Reliability and Content Validity of a Comprehensive Discourse Completion Test of Interlanguage Pragmatics for Academic Situations

Mehrdad Amiri, Parviz Birjandi
Islamic Azad University, Tehran, Iran

The present study aimed at investigating the reliability and content validity of a comprehensive test of Interlanguage Pragmatics (ILP) developed for the academic situations in the Iranian context. To do so, a meta-analysis of 50 articles which had reported on the tests of ILP was done and the criteria required for the selection of the speech acts of academic situations were developed. Then a Multiple-Choice Discourse Completion Test (MDCT) of interlanguage pragmatics was developed and validated through native checked situations and items based on meta-pragmatics procedures. Participants of the study were 50 M.A. students majoring in TEFL (Teaching English as a Foreign Language) studying at Iranian universities as well as 50 native speakers of English. The findings revealed that the test developed enjoyed high internal consistency and content validity. The present findings could contribute to the fields of second language testing and assessment in general and testing interlanguage pragmatics in particular.

Keywords: discourse completion test, interlanguage pragmatics, academic situations, MDCT, speech acts

Introduction

Tests of interlanguage pragmatic competence, which mainly have focused on various speech acts, have been developed since 1980s in various cultural settings (Jianda, 2007). The testing system developed to accomplish the purpose is labeled as Multiple-Choice Discourse Completion Test (MDCT). Preparing MDCTs requires a thorough meta-pragmatic assessment of situations, scenarios, and items.


Mehrdad Amiri, Ph.D. student, Department of Languages and Literature, Science and Research Branch, Islamic Azad University, Tehran, Iran.
Parviz Birjandi, Ph.D., Department of Languages and Literature, Science and Research Branch, Islamic Azad University, Tehran, Iran. (Corresponding Author: pbirjandi@yahoo.com.)
1 The MDCT developed will be sent to the interested individuals upon request.
been a matter of concern for the test developers (Jianda, 2006). Tests of ILP are used to measure the non-native speakers’ developmental knowledge of language pragmatics and appropriate use of second language. In case it is ignored, effective communication cannot be achieved, and that is why pragmatics and pragmatic testing come into the picture. In line with pragmatic resting research, developing an academic test of pragmatics, which mainly focuses on the academic situations, is a new research topic. The present study is an attempt to investigate the internal consistency and content validity of an MDCT developed for the ILP of academic situations. Therefore, the research questions formulated are as follows:

(1) Do the components of the MDCT developed at the academic level contribute to the reliability of the test?

(2) Does the MDCT developed have expert judgment validity based on content-related evidence?

**Method**

**Participants**

The participants of the study in the process of test development were 100 individuals: 50 M.A. students majoring in TEFL studying at Tehran universities as well as 50 educated native speakers of English dealing with academic activities in their respective countries who were requested to take part in the study. The 50 M.A. students taking part in this phase of the study, as the main participants, were selected based on the results of a standard version of TOEFL PBT (released by ETS in 2013) which was administered to ninety M.A. candidates of TEFL in Tehran universities. The students whose scores fell one SD below and above the mean were selected as the main participants of the study.

**Instrumentation**

Data collection instruments which were employed in the present study were a standard version of TOEFL PBT (2013) and a validated, researcher-made, and comprehensive academic MDCT of interlanguage pragmatics.

**Procedure**

First, a standard version of TOEFL test (released by ETS in 2013) was administered to 90 M.A. students majoring in TEFL in some universities of Tehran and the researchers could select 50 students whose scores fell one SD below and above the mean as the main participants of the study. Then, they were briefed about the purpose of the research and were asked to take part in the process of developing and validating the academic MDCT of ILP which included recognizing the situations, presenting essay type responses to the questions about various situations, taking tests, presenting think-aloud protocols, etc. Fifty North American native speakers of English dealing with academic activities were also asked to take part in the first phase of the study as mentioned above to accomplish the procedure of developing an academic version of MDCT of interlanguage pragmatics. These participants checked the situations presented for the discourse completion purpose and completed them. Then they answered the multiple choice test developed.

**Data Analysis**

Reliability

The first research question of the study was set to see if the components of the MDCT developed at the
academic level contribute the same to the reliability of the test. Table 1 displays the Cronbach’s alpha reliability indices for the total and components of MDCT. The total MDCT had a reliability index of .67. The reliability indices for the components of MDCT ranged from .55 for complaint to .78 for disagreement, considering the fact that there were only six items in each section.

Table 1
Reliability Statistics; Total Data

<table>
<thead>
<tr>
<th>Speech acts</th>
<th>Cronbach’s alpha</th>
<th>N of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apology</td>
<td>.646</td>
<td>6</td>
</tr>
<tr>
<td>Request</td>
<td>.664</td>
<td>6</td>
</tr>
<tr>
<td>Refusal</td>
<td>.747</td>
<td>6</td>
</tr>
<tr>
<td>Complaint</td>
<td>.554</td>
<td>6</td>
</tr>
<tr>
<td>Disagreement</td>
<td>.782</td>
<td>6</td>
</tr>
<tr>
<td>Total MDCT</td>
<td>.678</td>
<td>30</td>
</tr>
</tbody>
</table>

Thus, the first null-hypothesis as “components of the MDCT developed at the academic level do not contribute to the reliability of the test” was rejected. Then it was concluded that the test enjoyed relatively high reliability.

