-1.75	.20
6	197	1	1	-.41	-.40	1.0	-1.24	.10
7	417	2	3	-.31	-.34	1.0	-1.12	.06
8	749	3	6	-.20	-.28	1.2	-.89	.04
9	1228	5	12	-.19	-.21	1.1	-.74	.03
10	2239	10	21	-.17	-.15	1.0	-.78	.02
11	3246	14	35	-.12	-.09	.8	-.49	.02
12	4030	17	52	-.04	-.03	1.0	-.28	.01
13	3802	16	68	.03	.02	1.0	.05	.01
14	3539	15	84	.10	.08	.9	.12	.02
15	2223	10	93	.15	.14	1.0	.58	.02
16	1059	5	98	.20	.20	1.0	.91	.03
17	442	2	100	.20	.26	1.1	1.10	.04
18	107	0	100	.25	.32	1.1	1.71	.09
19	9	0	100	.38	.38	1.0	2.83	.33

Generally speaking, the WDCT was with high reliability and validity in the four facets which may influence the testing results. The 30 items could be used to test the EFL learners' ILP competence on speech acts in the Chinese EFL context.

Research Question 2: EFL Learners' Performances in the ILP Competence Test According to Level of Language Proficiency

The EFL learners showed significant differences in conducting speech acts according to level of language proficiency with $p < .01$. The variation pattern was "High>Medium>Low" in each aspect of the rating rubrics, and the students with higher language proficiency performed

better than the students with lower language proficiency. Generally speaking, the three groups achieved the highest scores in the use of the correct speech act. The students with high or medium language proficiency got the lowest scores in the aspect of typical expressions, while the students with low language proficiency got the lowest scores in the aspect of amount of speech and information. More information is presented in the following table.

Table 5: Variations of EFL learners' ILP competence according to level of language proficiency

	High (n=130)		Medium(n=130)		Low(n=130)		Sig. Level	Variation Pattern
	Mean	S.D.	Mean	S.D.	Mean	S.D.		
Correct speech act	3.46	.17	3.33	.21	3.19	.27	P<.05	High>Medium>Low
Typical expressions	3.10	.17	2.98	.20	2.83	.27	P<.05	High>Medium>Low
Amount of speech and information	3.13	.17	3.00	.21	2.84	.26	P<.05	High>Medium>Low
Degree of formality, directness and politeness	3.13	.17	3.01	.21	2.86	.27	P<.01	High>Medium>Low
Overall	3.21	.17	3.08	.20	2.93	.27	P<.05	High>Medium>Low

To illustrate the differences in EFL learners conducting speech acts with different levels of language proficiency, the following situation is taken as an example.

Situation: Your roommate plays music very loudly, so you can't go to sleep. You ask him/her to turn down the music.

In this situation, the speech act "request" is expected. To conduct this speech act, most students used the syntactic structures as "can you.....", "could you.....", "could you please.....", "would you mind.....", "would you like to.....", "please.....", "I would appreciate if.....". However, some students did not respond with the correct speech act, and "complaint" was conducted instead. The percentages of the students who conducted the wrong speech act were different in the three language proficiency groups. No student in the high language proficiency group conducted the wrong speech act. In contrast, 6.02% of the students in the medium level of language proficiency group conducted the speech act "complaint", and the percentage of the

students who conducted the speech act “complaint” was 18.55% in the low language proficiency group.

For example, S (student) 164 (low language proficiency) wrote “*I have to make complaints. However, if someone is playing music very loudly while you are sleeping, you will know what I feel now*”. In this example, S164 completely misinterpreted this situation, and she did not request the roommate to turn down the music, but complained about the loud music instead. This response could not fulfill the communicating purpose at all. The score she achieved in the aspect of correct speech was one point. Another example is shown by S153 (medium level language proficiency), who wrote “*I don’t want to complain but I can’t stand your playing music. Would you like to turn down the music?*”? In this response, the second sentence “*Would you like to turn down the music?*” was a “request”, and the first sentence “*I don’t want to complain but I can’t stand your playing music*” was a “complaint”. Although a “complaint” was included, the communicative purpose was fulfilled. The score for this response in the aspect of correct speech act was three points. An example of a five-point response is as follows: “*Excuse me, could you please turn down the music? It’s a little loud for me to go to sleep. Thank you*” (S362, high language proficiency).

