

Research Article

VALIDATION OF A MODIFIED PRAGMATIC TEST: C-TEST AS A MEASURE OF FORMULAIC COMPETENCE

Meisam Ziafar and *Parviz Maftoon

Department of English Language Teaching, Science and Research Branch, Islamic Azad University, Tehran, Iran

**Author for Correspondence*

ABSTRACT

As a less-trodden path, the assessment of pragmatic competence represents one of the most challenging areas in language testing. All the methods proposed so far suffer from certain validity aspects such as practicality and underrepresentation of the pragmatic construct. This study was an attempt to introduce C-test method, which has been primarily used as a measure of general language proficiency, into assessing one aspect of pragmatic competence, i.e. formulaic knowledge. A test of pragmalinguistic competence was adapted, modified and delivered to 215 university students in Iran and 19 native speakers. Results obtained revealed a higher level of reliability for C-test compared to multiple-choice discourse completion test (MDCT) in assessing the participants' knowledge about routines and yielded three interrelated components for the adapted pragmatic test. In line with previous studies it was found that general language proficiency was a determining factor in the participants' knowledge about speech acts. The highest overlap was found between implicatures and speech acts, and the lowest overlap existed between routines and implicatures. Further research is recommended on the effective use of C-test in measuring other aspects of pragmatic competence.

Keywords: *Pragmatic Competence, Practicality, C-test, Implicature, Routines, Speech Acts, MDCT*

INTRODUCTION

As an uncertain and ambiguous construct, pragmatic competence has always been a challenging facet of language tests (Liu, 2006; Roever, 2006). Grabowski (2008b) refers to pragmatic skill as “the waste basket into which all aspects of language that cannot be classified elsewhere are thrown” (p. 155). So far, attempt has been made to evaluate this skill through some forms of instruments such as written discourse completion task (WDCT), multiple-choice discourse completion task (MDCT), oral discourse completion task (ODCT), discourse role-play talks (DRPT), discourse self-assessment task (DSAT), and role-play self-assessments (RPSA). A review of literature suggests that three most important aspects of pragmatic competence, which represent the major components of pragmatic tests, are routines (Roever, 1994, 2006, 2012; Klein-Braley, 1985, 1997), speech acts (House & Kasper, 1987; Ikoma, 1993; Olshtain & Weinbach, 1987; Takahashi & Beebe, 1987; Trosborg, 1995) and implicatures (Bouton, 1988, 1990; Garcia, 2004; Roever, 2006). In order to prevent construct under-representation, McNamara and Roever (2006) propose the inclusion of at least these three types of pragmatic constructs in tests of pragmatics. Thus, a pragmatic test with high construct validity includes all these three constructs (Yamashita, 2008; Roever, 2006).

Implicature is defined as the difference between what is said and what is meant i.e. the meaning conveyed by the speaker (Grice, 1975). Grice (1975) introduces two types of implicature, i.e. conventional implicature and conversational implicature. Conversational implicature is based on the context and shared background knowledge of interlocutors. Conventional implicature on the other hand represents a similar entity apart from context. Garcia (2004) classified implicatures into specific implicatures and general implicatures. The first type includes individual speech units which necessitate inference on the part of the listener. General implicature necessitates the use of contextual expectations in order to discover what the speakers intend. In his test, Bouton (1988, 1994) considered two types of implicature to measure EFL learners' pragmatic knowledge: (1) idiosyncratic implicature (when one of the Gricean maxims are violated and one needs to resort to his/her background knowledge to infer the real meaning beyond the literal meaning of language)

Research Article

and (2) formulaic implicate. Bouton (1988) discovered that EFL students' performance in formulaic implicate was considerably weaker compared to their performance in idiosyncratic implicate. Bouton (1994) also came to the conclusion that formal instruction was more effective in promoting language learners' knowledge about formulaic implicate and that less formulaic implicates were less affected by formal instruction.

Some researchers' focus in testing pragmatic competence was more on the appropriateness of speech. Hudson *et al.*, (1995) designed a six-section test in order to assess the sociopragmatic competence of Japanese EFL learners based on the appropriateness of language use. The test consisted of MDCT, WDTC, role plays and two self-assessments based on three speech acts, i.e. request, apology, and refusal. They reported that finding clear distracters was a major challenging task in designing MDCTs. Liu (2006) developed a pragmatic test which was composed of MDCT, WDCT and DSAT instruments geared to evaluating Chinese EFL learners' knowledge about request and apology speech acts. MCDT was designed based on native speakers' responses taken as correct options and non-native speakers' responses as distracters. Liu did not find any significant difference between proficiency groups with regard to their sociopragmatic knowledge. According to McNamara and Roever (2006), there are some disadvantages to Liu's pragmatic test. First, since native speakers provided the correct options and non-native responses were used as distracters, participants may have provided their understanding about the idiomatic nature of the utterances rather than how appropriate they are. Second, it shows a hard-to-explain distinction between request and apology speech acts. Third, it is based on the responses of Chinese EFL learners and may not be suitable for other L1 groups.

