Assessing the Pragmatic Competence of Students

in an Intensive English Program

Altynai Abdukarimova and Anastasiya Bezborodova

Northern Arizona University

Abstract

A proficiency test was administered to 10 high- intermediate level ESL students at the Intensive English for Academic Purposes Program at Northern Arizona University. This report introduces a revised test description, methods, and results of statistical analysis. This report describes the proficiency test designed to examine students' pragmatic competence. The test focused on the students' level of proficiency in sociocultural and psychological meanings. The results collected from 10 test tasks were scored by two raters, the designers of the test. Raters scored the test based on the analytical rubric designed for this test and the scores for both tasks showed relatively good inter-rater reliability. From the analysis and discussion, it is seen that the overall quality of this test is acceptable, and this test achieved its purpose. However, there are some modifications that can be applied for future uses.

Keywords: interactive discourse completion test, sociocultural meaning, psychological meaning, pragmatic competence, proficiency test

Background

Assessment of second language pragmatic knowledge is an understudied area of pragmatics. However, few language teachers address this important issue in language teaching and testing. Thus, tests of pragmatic knowledge are few both in high-stakes testing as well as in most language programs. Some researchers have attempted to assess pragmatic knowledge, but they followed a speech-act framework often criticized for pragmatic construct under-representation (Yamashita, 2008; Roever, 2011). For this test, test designers choose interactive discourse completion test (IDCT) as their TLU task since it is important and practical for both the class learning and the coming competition (Roever, 2011). As students live in the target language setting, they have to be able to use the language not only linguistically appropriate, but also be able to apologize, request, refuse, and ask for permission. Bardovi-Harlig (1996) states, if a learner is linguistically competent but makes a pragmatic mistake, it is considered as more serious than a grammatical mistake.

The construct definition of a test is based on Purpura's (2004) theoretical model, which specifies that pragmatic knowledge includes contextual, sociolinguistic, sociocultural, psychological, and rhetorical meanings. Since test designers assume that students in Level 6 already know contextual, sociolinguistic, and rhetorical pragmatic meanings, this particular test will test only sociocultural meaning, which consists of cultural meanings and norms (e.g., use of speech acts and formulaic expressions) and psychological meaning, which includes affective stance (e.g., sarcasm, deference, politeness, importance, anger, humor, criticism) (Purpura, 2004). Thus, sociocultural and psychological meanings function as sub-constructs. The knowledge of grammatical form and grammatical meaning is assumed. Spelling and punctuation will not be tested.

In this test, test designers have made a hypothesis about the correlation between Part 1 and Part 2 of the test. That was, a test taker's ability in "speaking on paper" tasks, Part 1, would indicate his/her understanding of pragmatic relationship and performance in a role –play, Part 2. Therefore, if a test taker received a relatively good score in Part 1, we believed he or she would receive the same level of score in Part 2.

Research Questions

This study addresses the following research questions:

1. Are the students of high proficiency level, who live and study in the second language context, pragmatically competent?

2. Do they need additional instruction in the components of pragmatic knowledge?

Methods

There were 10 students involved in the test administration. They were from one section of Level 6, Listening and Speaking class in PIE. The participants were Chinese, Arabic, Brazilian, Korean, and one Taiwanese students, aged 20 - 28.

The test consisted of two parts; each part had different number of items and different assessing criteria. Part 1 consisted of two IDCT tasks given to a pair of students. The prompts were given visually, as students had to read the instructions on the handout and write a series of response exchanges on the paper in the form of dialogue. Students then switched roles, writing the same dialogue but with the opposite role assigned. This task was explained to students as "speaking on paper" (Ahn, 2005). Part 2 used one IDCT in role-play (Hudson, 2001). It was used in order to operationalize the reciprocal effect of context and meaning in the form of

conversation. For Part 2, the response was oral, where students working in the same pairs were

acting out only one dialog with emotions and feelings. Their oral responses were recorded by

themselves using PIE recorders. Students were familiar with this procedure and test designers did

not have to spend any time on explaining how to use recorders.

This task tested the following criteria: content, turn-taking, cohesion of argument,

structure, speech acts, message, and cooperation. All variables were rated on a five-point scale

(0-4). Therefore, the total number of items developed for Part 1 (sociocultural meaning) was two.

However, students were then switching roles and were asked to write two dialogues all in all.

Each dialogue was evaluated according to the criteria stated above. Also, students' ability to

cooperate was evaluated on three-point scale (0-2) in all three tasks.

Results

The statistics that were computed for the test were inter-rater reliability, descriptive

statistics, reliability and standard error of measurement, and description of masters and non-

masters.

The correlations between two raters were estimated for each criterion on the rubric. The

average correlation for task 1 was relatively high. It fell into the interval from .74, the lowest, to

1, the highest, with an exception for criterion of structure. It had the lowest correlation of .46.

The average correlation for task 2 was lower and fell into the interval from .58 to 1. However,

the structure in task 2 was significantly higher than for task one. The inter-rater reliability for

"role-play", task 3, was .69.