In the aspect of typical expressions, six patterns were demonstrated in all participants in this situation. They were “apology + request + explanation + gratitude”, “apology + request + explanation”, “request + explanation + gratitude”, “request + explanation”, “request + complaint”, “request”, and “complaint”. The first pattern, “apology + request + explanation + gratitude”, was considered as very appropriate in the aspect of typical expressions. This pattern was used by 25.34% of the students in the high language proficiency group, 8.56% of the students in the medium language proficiency group and 2.12% of the students in the low language proficiency group. The patterns “apology + request + explanation” and “request + explanation + gratitude” were considered as almost appropriate, the percentages of the students who used these two patterns were 42.36% in the high language proficiency

group, 28.29% in the medium language proficiency group, and 17.67% in the low language proficiency group. The patterns “request + explanation” and “request + complaint” were considered as generally appropriate, 34.22% of the students in the high language proficiency group, 58.43% in the medium language proficiency group, and 57.56% in the low language proficiency group used these patterns. The pattern “request” was evaluated as acceptable. No student in the high language proficiency group used this pattern. The percentage of students who used this pattern in the medium language proficiency group and the low language proficiency group were 5.34% and 16.21%, respectively. The last pattern “complaint” was thought as not appropriate at all. The percentages of students who used this pattern in the high, medium and low levels of language proficiency groups were 0.00%, 2.21% and 8.67%, respectively.

The pattern “apology + request + explanation + gratitude” is illustrated in S13’s (high language proficiency) response, *“I am sorry to interrupt you, but could you please turn down the music? It’s a little bit late. Thank you”*. The score for typical expressions was five points. The pattern “apology + request + explanation” or “apology + explanation + request” is as what S31 (high language proficiency) wrote *“Sorry, it’s time to sleep. Could you turn down the music?”* The score for this response was four points in the aspect of typical expressions. The pattern “request + explanation + gratitude” was also frequently used. For example, *“Would you mind turning down the music? It’s a little bit too loud for me to go to sleep. Thank you.”* (S103, medium language proficiency). This response also won a score of four points. The pattern “request + explanation” and “request + complaint” were used by the highest numbers of students in both the medium and the low language proficiency groups. An example of the pattern “request + explanation” is presented with S208’s (medium language proficiency) response *“would you mind turning down the music? I feel so tired that I want to go to sleep”*. This response received a score of three points. An example of the pattern “complaint + request” is illustrated in S284’s (low language proficiency) response *“I can’t bear your loud music, and please turn*

down it". Two points were given for this response. The response with the pattern "request" is as what S289 (low language proficiency) wrote "*Turn down the music*". This response received a score of two points. Although "request" was conducted, the expression was more like an order. The last pattern was "complaint", which was not the expected speech act at all, and the score for this pattern was one point only. For instance, S320 (low language proficiency) wrote "*The music is too loud to go to sleep. It bothers me a lot*".

For the aspect of amount of speech and information, the appropriate amount of speech and information was of high value. The speech and information should be related to the speech act that was expected to be elicited, so the speech and information which was irrelevant was not rated with high scores. The patterns used in the responses of the EFL learners could show the amount of speech and information to some extent. The pattern "apology + request + explanation + gratitude" was considered to be a very appropriate amount of speech and information, and such a response was very complete. For this situation, the patterns "apology + request + explanation" and "request + explanation + gratitude" were thought of as almost appropriate, and the patterns "request + explanation" and "request + complaint" were considered as generally appropriate. The pattern "request" was acceptable. However, any repetition of the speech or information was inappropriate, and one point would be deducted. For example, S385 (medium language proficiency) wrote "*I would be very appreciated if you could turn down your music, and I am really tired. Thank you very much*". In this response, the pattern was "request + explanation + gratitude", so the score should be four points based on the rating criterion, but "*I would be very appreciated*" and "*Thank you very much*" were repetitive, they shared the same function of gratitude. Thus, one point was deducted, and the score for this response in the aspect of amount of speech and information was three points. The pattern "complaint" was not appropriate at all, so however much speech and information was contained in the response, only one point would be given.