In a similar vein, Yamashita (2008) enumerates some shortcomings of DCTs including (1) their wordy structures; (2) numerous possible correct answers; (3) the difference between written and oral speech with regard to the length of responses; and (4) misleading test descriptions. MDCT has always revealed the lowest reliability in pragmatic research (Yamashita, 1996a, 1996b; Yoshitake, 1997). McNamara and Roever (2006) believe that due to the subjectivity of judgments made by native speakers about appropriateness with regard to sociopragmatic norms, in designing MDCT the most difficult task is developing tests with response options which can be considered as distracters. They further believe that DCTs do not clearly present actual speech, are not frequent and natural speech, deprive test takers of nonverbal and contextual cues, cannot predict successful pragmatic performance, are not effective tools in authentic conversation situations which are co-constructed by both interlocutors. They propose that DCT results should be merely considered as indicators of EFL learners' potentials in using their pragmatic knowledge to perform in real contexts. Despite all these shortcomings, as a practical form of assessment of pragmatic competence, multiple-choice test format has been used to test speech acts (Liu, 2006; Roever, 2005; Hudson *et al.*, 1995; Yamashita, 1996; Rose, 2000) implicature (Bouton, 1988, 1994; Garcia, 2004; Roever, 2001), and formulaic or routine expressions (Read, 2000; Roever, 2001).

Measuring Formulaic Competence

As a form of reduced redundancy test (Klein-Braley, 1997), C-test has received approval for its capability in testing language learners' proficiency (Dörnyei & Katona, 1992; Klein-Braley & Raatz, 1984; Klein-Braley, 1985). Although, to the best of the authors' knowledge, no C-test has been used in measuring pragmatic knowledge, it has frequently been employed in assessing language learners' formulaic and lexical knowledge (Read, 2000; Ebrahimi-Bazzaz *et al.*, 2012; Čolović-Marković, 2012; Schmitt *et al.*, 2004).

Ebrahimi-Bazzaz *et al.*, (2012) contend that C-test is useful in that it has a simple test format and thus easy to distribute, can be subjected to objective quick scoring, and represents a high degree of reliability. Klein-Braley (1985) emphasizes the authenticity of C-test as a true measure of the construct of general language proficiency. Oller (1973) considers C-test as a type of pragmatic language test and maintains that such test refers to any type of activity or task that supports language learners in processing stretches of segments in a language that complies with natural environmental restrictions and further necessitates the association of linguistic segments to extra linguistic environments through pragmatic mappings by the language learners.

Research Article

It seems that C-test lends itself to the measurement of EFL learners' pragmatic knowledge in general and their knowledge about routines (formulaic competence) in particular. C-tests can compensate for some of the shortcomings of MDCTs in assessing pragmatic competence. First, since they can be developed based on the contents of language corpora which provide authentic and real language, there is no room for the subjective judgments of native speakers as norms who provide correct and appropriate options for MDCTs. All these emphasize the objective and reliable nature of C-tests (Eckes & Grotjahn, 2006). Second, since they are not bound to any single first language (L1), C-tests developed to assess pragmatic competence can be validated and used for all EFL learners who talk in various L1s. Third, C-tests do not need to include many utterances and can measure one's pragmatic ability through simple contextualized sentences and co-constructed authentic conversations, and thus reduce the role of literacy as a construct-irrelevant factor in answering the pragmatic questions. This fact has been emphasized by Eckes and Grotjahn (2006) who take C-tests as 'economical measurement instruments' (p. 290).

The other benefit of C-test is that it limits the correct answer to one single choice without rejecting the appropriateness of any other possible correct option(s) in the specific situation. What is more, they are easily understood by test-takers and do not impose misleading test descriptions as is the case in MDCTs. Finally, since C-tests are taken from real discourse, they can be taken as authentic real speeches which have taken place in certain situations, with their frequency of use controlled through the use of corpora. The routine section in Roever's test also suffers from these shortcomings. He reported the lowest reliability for routines, and it showed the weakest relationship with the other two test facets. The routine section in Roever's original pragmatic test does not tap production and is more directed toward pragmatic comprehension. This study is an investigation of the validity of Roever's test and the extent to which a C-test can promote the reliability and validity of an already validated pragmatic test.

MATERIALS AND METHODS

Method

Participants

215 Iranian university students who majored in TEFL/TESL (Teaching English as a Foreign/Second Language) at M.A level participated in this study through providing their answers to an online pragmatic test issued through Google. There were 152 females and 63 male participants and 19 native-speakers (10 Americans, 1 Australian, 4 British, and 4 Canadian). The participants had no prior experience in learning pragmatics and had not lived in any native English speaking environment.

