

A STUDY OF THE EFFECT OF PROFICIENCY LEVELS IN PAIRED TAPE -MEDIATED ORAL TESTS

Ma Ruijuan
North University of China

Abstract

The present study focuses its attention on the influence of test takers language proficiency levels . It intends to investigate whether different language proficiency levels may result in differences in a candidate s performance in paired speaking tests that involve role play activity . Results from test scores and language analysis suggest evidence of a partner proficiency level effect for low proficiency group , with candidates achieving higher scores and gram matical accuracy when working with a high proficiency partner . However language analysis shows that there is no effect on gram matical complexity for them when they are paired at different proficiency levels . Furthermore for high proficiency groups there is no great change in terms of test scores , the analysis of gram matical complexity and gram matical accuracy whether their partner is a high - proficiency candidate or a low proficiency one .

Key words

performance variables ; tape -mediated oral tests ; gram matical complexity ; gram matical accuracy

1 . Introduction

Recent research in the area of language testing has highlighted the importance of performance variables . O Sullivan (2002a) saw test performance as being affected by a number of factors , related to the test taker , the interlocutor and the task . The effect of test taker characteristics on candidate performance in tests of language proficiency was studied by many researchers (Kunnan 1995 ; Purpura 1998 ; O Sullivan 2002a) . They focused on the effect of characteristics such as education level , sex , acquaintanceship and motivation , and their findings suggested that these can account for variability in test performance . However there are a number of aspects of proficiency tests that remain little explored . Language proficiency level as one performance variable , researchers paid little attention to it , especially the influence of one candidate s proficiency level on the other s performance . Anecdotal evidence from language teachers and testers suggests that the partner s proficiency level in an activity which involves interaction might affect the other candidate s performance on that task . This observation needs to be examined in a language testing situation . For both low proficiency and high proficiency candidates , when they are paired with a person considered to be a different proficiency level partner , will they perform differently ?

The related topic in this study remains largely unexplored . The following ones are only some studies about language proficiency level . Purpura (1998) studied the difference of strategy use between high - ability and low ability test takers . This study uses structural equation modeling to examine the relationships between strategy use and second language test performance with high and low ability test takers . The results showed that the metacognitive strategy use and second language test performance models produced almost identical factorial structures for each group , while the cognitive strategy use models were somewhat different . Norris et al . (2002) indicated that a specific task can make inference

about test takers' abilities to carry out the specific task. However, inference cannot be made about their abilities with other tasks. Many researchers investigated the relationship between interviewer competence and ratings (Morton et al., 1997; McNamara & Lumley, 1997). These studies found that raters rated higher when the interviewers were less competent or not so good at establishing rapport with the candidates.

The role of grammar in influencing candidates' performance was supported by Wilds (1975) and Raffaldini (1988). O'Sullivan (2002a) explored the scores awarded by trained raters and the language in terms of grammatical complexity and grammatical accuracy in pair task performance of test takers' familiarity with their partners. Based on the factors analyzed by O'Sullivan (2002a) in pair task performance, grammatical accuracy, grammatical complexity, and candidates' performance were investigated in the study.

This study will restrict other variables and mainly investigate the influence of language competence on participants' performance. Grammatical complexity and grammatical accuracy are important terms. They are used as basic criteria to judge performance differences in this study. According to Foster and Skehan (1996:30),

Grammatical complexity: Total number of clauses divided by total c-units. Here clauses are defined as being "either a simple independent finite clause or a dependent finite or nonfinite clause. A c-unit is defined as each independent utterance providing referential or pragmatic meaning. Thus, a c-unit may be made up of one simple independent finite clause or else an independent finite clause plus one or more dependent finite or nonfinite clause."

Grammatical accuracy: The percentage of error-free clauses. Where an error-free clause was defined as "[A] clause in which there is no error in syntax, morphology, or word order. Errors in lexis were counted when a word used was incontrovertibly wrong. In case of appropriacy, no error was recorded."

Based on the previous research, the following research questions will be specifically addressed in the present study:

- 1) In terms of test scores, how are candidates' performances influenced by different pairings for both groups?
- 2) In terms of grammatical complexity, how are candidates' performances influenced by different pairings for both groups?
- 3) In terms of grammatical accuracy, how are candidates' performances influenced by different pairings for both groups?