The last aspect was degrees of formality, directness and politeness. For formality, the students with higher language proficiency were more capable in choosing the correct words and verb forms. In addition, they were more cautious with face threatening expressions. Thus, in this situation, in order to show their indirectness and politeness, they used words such as “*please*”, “*could*”, “*would*”, “*might*”, sentence structures as “*could you please.....*”, “*Do you mind.....*”, and gratitude strategy by saying “*thank you*”, “*appreciate*” more frequently than students with lower language proficiency. The percentages of the students who used the above words and expressions in the high, medium and low language proficiency groups were 77.34%, 60.90% and 35.21%, respectively.

For example, S264 (high language proficiency) wrote “*Excuse me, do you mind turning down the music? It might be a little late. Thank you*”. The formality of this response was very appropriate, and the response was indirect and very polite, especially with the use of “*excuse me*”, “*do you mind.....*”, “*might*”, “*thank you*” to show the indirectness and politeness. The score of this response in the aspect of degrees of formality, directness and politeness was five points. Another example was in S242’s (medium language proficiency) response, in which she wrote “*Please turn down the music. I really can’t go to sleep. Thank you*”. The expression to request that the roommate to turn down the music “*Please turn down the music*” was more direct and impolite than “*Excuse me, do you mind turning down the music*” (S264), and the explanation “*I really can’t go to sleep*” (S242) showed a stronger degree of unhappiness than “*It might be a little late*” (S264). The use of word “*really*” was not a good choice. Thus, the score for this response in the aspect of degrees of formality, directness and politeness was three points. The next example was the response conducted by S187 (low language proficiency), and she wrote “*Ok, can you giving up playing music at this time?*” The formality of this response was not very appropriate, “*giving up*” did not fulfill the communication purpose of this situation, in which a request was required for turning down the music instead of turning off the music. In addition, a grammatical

mistake also existed in the structure “*Can you giving up.....*”. However, By saying “*can you.....*”, indirectness and politeness were shown, but not as appropriate as in S264’s and S187’s responses. The score for this response was two points in this aspect.

Discussion

Reliability and validity are complementary in the validation process of a test (Bachman, 1990). The reliability and validity of the WDCT in the present ILP competence test were high. This is in line with what Yamashita (1996a, 1996b), Hudson (2001) and Liu (2004, 2015) found, but different from Yoshitake’s (1997) findings. The high reliability and validity of the WDCT might be explained by the developing procedures involved in the present study. All the situations were independently developed for the Chinese EFL context by the researchers, so they were closely related to the daily life of the EFL learners. The native English speakers were also invited to investigate the possibilities of the situations. Both Chinese teachers and American teachers were invited to check the content validity. The efforts made in developing the items ensured the authenticity of the situations. Authenticity is seen as a critical quality of language tests and is said to have a great effect on the test takers’ performance (Bachman and Palmer 1996). Inconsistency might be found between elicitation through NSs and NNSs (Yamashita, 1996a). No such inconsistency was detected from the situations generated for this study. This would suggest that a combination of elicitation through both NSs and NNSs is a more practical way to construct the ILP competence test items. The present ILP competence test examined ILP knowledge since the situations happen both in China and in native English-speaking countries.

In the present study, it was found that the EFL learners’ ILP competence was strongly related to their level of language proficiency. There were significant differences in ILP competence among the three language proficiency groups and the variation pattern was high>medium>low. The results were in accordance with some of the

previous studies (Hoffman-Hicks, 1992; Yamanaka, 2003; Garcia, 2004; Rover, 2006; Xu, et. al., 2009; Liu, 2012; Naoko, 2013), but different from some others (Liu, 2004; Takahashi, 2005; Tian, 2013) who found that there was no relationship between ILP competence and level of language proficiency. Three factors may contribute to explain the differences of the EFL learners' ILP competence in relation to the level of language proficiency for the present study: 1) motivation, 2) out-of-classroom learning, and 3) the general low language proficiency of the participants.