Instruments

The original pragma linguistic test was designed to thoroughly determine language learners' pragma linguistic competence through including several aspects of pragma linguistic knowledge (Roever, 2006). Roever contends that all the three components of his test are related in that they are developed by means of pragmatic input and they promote language learners' awareness of how these pragmatic components function. There are three subcomponents in this test: implicate, routine, and speech acts. The implicate section consists of 4 items which measure language learners' knowledge about formulaic implicate and 8 items which measure their idiosyncratic implicate knowledge, all in multiple-choice format. Formulaic implicate items involved indirect criticism and Pope Q questions. Routines section consisted of 12 multiple-choice questions which tested language learners' situational bound expressions and functional routines.

Speech acts section consisted of 12 Multiple-Choice Discourse Completion Test (MDCT) questions with rejoinders which tested language learners' knowledge about three speech acts of request, apology, and refusal. Situational features, i.e. power and distance were kept constant and the degree of imposition differed for different items. Roever found that, in general, the factor analysis in his study revealed three components and that his test introduced a representative repertoire of pragmalinguistic knowledge with each section measuring a fairly independent aspect of that knowledge. Roever further reported a significant difference between native and non-natives' scores, different proficiency groups, and exposure versus non-exposure groups in his test.

Research Article

In order to compensate for the shortcomings of MDCTs already mentioned and promote the validity of pragmatic test, the first section of Roever's test was transformed into a C-test which similarly measures language learners' formulaic competence. On the other hand, as a way of eliminating the construct irrelevant variance of the learners' writing ability and literacy, instead of Written Discourse Completion Test (WDCT) section of Roever's original test, his 12-item MDCT for speech acts was adopted. The conversations embodying pragmatic C-test were taken from two American corpora (Corpus of Contemporary American English or COCA and Corpus of American Soap Operas). The C-test section can be found in appendix.

Procedure

In order to validate the newly developed three-section pragmatic test, it was sent online to 19 native speakers and 215 Iranian M.A students majoring in TEFL. The test was inserted into an online web-based document in Google Docs. Data were gathered through a spreadsheet which could be accessed through logging in Gmail. The accuracy of the modified section of the test in measuring language learners' formulaic competence was piloted through discussions with an English native speaker from Britain and also through obtaining online comments from some ELT professionals who received this test through their emails or in their profiles through LinkedIn.

The reliability analyses for the total test involved calculating Cronbach's alpha. Descriptive statistics and normality of distribution of the whole sample was investigated through SPSS software version 20. Exploratory factor analysis, with test items as variables, was carried out by SPSS in order to discover the subcomponents of the test. Next, in order to evaluate the amount of overlap between test sections, after ensuring that the assumptions of linearity and homoscedasticity were met, correlation and coefficient of determination were calculated. Tests of significance (independent-samples t-tests) were conducted in order to find significant differences between natives versus non-natives. Between-groups one-way ANOVA was conducted in order to find any possible significant difference between general proficiency groups. Samples were appointed to four proficiency levels based on their self-report of their own general proficiency. Finally, the newly designed pragmatic test in this study was subjected to correlation tests in order to find whether it was valid in predicting a significant degree of variation in the original test.

RESULTS AND DISCUSSION

Result

Test Reliability

The Cronbach alpha value for the Routine, Implicate and speech act sections revealed adequate internal consistency for the subscales and the pragmatic test in too. Results are quite compatible with what Roever (2006) reported for the reliability of his scale (Table 1). Considering the low reliability reported for MDCT tests (Hudson et al., 1995; Yamashita, 1996; Yoshitake, 1997; Brown, 2001), it seems that C-test is presents a more reliable measure of EFL learners' formulaic competence.

Table 1: Reliability reports

| Section | Roever's Pragmatic Test | Adapted Pragmatic Test |
|--------------------|-------------------------|------------------------|
| <i>Implicature</i> | 0.82 | 0.89 |
| <i>Routines</i> | 0.73 | 0.87 |
| <i>Speech acts</i> | 0.89 | 0.86 |
| <i>Total</i> | 0.91 | 0.92 |

Three Separate Pragmatic Factors: Distinction versus Overlap

In order to investigate the degree of overlap between the three sections, correlations and coefficients of determination were calculated (Table 2). The sections of the test were moderately correlated. Speech acts and implicature indicate the strongest overlap and explaining almost a third of each other's' variance. Routines and implicature show the weakest overlap and account for 14 percent of each other's variance. This is quite in line with Roever's (2006) findings. Similar to Roever's findings, these results suggest that

Research Article

all three sections of the newly developed test measure the same kind of knowledge to an acceptable degree, when they evaluate specific unique constructs.

Table 2: Factors Correlations and Coefficients

| | Routines (C-test) | Implicature | Speech acts |
|--------------------|-------------------|-------------|-------------|
| <i>Routines</i> | - | 0.14 | 0.22 |
| <i>Implicature</i> | 0.37 | - | 0.33 |
| <i>Speech acts</i> | 0.47 | 0.57 | - |

Note: Correlations are presented on the lower left, and the coefficients of determination are indicated on the upper right.