2. Research design

2.1 Participants

The subjects participating in this study are 66 second-year English majors from the School of Foreign Languages at Shanxi University. Of the 66 subjects, 56 are female and 10 are male. Their previous study in English ranges from 8 to 10 years and they have taken almost the same courses during their two-year college English study. They took part in Oral Test, TEM-4 (Test for English Majors Grade 4) just several weeks before this study, so this can guarantee no obvious change in their oral language proficiency levels to exclude the learning effect in their final scores. According to their scores of Oral Test, TEM-4 and the purpose of this study, the subjects are divided into low proficiency and high proficiency groups and paired with different partners. Then the Simulated Oral Proficiency Test (SOPT) was investigated. All the data from the SOPT were analyzed for scores and language in terms of grammatical complexity and grammatical accuracy.

2.2 Tasks

The students were asked to participate in two activities, one alone (not rated) and the other with a partner to play a role (see Appendix).

This role play task was expected to be the most cognitively challenging as the topic was quite clear and it involved a degree of negotiation and compromise. The language accuracy might be affected as the students focused more on their response to the role while the output was expected to be more interactive in nature. It was also assumed to suit either male or female candidates, or to be neutral. All candidates were college students and each student's partner was from the same class, so they were familiar with the topic as well as the partner and had common interest. Of course they shared the same culture background and had something to say about the role play task.

Performance on role play task was rated using an analytic rating scale and a holistic rating scale at the same time. With four levels, it included the following scoring criteria in both rating scales: pronunciation, grammar, vocabulary, fluency, and content.

2.3 Procedures

Participants took the SOPT in three laboratories during the same two-hour class period at Shanxi University. They were paired according to the required pairings in this study and asked to finish the test with their partners within four minutes. All the SOPT data were tape recorded and the interactions were transcribed and analyzed for the analysis of the language.

2.4 Rating and reliability

Trained raters were asked to score the performances from the tapes. All performances were scored five times, with two raters using the holistic rating scale and three raters using the analytic rating scale. The rating scales adopted in this study were from Oral, TEM 4 and they were revised slightly according to the requirement of this study. To calculate the final score for an individual's performance, the scores from the holistic scale and the analytic scale were averaged. The reliability is shown in Table 1 and it indicates a significant correlation as well as a high reliability.

Table 1. Correlation analysis for all raters

	R1	R2	R3	R4	R5
R1	1.000	.748	.856	.907	.763
R2	.748	1.000	.812	.742	.897
R3	.856	.812	1.000	.819	.819
R4	.907	.742	.819	1.000	.770
R5	.763	.897	.819	.770	1.000

3. Results and discussion

3.1 Differences in measured test scores

After obtaining the high reliability, One-Way ANOVA of test scores for both groups were carried out. The results of the present study were analyzed by using SPSS. In all the analyses reported below, the level of confidence for rejection of the null hypothesis was .05 and actual p values were reported.

Table 2 shows that when the low proficiency participants are paired with high proficiency partners, the minimum and maximum scores as well as the mean scores are higher than those scores when they are paired with low proficiency partners. It indicates a tendency of higher scores for low proficiency participants when they are paired with high proficiency partners. For the high proficiency group, Table 2 describes similar scores in terms of mean, minimum and maximum scores. Though when they are paired with high proficiency partners the mean score is a little higher than the other circumstance, there is no significance. Table 3 can be concluded that low proficiency participants will perform differently when paired at different proficiency levels ($p = .006 < .05$). But for high proficiency participants, Table 3 indicates that there is no significant difference ($p = .645 > .05$).

Table 2. Descriptive statistics of test scores for low proficiency and high proficiency groups

	pairings	N	Mean	Minimum	Maximum
Low proficiency group	Low -Low proficiency	12	61 ,9667	50 ,40	74 ,60
	Low -High proficiency	13	69 ,8769	57 ,60	81 ,60
High proficiency group	High -Low proficiency	13	71 ,8145	62 ,00	82 ,00
	High -High proficiency	28	72 ,7429	63 ,00	83 ,20

Table 3. One -Way ANOVA of test scores for low proficiency and high proficiency groups

Low proficiency group		High proficiency group	
F	Sig .	F	Sig .
9 ,063	,006	,216	,645

3.2 Differences in language characteristics in terms of grammatical complexity and accuracy

In order to explore the effect on the complexity and accuracy of the language produced by participants under different conditions ,the performances were transcribed and analyzed . All the oral performances were transcribed ,analyzed and coded for clauses ,error free clauses . In this part total numbers of clauses were divided by clauses with errors and error free clauses . The samples were the same as the analysis of test scores ,the transcription data were coded and the figures for grammatical complexity and accuracy were then calculated and analyzed using One -way ANOVA .