The first factor which relates to the relationship between ILP competence and the level of language proficiency is motivation. According to Ellis (2008), motivation refers to efforts, desire and attitude in language learning. Ushioda (2008) points out that good learners have high motivation. Learners who have experienced success in language learning are highly motivated to learn (Yule, 1996). Niezgodna and Rover (2001) and Shao, Zhao and Sun (2011) report a positive correlation between motivation and ILP competence. Manolopoulo-Sergi (2004) argues that the way in which the learners input, integrate intake and process output in the interlanguage system is influenced by motivation. Students with lower language proficiency might only be able to attend to some surface characteristics of L2 pragmatic input and produce output in a manner which just delivers information. However, students with higher language proficiency might be able to process L2 pragmatic input in a manner which could express the ideas in a more effective and appropriate way.

In the present study, the students with different levels of language proficiency showed different motivations in L2 learning and L2 communication. In the interview, I (Interviewee) 3 (high level of language proficiency) said "*I really want to learn English well. I think pragmatics is very important in language learning and I feel proud when I can use good English to communicate with native speakers*". I23 (high language proficiency) also mentioned that "*when I was in high school, English was my favorite subject, so I spend a lot of time on it and became an English major student*". Students in the medium level of

language proficiency mentioned that *“I work hard to pass the examinations”* (I7, I8) and *“my motivation in learning English is not so high, I may not use English in my job in the future, so I just fulfill the requirements of the teachers”* (I2, I17, I21). On the contrary, students with low language proficiency had different opinions. For example, I11 said *“I don’t like English. English was not my choice as a major, but I failed in the college entrance exam for another major, so I was transferred to be an English major”*. I19 mentioned *“to be frank, my interest is not in English, and to learn English is to make my parents happy”*.

The second factor which may explain the relationship between ILP competence and language proficiency is out-of-classroom learning. In the interview, the students with high language proficiency (I1, I3, I9, I10) reported that they made great efforts in learning English after class. They spent a large amount of time watching English movies, reading English newspapers and novels, and making friends with native English speakers. They benefited more from out-of-classroom learning than from the textbooks and in-classroom teaching and learning interlanguage pragmatics. I1 said *“I learnt typical expressions and English routines through watching movies”*. I10 mentioned *“I think I could immerge myself with native speakers when I communicate with the native speakers, and I learn a lot about their culture”*. On the contrary, the students with low language proficiency (I12, I14, I15, I19, I20) reported that they seldom watched movies, read newspapers or novels in English, even less did they communicate with native speakers. It is due to their poor grammar and limited vocabulary, and they could not understand most English reading materials or native English speakers. To fulfill the requirements of the teachers was not easy for them. Those students with low language proficiency also reported that what they learnt in class was far less than enough to communicate with native English speakers or to finish the ILP competence test. It can be concluded from the interviews that the students with higher language proficiency would have more time and interest in out-of-classroom learning, while the students with lower

language proficiency felt that it was difficult to cope with the in-classroom tasks and to fulfill the teachers' requirements. Out-of-classroom learning is more helpful in improving the EFL learners' ILP competence, while in-classroom study might not be very beneficial in enhancing the ability in interlanguage pragmatics. Hence, out-of-classroom learning might be a factor which relates to ILP competence.