The Construct of the Test: The Test Components

In order to have a more clear understanding of the test construct, the 36 items of the Roever’s adapted Pragmatic Test were subjected to principal components analysis (PCA) using SPSS Version 20. Prior to performing PCA, the suitability of data for factor analysis was investigated. Inspection of the correlation matrix revealed the presence of many coefficients of .3 and above. The Kaiser-Meyer-Oklin value was .88, exceeding the recommended value of .6 (Kaiser, 1970) and Bartlett’s Test of Sphericity (Bartlett, 1954) reached statistical significance (Sig = .000), supporting the factorability of the correlation matrix.

Principal component analysis revealed the presence of seven components with eigenvalues exceeding 1, explaining 27.8%, 9.6%, 6.1%, 3.7%, 3.1%, 3% and 2.8% of variance respectively. An inspection of the screeplot revealed a clear break after the third component. Using Catell’s (1966) scree test, it was decided to retain three components for further investigation (Figure 1). This was further supported by the results of Parallel Analysis (as presented in table 3), which showed only three components with eigenvalues exceeding the corresponding criterion values for a randomly generated data matrix of the same size (36 variables X 215 respondents). The unrotated component matrix also revealed that most items load strongly on Components 1, 2, and 3. This fact further suggests that a three-factor solution is likely to be more appropriate. What is more, the Pattern Matrix shows ten items loading above .5 on Component 1, nine items loading above .4 on Component 2, and nine items loading above .3 on Component 3. Other Component explanations produce less than four items loading strongly on more than three components (c.f. Appendix A).

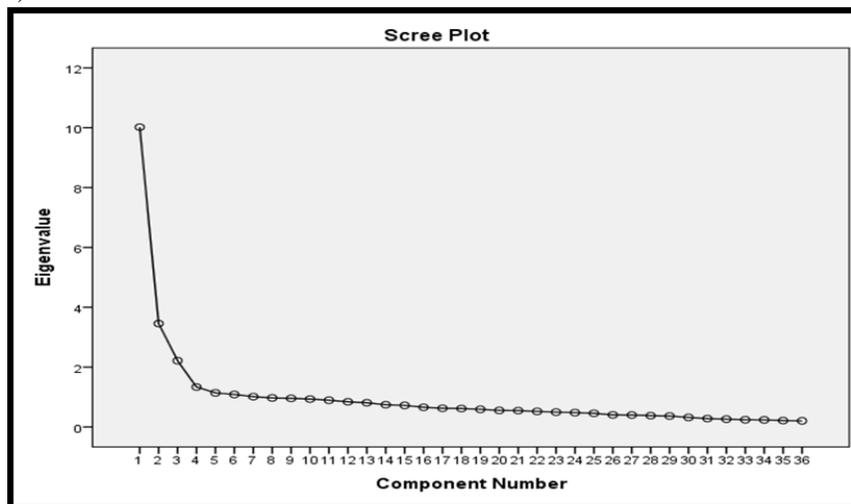


Figure 1: Scree Plot for Pragmatic Test Components

Research Article

Table 3: Comparison of eigenvalues from PCA and criterion values from parallel analysis

| Component number | Actual eigenvalue from PCA | Criterion value from parallel analysis | Decision |
|------------------|----------------------------|--|----------|
| 1 | 10.01 | 1.8656 | Accept |
| 2 | 3.456 | 1.7558 | Accept |
| 3 | 2.215 | 1.6663 | Accept |
| 4 | 1.336 | 1.6004 | Reject |
| 5 | 1.141 | 1.5395 | Reject |
| 6 | 1.087 | 1.4834 | Reject |
| 7 | 1.015 | 1.4248 | Reject |

The three-component solution explained a total of 43.58% of the variance, with Component 1 contributing 27.8%, Component 2 contributing 9.6% and Component 3 contributing 6%. To aid in the interpretation of these three components, oblimin rotation was performed. The rotated solution showed the presence of simple structure (Thurstone, 1947), with three components showing a number of strong loadings and all variables loading substantially on only one component. There were moderately weak correlations between these three factors (Table 4). The results of this analysis supports the interpretation of routines, implicatures and speech acts as separate aspects of pragmatic competence as suggested by Roever (2006), although he also found one extra component which accounted for difficult items.

Table 4: Component Correlation Matrix

| | Routines (C-test) | Implicature | Speech acts |
|--------------------|-------------------|-------------|-------------|
| <i>Routines</i> | - | - | - |
| <i>Implicature</i> | 0.27 | - | - |
| <i>Speech acts</i> | 0.41 | 0.28 | - |

Natives versus non-natives

In order to investigate the presence of possible significant statistical difference between native speakers (n=19) and nonnative speakers (n=215), who were classified into four proficiency groups (A2, B1, B2, C1), the entire Participants' scores for three different test sections and the total pragmatic test were subjected to one-way ANOVA tests.