Validity

The second research question of the study aimed at investigating expert judgment validity of the MDCT developed based on content-related evidence. Berk (1990) argues that content-related evidence of validity is a central concern during “instrument” development and expert professional judgment should play an integral part in developing the test items and definition of what is to be measured. A Multivariate Analysis of Variance (MANOVA) was run to compare speech acts of request, refusal, apology, complaint, and disagreement according to the native speakers’ performance. Based on the results, the native speakers had the highest mean on request (M = 14.74, SE = .30, 95% CI [14.13, 15.34]). This was followed by refusal (M = 14.32, SE = .25, 95% CI [13.79, 14.84]), apology (M = 12.88, SE = .34, 95% CI [12.18, 13.57]), complaint (M = 12.66, SE = .29, 95% CI [12.06, 13.25]), and disagreement (M = 12.44, SE = .29, 95% CI [11.84, 13.03]). Since the assumption of sphericity-homogeneity of variances for the differences between any two tests was not met (Mauchly’s W = .68, p < .05), the results of the multivariate test was reported. Based on the results displayed in Table 2, (F (4, 46) = 10.74, p < .05, Partial $\eta^2 = .48$ representing a large effect size), it was concluded that there were significant differences between the native speakers’ performance on the MDCT test and its components. Thus, the second null-hypothesis as “the MDCT developed does not have expert judgment validity based on content-related evidence” was rejected.

Table 2
Multivariate Tests; MDCT (Native Speakers)

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pillai’s trace</td>
<td>.483</td>
<td>10.745</td>
<td>4</td>
<td>46</td>
<td>.000</td>
<td>.483</td>
</tr>
<tr>
<td>Wilks’ lambda</td>
<td>.517</td>
<td>10.745</td>
<td>4</td>
<td>46</td>
<td>.000</td>
<td>.483</td>
</tr>
<tr>
<td>Hotelling’s trace</td>
<td>.934</td>
<td>10.745</td>
<td>4</td>
<td>46</td>
<td>.000</td>
<td>.483</td>
</tr>
<tr>
<td>Roy’s largest root</td>
<td>.934</td>
<td>10.745</td>
<td>4</td>
<td>46</td>
<td>.000</td>
<td>.483</td>
</tr>
</tbody>
</table>

A post-hoc comparison test was run to show the whereabouts of the test components. The results of
RELIABILITY AND CONTENT VALIDITY

post-hoc comparison test indicated that: (1) There was not any significant difference between the native speakers’ means on request (M = 14.74) and refusal (M = 14.32) (MD = .42, p > .05, 95 % CI [-.53, 1.37]); (2) The native speakers’ mean on request (M = 14.74) was significantly higher than their mean on apology (M = 12.88) (MD = 1.86, p < .05, 95 % CI [.77, 2.94]); (3) The native speakers’ mean on request (M = 14.74) was significantly higher than their mean on complaint (M = 12.66) (MD = 2.08, p < .05, 95 % CI [.73, 3.42]); (4) The native speakers’ mean on request (M = 14.74) was significantly higher than their mean on disagreement (M = 12.88) (MD = 1.86, p < .05, 95 % CI [.77, 2.94]); (5) The native speakers’ mean on refusal (M = 14.32) was significantly higher than their mean on apology (M = 12.88) (MD = 1.44, p < .05, 95 % CI [.35, 2.52]); (6) The native speakers’ mean on request (M = 14.32) was significantly higher than their mean on complaint (M = 12.66) (MD = 1.66, p < .05, 95 % CI [.57, 2.74]); (7) The native speakers’ mean on request (M = 14.74) was significantly higher than their mean on disagreement (M = 12.44) (MD = 1.88, p < .05, 95 % CI [.79, 2.96]); (8) There was not any significant difference between the native speakers’ means on apology (M = 12.88) and complaint (M = 12.66) (MD = .22, p > .05, 95 % CI [-.93, 1.37]); and (9) There was not any significant difference between the native speakers’ means on complaint (M = 12.66) and disagreement (M = 12.44) (MD = .22, p > .05, 95 % CI [-.69, 1.13]).

Discussion

The findings of the present study revealed that the MDCT developed enjoyed high reliability and expert judgment validity. It was also found that components of the MDCT developed at the academic level highly contributed to the reliability of the test, though their contribution was not the same. The results revealed significant differences between the components of pragmatic test (MDCT) in the academic level. In fact, native speakers had the highest mean on request speech act followed in turn by the speech acts of refusal, apology, complaint, and disagreement. Secondly, findings revealed that the MDCT developed enjoyed high expert judgment validity based on content-related evidence. According to Berk (1990), content-related evidence of validity is a central concern during “instrument” development and expert professional judgment should play an integral part in developing the test items and definition of what is to be measured. The present study findings revealed a significant relationship between the rank orders of speech acts in the pragmatic test (MDCT), as the highest correlation was between apology and disagreement which was followed by the relationships between complaint and disagreement, request and refusal, request and apology, and refusal and apology. However, the MDCT of request and apology did not show any significant relationships with complaint and disagreement, nor did the MDCT of apology show significant correlation with that of complaint. All in all, the items of the speech acts developed and used in the present study enjoy the total high reliability of .67, based on Cronbach’s Alpha (α = 0.67).

Conclusion

It was revealed that components of the developed MDCT at the academic level highly contributed to the reliability of the test, though their contribution was not the same. This signifies that internal consistency of components of a test is as important as the reliability of the whole test (Bachman & Palmer, 2010). Stability of the test results in different administrations also revealed high reliability of the test developed (Bachman, 1990), meanwhile, the native speakers’ data had a significantly higher reliability index. The reason might lie in the fact
that familiarity of the native speakers with real language situations and the speech acts used in such situations to get the meaning conveyed is surely higher than that of the non-native speakers. Taguchi (2006) argued that native speakers have got mastery over the situations and the communicative acts taking place. Berk (1990) argues that content-related evidence of validity is a central concern during “instrument” development and expert professional judgment should play an integral part in developing the test items and definition of what is to be measured. As expert judgment validity of the developed MCDT was proved, it can be concluded that the test enjoys a high content validity.

References