The third factor which could explain the relation between ILP competence and level of language proficiency might be the general low language proficiency of the students in the Guizhou Province. Although some researchers (Liu, 2004; Takahashi, 2005; Tian, 2013) found no relationship between ILP competence and language proficiency, it might be because the EFL learners' language proficiency in those studies had reached a level in which vocabulary, grammar and syntax were not obstacles in understanding. Chen (2007) mentions that the development of pragmatic competence depends on linguistic competence, but this method could only be applied to the learners whose general linguistic competence is not high. In the present study, the mean score of TEM-4 for all the participants was 49.44 and only 15.90% passed (over 60 points), while the mean score of the test in the same year for the students of all comprehensive universities in China was 62.47 and 65.10% test takers passed. The great distance of language proficiency between the participants in the Guizhou Province and in the whole country shows that the language proficiency of the 390 students was really low in general. Their level of language proficiency had not reached a level in which understanding texts would not be difficult for them. The data in the interview also confirmed this. Some interviewees (I2, I4, I11, I16) mentioned that to understand and comprehend the items was still difficult for them, and a few of them (I5, I9, I21) reported that there were new words and unfamiliar expressions for them in the test.

It is understandable that the general language proficiency of the participants was lower than the average level of the whole country. Among the 390 participants, only 194 of them were Han, the majority people in China, and the other 196 were minority people from Miao,

Gelao, Shui, Tujia, Chuanqing, Hui and so on. For these minority students, their first languages (L1) were their minority languages, which are different from Mandarin, the official language of China, in characters, pronunciation, and syntax. Mandarin was their second language (L2), and English was their third language (L3). Their acquisition of L3 had been influenced both by the mother tongue transfer and L2 transfer. Bialystok (2002) mentions that bilingualism has clear effects in the cognitive and intellectual development of language learning. Previous researchers have reported a lower ability in learning a new language for bilingual speakers than monolingual speakers (August and Hakuta, 1997; Hakuta, Butler and Witt, 2000). Thus, bilingualism might bring a more negative transfer to the EFL learners. Accordingly, in the present study, it was reasonable that the participants' general low level of language proficiency would have an influence on the EFL learners' ILP competence.

From the above discussion, it can be concluded that through careful planning and designing, WDCT could be a reliable and valid method in testing EFL learners' ILP competence. The level of language proficiency was a factor which was strongly related to ILP competence in the present study. Although some previous researchers achieved different findings, it might be because the participants were influenced by other variables, such as length of residence in a target language country, exposure to the target culture, exposure to specialized courses and so on. Since the relationship between ILP competence and language proficiency is still quite controversial, further research is needed.

Conclusion

The present study investigated the reliability and validity of WDCT in testing ILP competence in a Chinese EFL context, as well as the EFL learners' performances in the ILP competence test according to the level of language proficiency. Speech acts were the main concern in the present study. Altogether 390 Chinese EFL university students participated in the study. Another 100 university students and 33

native English speakers helped to develop the research instrument. The results show that the WDCT, including ten speech acts and 30 situations, were with high reliability and validity. In addition, significant differences were found in the ILP competence test among the students with different language proficiencies.

In the present study, the students with a higher language proficiency were with higher ILP competence. The students with a higher level of language proficiency reported that they employed a number of out-of-classroom learning methods in improving their ILP competence, which formed a virtuous circle for their language learning. The students with a lower level of language proficiency were less motivated to learn due to their limited vocabulary and poor grammar. In order to help the lower language proficiency learners improve their ILP competence, the teachers could encourage them by assigning some tasks that they can fulfill with less difficulty. When the students feel a sense of success, they will be more motivated to learn. In addition, the teachers should encourage them to increase their vocabulary and enhance their grammar, and certain learning plans could be made with the help of the teachers. Only when the students are equipped with enough vocabulary and grammar, can they be involved in language learning more actively. In addition, the teachers could recommend some learning materials and methods to the EFL learners.

Although great endeavors have been made in the present study, limitations still exist. First, the fundamental concern in constructing items of interlanguage tests is that the items are representative of real-world language use (Wolfson and Judd, 1983). Although the present study made great efforts to guarantee authenticity, which can be reflected in all the steps involved in the development of the WDCT, it was built up with a limited number of EFL learners, native English speakers and English teachers. Second, although the present study has covered ten speech acts, other fields in ILP knowledge were excluded. In addition, the classification of the students depended on the mode and range of the TEM-4 scores, so the students in the high

language proficiency group may not achieve really high scores since the language proficiency for all participants was not high.