Table 5: Robust Test of Equality of Means

| Sections | | Statistic | df1 | df2 | Sig |
|-----------------------|----------------|-----------|-----|--------|-------|
| <i>C-Test</i> | Welch | 52.30 | 4 | 81.52 | 0.000 |
| | Brown-Forsythe | 16.05 | 4 | 157.45 | 0.000 |
| <i>Implicature</i> | Welch | 64.75 | 4 | 84.86 | 0.000 |
| | Brown-Forsythe | 13.64 | 4 | 182.70 | 0.000 |
| <i>Speech acts</i> | Welch | 36.52 | 4 | 83.66 | 0.000 |
| | Brown-Forsythe | 11.23 | 4 | 165.28 | 0.000 |
| <i>Pragmatic Test</i> | Welch | 85.98 | 4 | 84.12 | 0.000 |
| | Brown-Forsythe | 21.06 | 4 | 153.48 | 0.000 |

Research Article

The results are presented in table 5. Since the assumption of the homogeneity of variance was violated (the Levene statistic revealed unequal variance) for all variables, Robust Tests of Equality of Means (Welch and Brown-Forsythe) statistic and Tamhane *post hoc* test were inspected (Table s). Tamhane Post hoc test revealed that there existed significant differences ($p <.001$) between the native speakers group and all the other proficiency groups for each pragmatic section (routines, implicature, and speech acts) and the pragmatic test in too.

Proficiency Groups

With regard to the relationship between general proficiency and pragmatic competence, Roever (2005) showed that EFL learners’ comprehension increases with L2 proficiency. In order to investigate the presence of possible significant statistical difference between proficiency groups (A2, B1, B2, C1), the entire participants’ scores for three different test sections and the total pragmatic test were subjected to one-way ANOVA tests. Except for the implicature section which violated the assumption of the homogeneity of variance (the Levene statistic revealed unequal variance), the other sections showed to have equal variance. The results are presented in table 6.

Table 6: Robust Test of Equality of Means

| Sections | | Statistic | df1 | df2 | Sig |
|-------------|----------------|-----------|-----|--------|-------|
| Implicature | Welch | 3.74 | 3 | 78.64 | 0.014 |
| | Brown-Forsythe | 3.64 | 3 | 177.21 | 0.014 |

Tamhane Post hoc test revealed that there were no significant differences between the proficiency groups for implicate section. In routine section, Tukey HSD *post hoc* test revealed that the C1 (Advanced) group outperformed the first two proficiency groups (A2 and B1), but not the B2 group. This is in contrast with Roever’s (2012) findings that proficiency cannot be a predictor of ELF learners’ knowledge about routine formulas. This discrepancy can be accounted for by taking C-test as an integrative assessment of the general language proficiency construct as suggested by Eckes and Grotjahn (2006). Further, this may be because of the fact that Roever had taken year of instruction as the indicator of general proficiency instead of giving a proficiency or assessment test in order to form genuine proficiency groups. For speech section, it was found that both C1 and B2 groups were significantly different from A2 group (Table 7). This is quite in line with other studies (Trosborg, 1987; Takahashi & Dufon, 1989; Hill, 1997; Rose, 2000; Achiba, 2002; Roever, 2006). And finally as a general pragmatic assessment, it can be claimed that the participants’ pragmatic knowledge increased with proficiency (B2 and C1 group pragmatic knowledge was significantly different from both A2 and B1 proficiency groups). Results of proficiency group differences are presented in table 7 and the descriptive are shown in Appendix B.

Table 7: ANOVA of group differences in terms of proficiency

| Section | | Sum of Squares | df | Mean Square | F | Sig |
|----------------|----------------|----------------|-----|-------------|------|-------|
| Routines | Between Groups | 168.38 | 3 | 56.13 | 3.94 | 0.009 |
| | Within Groups | 3006.25 | 211 | 14.25 | - | - |
| | Total | 3174.62 | 214 | - | - | - |
| Speech acts | Between Groups | 169.46 | 3 | 56.49 | 4.13 | 0.007 |
| | Within Groups | 2884.26 | 211 | 136.70 | - | - |
| | Total | 3053.72 | 214 | - | - | - |
| Pragmatic Test | Between Groups | 1464.162 | 3 | 488.05 | 5.90 | 0.001 |
| | Within Groups | 17430.72 | 211 | 82.61 | - | - |
| | Total | 18894.884 | 214 | - | - | - |

Research Article

Idiosyncratic versus Formulaic Implicature

After conducting a paired-sample t-test, it was also found that the pragmatic performance of 215 Iranian participants in idiosyncratic implicature ($M = 4.31$, $SD = 2.75$) items was significantly higher than their performance in formulaic implicature ($M = 2.16$, $SD = 1.51$) items, $t(214) = 17.47$, $p < .0005$ (two-tailed), $Sig = 0.000$. The mean difference was 2.15 with a 95% confidence interval ranging from 1.90 to 2.39. The eta squared statistics (.59) indicated a large effect size. This is quite in keeping with Bouton's (1988) findings.