The descriptive statistics for the test scores on three tasks and the total score are

illustrated in table below including number of items, the mean, standard deviation, minimum,

and maximum of the scores (Table 2). The relationship is also depicted in the histogram (Figure

1). This test was taken by ten test takers. Among these ten people, the highest score they received

for the total of tasks was 52.50, and the lowest one was 37.50. The mean of all these scores was

46.5, and the Standard Deviation (SD) was 5.17. The maximum possible score for Task 1 and

Task 2 was 26 and 56 for both tasks.

The highest score in Task 1 was 25 and the lowest one was 15. Task 2 general

performance was higher and minimum score is 19. Though, the highest score is the same with

Task 1, 25 points. The mean of Task 1 was 20.6, and SD was 3.33. From the table above it is

seen that the mean of two tasks was close to each other. However, students performed better on

Task 2. The mean of Task 2 was 22.7 and SD was 2.48. On Task 3 some students got the highest

score of 4 and the lowest was 2.5 with SD .48 and mean 3.2

Table 1

Descriptive Statistics

Test	K	Mean	Standard Deviation	Minimum	Maximum
Total	3	46.50	5.17	37.50	52.50
Task 1	1	20.60	3.33	15.00	25.00
Task 2	1	22.70	2.42	19.00	25.00
Task 3	1	3.20	.48	2.50	4.00

Note. N=10

The reliability (r), standard error of measurement (SEM), and agreement coefficient (p_o)

were calculated and the results are illustrated in the table below (Table 3). Reliability result for

the test is equal .53, which means that the test was not reliable. Based to this reliability score, the

SEM was 3.54. Accordingly to Subkoviak (1988), with the test reliability, the agreement

coefficient was .79.

Relevance to PIE and Second Language Learning

The decisions that were made on the interpretations of the results were low-stakes. There was not negative or positive classification since test takers were not classified on the results of their performances. The entirely decision, that was made, was whether or not to focus on pragmatic competence for a particular group of test takers, level 6 at PIE. According to the results of the test, the decision was made that level 6 PIE students did not need pragmatic competence to be addressed, since they were able to perform their knowledge and ability in pragmatics. However, it would be interesting to conduct research for lower level students and to see if pragmatic competence should be addressed in their classes. If this type of research is conducted with low level students, and if the results show that low level students need explicit pragmatic instruction, then the language program and the institution might be affected in terms of developing new curriculum and incorporating pragmatics to the program as well as training teachers how to instruct pragmatics.

Related Readings

Alderson, J. C. (1991). Bands and scores. In J. C. Alderson & B. North. *Language testing in the*

     1990s. London: Macmillan.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. New York, NY: Oxford

     University Press.

Bardovi-Harlig, K. (1996). Pragmatics and language teaching: Bringing pragmatics and

     pedagogy together. *Pragmatics and language learning, 7,* 21-39.

Hudson, T. (2005). Trends in Assessment Scales and Criterion-Referenced Language

     Assessment**.** *Annual Review of Applied Linguistics, 25,* 205-227.

Jungheim, N. O. (2001). The unspoken element of communicative competence: Evaluating

     language learners' nonverbal behavior. In T. Hudson & J. D. Brown (Eds.), *A focus on*

     language test development: Expanding the language proficiency construct across a

     variety of tests (pp. 1-34). Honolulu: University of Hawai'i, Second Language Teaching

     and Curriculum Center.

Korsko, P. (2004). *The narrative shape of two-party complaints in Portuguese: A discourse*

     analytic study. Unpublished doctoral dissertation, Teachers College, Columbia

     University, New York City.

Lightbown, P., & Spada, N. (2006). *How languages are learned* (3rd ed.). New York, NY:

     Oxford University Press.

Liu, J. (2007). Developing a Pragmatics Test for Chinese EFL Learners. *Language Testing*, *24*,

     391-415.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–
        103). New York: American Council on Education and Macmillan Publishing Company.

Miller, M. D., Linn, R., & Gronlund, N. (2009). *Measurement and evaluation in teaching.* (Tenth
        Edition). Upper Saddle River, NJ: Merrill, Prentice Hall.

Purpura, J. (2004). *Assessing grammar.* New York: Cambridge University Press.

Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing,
        28(4),* 463-481.

Stoynoff, S., & Chapelle, C. (2005). *ESOL Tests and Testing*. Alexandria, VA: TESOL.

Subkoviak, M. (1988). A practitioner's guide to computation and interpretation of reliability
        indices for mastery tests. *Journal of Educational Measurement, 25,* 47-55.

Wolfson, N. (1989). Perspectives: Sociolinguistics and TESOL. Cambirdge, MA.

Yamashita, S. (2008). Investigating interlanguage pragmatic ability: What are we testing? In E.
        Alcon Soler & A. Martinez-Flor (Eds.), *Second language acquisition research series:
        Vol: 30. Investigating pragmatics in foreign language learning, teaching and testing* (pp.
        201-223). Bristol, UK: Multilingual Matters.