Because of the limitations, some suggestions for future research can be made. First, it is suggested that in order to obtain more authentic elicitation of situations, a corpus on spoken language could be established. Second, it is recommended that other fields in ILP knowledge could be included in ILP competence testing, such as implicature and routines. Last but not least, the situations in the present study were developed in the Chinese EFL context, when they are used in other EFL contexts, reliability and validity should be rechecked, and revisions and replacement of the situations might be needed.

Acknowledgements

The present study is supported in part by a grant from Suranaree University of Technology, Thailand. This study would not be complete 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 the Guizhou Province of China and from Suranaree University of Technology in Thailand.

The Authors

Lan Xu is currently a Ph.D. in English language studies. She has been a teacher in the School of Foreign Languages, Guizhou University, China since 2001. Her fields of interests include second language acquisition, pragmatics and language testing. Email: xulanyu80@sina.com

Anchalee Wannaruk is currently an associate professor, Ph.D. advisor and the chair of the School of Foreign Languages, Institute of Social Technology, Suranaree University of Technology, Thailand. She received her Ph.D. from the University of Illinois in 1997. Her research

interests include pragmatics, language testing and corpus-based language learning. Email: wannaruk@sut.ac.th

References

- Ahn, R. C. (2005). Five measures of interlanguage pragmatics in KEL learners Doctoral Dissertation, University of Hawaii.
- Arghamiri, A. & Sadighi, F. (2013). The impact of metalinguistic knowledge and proficiency level on pragmatic competence of Iranian EFL learners. *IJLLALW*, 4(1), 181-192. Retrieved from <http://www.ijllalw.org/finalversion4114.pdf>.
- August, D. & Hakuta, K. (Eds.). (1997). *Improving schooling for language minority children: A research agenda*. Washington, DC: National Academy Press
- Bachman, L. F. (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.
- Bialystok, E. (2002). Acquisition of literacy in bilingual children: A framework for research. *Language Learning*, 52(1), 159-199. doi: 10.1111/1467-9922.00180.
- Chen, B. G. (2007). A study of the correlation between language ability and pragmatic competence of college EFL learners. *Journal of Yili Normal University*, 3, 115-117.
- Cohen, A.D. & Olshtain, E. (1981). Developing a measure of sociocultural competence: The case of apology. *Language Learning*, 31(1), 113-134. doi: 10.1111/j.1467-1770.1981.tb01375.x.
- Ellis, R. (2008). *The study of second language acquisition*. Oxford: Oxford University Press.
- Garcia, P. (2004). Pragmatic comprehension of high and low level language learners. *TESL-EJ*, 8(2), 1-15. Retrieved from <http://www.tesl-ej.org/wordpress/issues/volume8/ej30/ej30a1/>.
- Hakuta, K., Butler, Y. G. & Witt, D. (2000). How long does it take English learners to attain proficiency? (Policy Report 2000-

- 1).California: The University of California Linguistic Minority Research Institute. Retrieved from <http://www.stanford.edu/~hakuta/Docs/HowLong.pdf>.
- Hoffman-Hicks (1992). Linguistic and Pragmatic Competence: Their Relationship in the Overall Competence of the Language Learner. *Pragmatics and Language Learning*, 3, 66-80. Retrieved from <http://files.eric.ed.gov/fulltext/ED395522.pdf>.
- Hudson, T. (2001). Self-assessment methods in cross-cultural pragmatics. In T. Hudson and J. D. Brown (Eds.), *A Focus on language test development* (pp: 57-74). Honolulu, HI: University of Hawaii Press,.
- Hudson, T., Detmer, E. & Brown, J.D. (1995). *Developing prototypic measures of cross-cultural pragmatics*. (Technical Report7). Honolulu, HI: University of Hawaii, Second Language Teaching and Curriculum Center.
- Kasper, G. &Blum-Kulka, S. (1993). *Interlanguage pragmatics*. Oxford: Oxford University Press.
- Kasper, G. &Rose, K. R. (1999). Pragmatics and SLA. *Annual Review of Applied Linguistics*, 19, 81-104. doi: 10.1017/s0267190599190056.
- Li, Y. & Jiang, Z. H. (2014). An Investigation of English Majors' ILP Competence. *Foreign Language Education*, 35(5), 68-71.
- Linacre, J. M. (2014). *A User's Guide to FACETS: Rasch-Model computer programs*. Retrieved from <http://www.winsteps.com/a/facets-manual.pdf>.
- Liu, J. D. (2004). Measuring interlanguage pragmatic knowledge of Chinese EFL learners(Doctoral Dissertation, City University of Hong Kong, 2004).
- Liu, J. D. (2012). Chinese EFL learners' English proficiency and their pragmatic competence development. *Foreign Languages in China*, 9(1), 64-70.
- Liu, Y., (2015). An Empirical Study on the Development Pattern of Chinese EFL Learners' Pragmatic Competence. *Foreign Language Research*, 182(1), 114-119.