C-test versus Multiple-choice Test for Routines

In order to discover concurrent validity, the newly adapted test and the original pragmatic test were subjected to correlation statistics. There was a strong, positive correlation between the two variables, $r = .57$, $n = 215$, $p < 0.0005$, with high levels of c-test scores associated with high scores of multiple-choice test. The coefficient of determination equals .32 which means that two types of routine test accounts for 32 percent of each other's variance, which is quite high and shows that the two variables overlap to a considerable degree.

On the other hand, the original pragmatic test showed to be significantly correlated with the newly adapted pragmatic test. There was a strong positive correlation between the two tests, $r = .95$, $n = 215$, $p < .0005$. The coefficient of determination amounts to .90 which shows that 90% of variance in each variable can be explained by the other variable. These findings suggest that the newly adapted test is highly predictive of the participants' performance on the original pragmatic test prepared by Roever (2006).

Discussion

The purpose of this study was to investigate the validity of an adapted pragmatic test with Iranian participants. Reliability is a prerequisite to test validity, thus as the first step in this research the reliability of separate test sections and the whole test was ensured. It can be seen that the low reliability reported for the routine section in Roever's (2006) research has been compensated by introducing a C-test section. Factor analysis and reliability test revealed that when the three pragmatic subcomponents showed to be moderately correlated, they also explained variance of their own. It was also discovered that native speakers outperformed other non-native proficiency groups in each section of the pragmatic test and also in the adapted pragmatic test as a whole.

Similar to Roever's findings, the highest overlap existed between implicates and speech acts, and the lowest overlap was found between routines and implicates. As claimed by Roever, the relationship between speech acts and routines can be explained by the formulaic nature of speech acts. Roever believes that the weak relationship between routines and implicates is due to the fact that when the first one is subordinated to general proficiency and deduction, the second one depends on experience and memorization. A further claim can be made that although implicate is more of a receptive nature, the other two pragmatic components represent more productive pragmatic constructs. In the present study, routines were measured through C-test which required even more production by the participants. Dissimilar to Roever's findings, in this study it was found that except for the implicate section, proficiency groups were significantly different in their score of routine and speech act sections. Although the nature of C-test as a measure of general proficiency may be responsible for this discrepancy, it seems that in addition to taking proficiency as the major determining factor in contrasting routines with the other two constructs, the productive versus receptive nature of these constructs should also be taken into account. This difference is quite conceivable considering the general contrast between pragmatic passive knowledge and active performance.

Regarding implicate, the results found are consistent with Bouton's (1988) findings. The participants performed better on idiosyncratic implicates compared with formulaic implicates. On the other hand, the participants with higher proficiency were not more successful in answering the implicate section. This is not consistent with Roever's findings. The difference between Roever's findings and the results in this study can be attributed to the fact that when Roever's routine test was in the form of multiple-choice, the C-test format of the routine test in this study called for more general proficiency with regard to vocabulary and word dictation. Thus, in this study the knowledge about routines was influenced by the

Research Article

general proficiency of the participants and this is in keeping with the belief that C-tests measure general language proficiency (Dörnyei and Katona, 1992; Kelin-Braley and Raatz, 1984; Klein-Braley, 1985).

Since in this study the multiple-choice version of speech act section was used instead of WDCT it can be claimed that the influence of proficiency on pragmatic production in Roever's DCT has been shifted from the more receptive speech act section to the productive C-test section in this study. Thus dictation, vocabulary and writing ability may represent construct-irrelevant variance in the participants' performance in C-test section. Considering the simplicity and high frequency of missing words and providing the number of missing letters for each word in the C-test section, it can be claimed that the construct-irrelevant variance has been moderately kept under control. What is more, since all the participants in this study were majoring in TEFL/TESL at MA level, part of this discrepancy in the results of the implicate section can be attributed to the fact that the participants in this study promoted themselves in both general proficiency and knowledge about routines through their exposure to ELT materials.

On the other hand, since they were more aware of the indirect and illocutionary forces behind utterances, they less resorted to their general proficiency in detecting the indirect meanings.

The pragmatic test in this study showed to have concurrent validity. It was found that there was a moderately strong relationship between C-test and MDCT tests of routines with a considerable amount of overlap between the two types of routine tests. It was also discovered that the two versions of the pragmatic test overlapped to a high degree and that the newly adapted test is highly predictive of the participants' performance on the original pragmatic test developed by Roever (2006).

With regard to speech acts, the results in this study are in agreement with the findings in similar studies. It was discovered that general proficiency was a determining factor in the participants' knowledge about speech acts. Low proficiency learners are expected to be less capable of recognizing and using appropriate speech acts which appropriately fit the situation. This is quite understandable and attributable to test method, since high proficiency language learners are expected to be better analyzers of linguistic features in any speech and this provides them with higher chances of opting for the best option. On the other hand, since the participants in this study were TEFL/TESL university students at MA level, who might have had experience in teaching these speech acts.