Table 4. One -Way ANOVA of grammatical complexity and accuracy for low proficiency and high proficiency groups

	Low proficiency group		High proficiency group	
	F	Sig .	F	Sig .
gram matical complexity	,285	,598	,032	,859
gram matical accuracy	181 ,583	,000	,066	,799

Table 4 shows that there is no statistical significance in gram matical complexity for low proficiency group and high proficiency group when they are paired at different proficiency levels($p = .598 > .05$, $p = .859 > .05$). Table 4 indicates great statistical change in grammatical accuracy for low proficiency group in different conditions ($p = .000 < .05$) ,and no statistical significance for high proficiency ($p = .799 > .05$).

4 . Conclusion

Oral tests are influenced by many performance variables . This study intends to investigate the influence of language proficiency levels in a tape -mediated speaking test . In the SOPTstudy ,to test this performance variable ,all other variables were controlled and tested to make sure that any difference in interactions came from the influence of language proficiency levels .

4.1 For low proficiency group

The findings indicate that significant differences in measured test scores occur when low proficiency candidates engage in one-to-one interactions with high proficiency partners. This study suggests that for low proficiency candidates, they will perform differently when paired at different proficiency levels and will achieve higher ratings when working with a high proficiency partner. There is no significant change in the participants' grammatical complexity in different pairing conditions for low proficiency group. For language grammatical accuracy, low proficiency group shows great statistical significance under different conditions. And there is a tendency of higher grammatical accuracy for low proficiency participants when they are paired with high proficiency partners.

4.2 For high proficiency group

The findings show that there is no significant change in the measured test scores in different pairing conditions for high proficiency group. It also indicates no statistical significance in the participants' grammatical complexity and accuracy when they are paired with different candidates.

4.3 Further explanations about the findings

The findings relating to no change of grammatical complexity could be explained from three aspects: 1) In high stakes tests, test takers will use their language conservatively; 2) There is not sufficient mention of grammatical complexity in rating scales; 3) The proficiency level of the participants in this study is relatively low. Perhaps it can be rationalized that lack of variation in the complexity of the language under the two conditions for both low and high proficiency groups — since complexity is associated with risk taking and is not associated with formality of interactions.

The findings relating to the accuracy of the language are also of interest. It should be recognized that grammatical accuracy is just one aspect of overall linguistic performance. One aim of the present study is to provide empirical evidence for language testing. It was thought that for low proficiency group, superiority in performance when interacting with a high proficiency candidate would be seen, and this superiority would be seen in both the scores awarded by the raters and in the accuracy of the specific language employed by the participants. The results of the analyses of the accuracy of the language used in different interactions indicate a level of proficiency effect for low proficiency group. Though accuracy is normally associated with formality, in this case it cannot be claimed that for low proficiency group the interaction involving high proficiency partner is more formal than the one involving two low proficiency candidates. We can hypothesize that high proficiency candidates do not dominate the interactions. High-proficiency candidates will use strategies and accommodate low proficiency candidates to carry out the interactions.

4.4 Implications for oral language teaching

This study, which deals with the influence of different proficiency levels in oral language tests, may throw some light on the way to help the teaching of Chinese EFL learners' oral pair work. According to some findings in this study, generally speaking, if one talks with a partner of a higher language proficiency level, scores will be higher especially in a testing situation. Hence, when a teacher designs and carries out oral pair work in oral language class, students can be arranged according to teaching purposes. If a teacher wants to improve low proficiency students' oral competence, he/she can arrange them to practice with high proficiency partners to improve their oral language competence. This study shows that they will improve their interactions in terms of scores and grammatical accuracy. High-proficiency students can be arranged with either a high proficiency or low proficiency partner because their performances will not be influenced by different pairings according to the findings of this study.

4.5 Limitations of the present study

Firstly, the samples are relatively small. In this study the participants are limited and all the students of English majors in one grade took part in the study. If the participants can be increased and chosen from more colleges and universities, the findings will be more convincing. So just as the above mentioned, the results of this study may not be readily generalizable outside the context.

Secondly, the division of the groups is not so clear. In this study, participants were divided into two groups: low proficiency group and high proficiency group. Due to lack of better students whose scores of Oral Test, TEM 4 are 4 points, those students of 2 point are regarded as high proficiency group and those students whose scores are 1 point belong to low proficiency group. If 4 point students are classed as high-proficiency group and 1 point students as low proficiency group, the results might be more clear and confirming.