- Manolopoulo-Sergi, E. (2004). Motivation within the information processing model of foreign language learning. *System*, 32, 427-442. doi: 10.1016/j.system.2004.04.002.
- Naoko, T. (2013). Production of routines in L2 English: Effect of proficiency and study-abroad experience. *System*, 41(1), 109-121. doi: 10.1016/j.system.2013.01.003.
- Niezgoda, K. & Rover, C. (2001). Pragmatic and grammatical awareness in Rose, K.R. and Kasper, G.(Eds.), *Pragmatics in Language Teaching* (pp:63-79). Cambridge: Cambridge University Press.
- Pollitt, A. & Hutchinson, C. (1987). Calibrating graded assessments: Rasch partial credit analysis of performance in writing. *Language Testing*, 4(1), 72-92.
- Rose, K. R. (1994). 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. & 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. (2006). Validation of a web-based test of ESL pragma linguistics. *Language Testing*, 23(2), 229-256. doi:10.1191/0265532206lt329oa.
- Rover, C. (2014). *Testing ESL pragmatics*. New York: Peter Lang.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139173438.
- Shao, L. J. Zhao, Y. R. and Sun, Q. Y. (2011). Correlation between the English Learner's Learning Motivation, Cultural Identity and the Development of the Learners' Pragmatic Competence. *Hebei Normal University of Science & Technology(Social Sciences)*, 10(2): 88-91.
- Takahashi, S. (2005). Pragma linguistic Awareness: Is it related to motivation and proficiency? *Applied Linguistics*, 26(1), 90-120. Retrieved from <http://applied.oxfordjournals.org/content/26/1/90.full.pdf>.

- Tian, Y. P. (2013). A Study of Cross-culture Pragmatic Skills of the Minority College Students in Qinghai. *Journal of Qinghai Nationalities University*, 39(2), 144-147.
- Ushioda, E. (2008). *Motivation and good language learners*. Cambridge: Cambridge University Press. Retrieved from http://www.lenguasvivas.org/campus/files/0_28/motivation_and_lgelearners.pdf.
- Wolfson, N. & Judd, E. (1983). *Sociolinguistics and Language Acquisition*. New York: Newbury House Publishers.
- Xu, W., Case, R. E. & Wang, Y. (2008). Pragmatic and grammatical competence, length of residence, and overall L2 proficiency. *System*, 37(2), 205-216.
doi: 10.1016/j.system.2008.09.007.
- Yamanaka, J. E. (2003). Effects of proficiency and length of residence on the pragmatic comprehension of Japanese ESL learners. *Second Language Studies*, 22(1), 107-175.
Retrieved from <http://scholarspace.manoa.hawaii.edu/handle/10125/40660>.
- 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.
- 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.
- Yue, X. B. (2015). Application of Scaffolding Instruction in releasing Communication Anxiety. *Journal of the Chinese Society of Education*, 1, 14-17
- Yule, G. (1996). *Pragmatics*. Oxford: Oxford University Press.