Limitations

The major limitation of this research was that the participants voluntarily decided to take part in this study. Those who participated may be different from others with regard to their motivation and interest in pragmatic test. Considering the fact that practicality has been a major concern in recent arguments on language assessment (Qin & Pan, 2011), this study is limited in that it does not take practicality into consideration and this makes the obtained results tentative and the conclusions drawn subject to further complementary investigation. As is the case with Roever's pragmatic test (and all other language proficiency tests), the present study was not comprehensive with regard to bringing all aspects of pragmatic competence into play. What is more, despite all the advantages in assessing part of the participants' pragmatic competence, C-test section has introduced the construct irrelevant variance into the results. The two other sections also suffer from all the shortcomings pertained to MDCTs. Changing all sections of the pragmatic test into C-test may be advantageous in eliminating some of the major shortcomings of pragmatic tests, but this contention await further research. Finally the participants' familiarity with Web-based browsers, as one possible construct-irrelevant variable, was not taken into control.

Conclusion

This study was performed in an attempt to validate an adapted pragmatic test developed by Roever (2006). In general, the reliability of the test was confirmed, its construct and concurrent-oriented validity was ensured and its adaptability with related research was investigated. Although the use of C-test as one possible form of pragmatic test was proposed for the first time, it is recommended that further research be done in order to better delineate the effectiveness of this test format in promoting the validity of pragmatic tests.

Research Article

Given the fact that any pragmatic test, at best, suggests the potentials of language learners in effective language use, more performance oriented pragmatic tests are required which promote the quality of inferences made about language learners' actual use of their pragmatic knowledge based on pragmatic test results. Although Grabowski (2007, 2008a) provides us with such a test, due to practicality concerns, the need is still felt for further research which satisfies major validity concerns such as construct underrepresentation, construct irrelevant variables, practicality, reliability, etc.

Considering the effectiveness of C-tests in providing the opportunity for both higher objectivity and more real and authentic performance in language testing based on real situations which require less subjective and uncertain mediation by native speakers, a new pragmatic test can be developed which takes advantage of C-test for all the sections of pragmatic tests including routines, implicates and speech acts. Research on the role of C-tests in assessing pragmatic competence can be promising in bridging the gaps between practicality issues and validity underrepresentation challenges in the field of pragmatic tests. This study can be replicated with larger and more diverse participants.

REFERENCES

- Achiba M (2002).** *Learning to Request in a Second Language: Child Inter Language Pragmatics*. Clevedon: Multilingual Matters.
- Bartlett MS (1954).** A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society* **16**(Series B) 296-8.
- Bouton L (1988).** A cross-cultural study of ability to interpret implicatures in English. *World Englishes* **17** 183-196.
- Bouton L (1994).** Conversational implicature in the second language: Learned slowly when not deliberately taught. *Journal of Pragmatics* **22**(2) 157-67.
- Bouton L (1994).** Conversational implicature in the second language: Learned slowly when not deliberately taught. *Journal of Pragmatics* **22**(2) 157-67.
- Bouton LF (1990).** The effective use of implicature in English: Why and how it should be taught in the ESL classroom. In: *Pragmatics and Language Learning*, edited by Bouton LF and Kachru Y, University of Illinois, Urbana-Champaign: Division of English as an International Language 43-51.
- Brown JD (2001).** Six types of pragmatics tests in two different contexts. In: *Pragmatics in Language Teaching*, edited by Rose K and Kasper G (Cambridge University Press) New York 301–25.
- Čolović-Marković J (2012).** The effect of explicit instruction of formulaic sequences on second language writers (Unpublished doctoral dissertation). University of Utah, USA.
- Dörnyei Z and Katona L (1992).** Validation of the C-test amongst Hungarian EFL learners. *Language Testing* **9**(2) 187–206.
- Ebrahimi-Bazzaz F, Ab-Samad A, Arif-bin-Esmail I and Noordin N (2012).** Measuring collocational competence of Iranian learners by using C-test. *The Iranian EFL Journal* **8**(6) 227-240.
- Eckes T and Grotjahn R (2006).** A closer look at the construct validity of C-tests. *Language Testing* **23**(3) 290-325.
- Garcia P (2004).** Pragmatic comprehension of high and low level language. *TESLEJ* **8**(2).
- Grabowski KC (2007).** *Reconsidering the Measurement of Pragmatic Knowledge using a Reciprocal Written Task Format*. Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics **7**(1) 1-48.
- Grabowski KC (2008b).** Measuring pragmatic knowledge: Issues of construct, underrepresentation or labeling? *Language Assessment Quarterly* **5**(2) 154-159.
- Grice HP (1975).** Logic and conversation. In: *Syntax and Semantics, : Speech Acts*, edited by Cole P and Morgan JL (Academic Press) New York **3** 41-58.
- Hill T (1997).** The development of pragmatic competence in an EFL context. Unpublished doctoral dissertation, Temple University, Japan.