Thirdly, the purpose of this study is to explore the effect of one variable on candidates performance, so the influence of other variables must be controlled well. From the design and execution of this study, other variables were controlled to minimum. But it is difficult and it cannot be concluded that other variables played no role in the final findings. Though efforts are made to do this, the control of other variables perhaps are not so perfect.

4.6 Suggestions for further research

Despite its limitations, the results of this study confirm that any test format that employs tasks requiring interaction between individuals is in need of rather urgent and extensive study. Though the results may not be generalizable readily outside the context, within that context, it appears to be predictable and significant. In terms of performance variables there are some suggestions.

Firstly, when performance variables are examined, the measure of the specific linguistic performance and the rated performance should be taken into account.

Secondly, a similar approach can be used to other variables proposed in the introduction part (personality, status, sex and age, types of tasks, task difficulty, and culture background etc).

Thirdly, in this kind of study, the author should examine closely how and what aspects these variables affect performance. More qualitative investigation about the language should be investigated.

Lastly, further research about this performance variable should investigate the reasons why high-proficiency candidates are not influenced by low proficiency partners performance. Further studies should focus attention on the accommodations and communicative strategies used by high proficiency groups.

References

- Bachman, L.F. 1990. *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, L.F. & A.S. Palmer. 1996. *Language Testing in Practice*. Oxford University Press.
- Foster, P. & P. Skehan. 1996. The influence of planning and task type on second language performance. *Studies in Second Language Acquisition* 18 :299-323.
- Kunnan, A.J. 1995. *Test Taker Characteristics and Performance: A Structural Modeling Approach*. Cambridge University Press.
- McNamara, T.F. & T. Lumley. 1997. The effect of interlocutor and assessment mode variables in overseas assessment of speaking skills in occupational settings. *Language Testing* 14 :140-156.
- Morton, J., G. Wigglesworth & D. Williams. 1997. Approaches to the evaluation of interviewer performance in oral interaction tests. In G. Brindley & G. Wigglesworth (eds.). *Access: Issues in English Language Test Design and Delivery*. Sydney: National Centre for English Language Teaching and Research. pp. 175-196.
- Norris, J.M. et al. 2002. Examinee abilities and task difficulty in task based second language performance assessment. *Language Testing* 19 :395-418.
- O'Sullivan, B. 2000a. *Towards a model of performance in oral language testing*. Unpublished PhD dissertation, University of Reading.
- O'Sullivan, B. 2000b. Exploring gender and oral proficiency interview performance. *System* 28 :1-14.
- O'Sullivan, B. 2002a. Learner acquaintanceship and oral proficiency test pair task performance. *Language Testing* 19 :277-295.
- O'Sullivan, B. 2002b. Using observation checklists to validate speaking test tasks. *Language Testing* 19 :32-51.
- Purpura, J.E. 1998. Investigating the effects of strategy use and second language test performance with high- and low ability test takers: A structural equation modelling approach. *Language Testing* 15 :333-379.

- Raffaldini , T . 1988 . The use of situation tests as measures of communicative ability . Studies in Second Language Acquisition 10 :197-216 .
- Wilds , C . P . 1975 . The oral interview test . In R . L . Jones & B . Spolsky (eds .) . Testing Language Proficiency . Arlington : Center for Applied Linguistics , 29-44 .
- 邹申 , 2005 , 《语言测试》 . 上海 : 上海外语教育出版社 .

Appendix

Test

Instructions

(The instructions are given in writing in the target language . Specification of procedures and tasks is explicit .)

- 1 . This is a test of your oral language proficiency .
- 2 . After you read the directives , finish Task 1 and Task 2 .
- 3 . Try to make your conversation as meaningful and appropriate as possible . Your response will be scored on pronunciation , grammar , vocabulary , fluency and content .
- 4 . You have 10 minutes to finish both the tasks .

Task 1

For this part , you are going to make a self-introduction for about 3 minutes according to the questions recorded on the tape .

Task 2

Hold a dialogue of about 4 minutes based on the following situation . Each of you should read your role card only . You have 3 minutes to prepare by yourself .

Role card for Candidate A :

You are a college student who is busy preparing for your final examinations . You need a quiet place to review your lessons . But your neighbor (Candidate B) always has his TV on loudly . It's very late at night . You cannot put up with the noise and go to complain . Begin by knocking at the door .

Role card for Candidate B :

You are a retired worker whose only recreation is watching TV . Since you are a little deaf , you always have your TV on loudly . It's late at night and you are enjoying your favorite TV series . Suddenly there is someone knocking at the door .

(Zou 2005 : 27)