Research Article

- House J and Kasper G (1987).** Interlanguage pragmatics: requesting in a foreign language. In: *Perspectives on Language in Performance*, edited by Lo`rscher W and Schulze R, Narr, Tu`bingen 1250-1288.
- Hudson T, Detmer E and Brown JD (1995).** Developing prototypic measures of cross-cultural pragmatics (Tech. Rep.No. 7). Honolulu: University of Hawai`i, Second Language Teaching and Curriculum Center.
- Ikoma T (1993).** Sorry for giving me a ride: The use of apologetic expressions to show gratitude in Japanese. Unpublished PhD thesis, University of Hawai`i.
- Kaiser HF (1970).** A second generation little jiffy. *Psychometrika* **35** 401-415.
- Klein-Braley C (1985).** A cloze-up on the C-test: A study in the construct validation of authentic tests. *Language Testing* **2**(1) 76-104.
- Klein-Braley C (1997).** 1997: C-Tests in the context of reduced redundancy testing: an appraisal. *Language Testing* **14** 47–84.
- Klein-Braley Cand Raatz U (1984).** A survey of research on the C-Test. *Language Testing* **1**(2) 134-146.
- Liu J (2006).** *Measuring Interlanguage Pragmatic Knowledge of EFL Learners*. Frankfurt: Peter Lang.
- McNamara T and Roever C (2006).** *Language Testing: The Social Dimension* (MA: Blackwell) Malden.
- Oller JW Jr (1973).** Cloze tests of second language proficiency and what they measure. *Language Learning* **23**(1) 105 -1 8.
- Olshtain E and Weinbach L (1987).** Complaints: A study of speech act behavior among native and nonnative speakers of Hebrew. In: *The Pragmatic Perspective*, edited by Papi M and Vershueren J (Amsterdam: Benjamin) 195-208.
- Qin DD and Pan M (2011).** A practical approach to fostering language assessment literacy practical language testing, by Glenn Fulcher. *Language Assessment Quarterly* **8**(3) 309-312.
- Read J (2000).** *Assessing Vocabulary* (CambridgeUniversity Press) Cambridge.
- Roever C (1994).** Linguistische Routinen beim Fremdsprachenerwerb des Englischen: Bedeutung, Probleme und Perspektiven [Linguistic routines in the acquisition of English as a foreign language: significance, problems and perspectives]. Unpublished MA thesis, University of Duisburg, Germany.
- Roever C (2001).** A web-based test of interlanguage pragmalinguistic knowledge: Speech acts, routines, and implicatures. Unpublished PhD thesis, University of Hawai`i.
- Roever C (2006).** Validation of a web-based test of ESL pragmalinguistics. *Language Testing* **23**(2) 229-255.
- Roever C (2012).** What learners get for free: learning of routine formulae in ESL and EFL environments. *ELT Journal* **66**(1) 10-21.
- Rose K (2000).** An exploratory cross-sectional study of interlanguage pragmatic development. *Studies in Second Language Acquisition* **22** 27–67.
- Schmitt N, Dornyei Z, Adolphs S and Durow V (2004).** Knowledge and acquisition of formulaic sequences: A longitudinal study. In: *Formulaic Sequences: Acquisition, Processing and Use*, edited by Schmitt N (Amsterdam: John Benjamins) 55-86.
- Takahashi S and DuFon MA (1989).** Cross-linguistic influence in indirectness: the case of English directives performed by native Japanese speakers. *ERIC Document* ED370439.
- Takahashi T and Beebe LM (1987).** The development of pragmatic competence by Japanese learners of English. *JALT Journal* **8** 131-155.
- Thurstone LL (1947).** *Multiple-factor Analysis* (University of Chicago Press) Chicago.
- Trosborg A (1987).** Apology strategies in natives/non-natives. *Journal of Pragmatics* **11** 147–67.
- Trosborg A (1995).** *Interlanguage Pragmatics* (Mouton de Gruyter) Berlin.
- Yamashita S (2008).** Investigating interlanguage pragmatic ability: What are we testing? In: *Investigating Pragmatics In Foreign Language Learning, Teaching And Testing*, Soler EA and Matinez-Flor A (Multilingual Matters) New York 201-221.

Research Article

Yamashita SO (1996a). Comparing six cross-cultural pragmatics measures. Unpublished PhD thesis, Temple University, Philadelphia.

Yamashita SO (1996b). Six measures of JSL Pragmatics (Technical Report, 14). Honolulu: University of Hawai'i at Manoa, Second Language Teaching and Curriculum Centre.

Yoshitake SS (1997). Measuring interlanguage pragmatic competence of Japanese students of English as a foreign language: A multi-test framework evaluation. Unpublished doctoral dissertation, Columbia Pacific University, Novata, CA.