Interlocutor Proficiency in Paired Speaking Tests

Huei-Chun Teng

National Taiwan University of Science and Technology

The study was aimed at investigating interlocutor proficiency in paired speaking tests. The participants were 20 EFL students from a university in Taiwan. 10 English major students were designated as learners of higher proficiency; the other 10 non-English major students were designated as learners of lower proficiency. A role-play task was adopted for the paired speaking test. The scoring rubric used a scale of 1 to 5 in four subcategories: grammar and vocabulary, pronunciation, fluency, and content. Results showed that lower proficiency participants performed significantly better in fluency, scored higher, and produced more language when talking to partners who had higher EFL proficiency. Most of the interactions in the paired oral test were collaborative, with high mutuality and equality. Through providing empirical evidence, this study can shed some light on the valid administration of EFL paired speaking tests.

　本研究ではペア・スピーキングテストにおける対話者の習熟度を実証的に調査した。20名のEFL台湾人大学生を対象とし、うち10名は英語専攻の高レベル学習者、残り10名は非英語専攻の低レベル学習者であった。ペア・スピーキングテストはロールプレータスクを用いて行った。採点は1から5の5段階で行い、文法及び語彙、発音、流暢さ、内容の4つのサブカテゴリーを設けた。その結果、流暢さにおいて、低レベルの参加者は低レベルのパートナーと話したときよりも高レベルのパートナーと話したときに有意な上達が見られた。また、高レベルのパートナーと話したとき、低レベルの参加者は高得点を獲得した上、より多く発話をした。ペア・オーラルテストの際、インタラクションのほとんどは協力的なものだった。実験的証拠を通じ、本研究はEFLペア・スピーキングテストが有効である可能性を示している。

THERE HAS been increasing use of pair work in second language teaching with the popularity of the communicative approach. The emphasis on oral communication in ESL classrooms has resulted in the growth of paired oral assessment. According to previous research (Bachman, 2001; Kenyon, 1992; McNamara, 1996; Skehan, 1998), a paired oral assessment can be regarded as a performance that is the result of an interaction among candidates as interlocutors, includes tasks, and is judged by raters who apply a rating scale and produce scores. The performance in actual assessment may vary in response to various tasks, or different examinees and interlocutors.

Taylor (2000) pointed out a number of advantages associated with the paired speaking test. First, it makes use of Communicative Language Teaching (CLT) in the classroom and can produce positive washback. Second, it provides various samples of interaction. Third, it is natural for interactions to be slightly asymmetric. Finally, it can make speaking tasks more authentic. Moreover, Együd and Glover (2001) argued in favor of the paired format in terms

of candidate preferences, performance opportunities, quality interaction, washback, and training. However, in spite of the potential benefits of the paired test format, a major concern has been the question of whether it matters who is paired with whom (Fulcher, 2003). Brown and McNamara (2004) claimed that the interlocutor factor is potentially more complex in paired or group oral speaking tests. Thus, this study was aimed at investigating interlocutor proficiency in paired speaking tests by addressing the following research questions:

1. Does the proficiency level of one speaker influence the performance of the other speaker in a paired speaking test and is this reflected in the scoring?

2. What impact does the interlocutor proficiency have on the discourse produced in the test?

Literature Review

In terms of the interaction in paired speaking tests, Canagarajah (2006) suggested that paired interaction broadens the notion of speaking from that of traditional one-to-one interviews. It could be reasonably argued that the paired candidate format changes the notion from one of *speaking* to one of *interactional competence* or even *intercultural communicative competence*. However, while the unpredictability and dynamic nature of the interaction forms the basis of favorable claims by proponents of the paired speaking test, it has also been argued that this unpredictability may compromise test reliability (Brown, 2003). Macqueen and Harding (2009) indicated that the paired format has been criticized for introducing a range of factors that may cause threats to test validity. According to Foot (1999), these factors include the potential for breakdowns in communication between candidates who struggle with each other's variety of speech and the effects of variables such as age, personality, and social class. Besides, Taylor and Wigglesworth (2009) argued that the co-constructed

nature of the interaction and the fact that co-participants' contributions are inextricably linked raises issues of construct definition, reliability, and fairness. Nevertheless, Brooks's (2009) study showed that in the paired format test takers demonstrated their facility in negotiating meaning and communicating with another language learner, co-constructing good and rich performances through their interaction. He proposed that perhaps rather than being viewed as a threat to construct validity, variability in paired interaction can be embraced as being more reflective of real-world communication.

For the past two decades, research has looked at a number of variables associated with the interlocutors, such as acquaintanceship, age, gender, interaction style, personality, proficiency level, and status. Among these interlocutor variables, the current study specifically investigated proficiency. Regarding the effect of interlocutor proficiency on test performance, findings presented in previous research are mixed. For example, in Iwashita's (1996) study, test takers in both groups of high and low proficiency produced more turns and c-units in a two-way task when working with a high-proficiency partner. It was proposed that interlocutor proficiency may influence both the amount of talk produced and scores received in a speaking task. In contrast, Bonk and Van Moere (2004) found that neither the groups' mean proficiency level nor the amount of variation in proficiency levels present in the group appeared to change the scores examinees were expected to receive.

Nakatsuhara (2004) studied the discourse produced by various combinations of higher and lower proficiency candidates performing a problem-solving task and found no differences in the features of interactional contingency, goal orientation, and quantity dominance. The results suggested that differences in proficiency level among candidates had little effect on conversation type. Davis (2009) examined the influence of interlocutor proficiency on speaking performance in a group of Chinese

college freshmen. Findings showed that interlocutor proficiency level had no significant effect on speaking performance, but lower level examinees produced more language when working with a higher level partner. He proposed that examinees' proficiency differences need not preclude use of the paired oral test format.

Paired candidate speaking tests have received relatively little attention in the language testing literature until recently. According to Lazaraton (1996), the role of the interlocutor in the oral assessment context deserves much more empirical attention than it has received to date. To fill the gap of previous research, this study examined the interlocutor proficiency in the context of EFL paired speaking tests for Taiwanese college students.

Methodology

Participants

Participants in the study were 20 students at a university in northern Taiwan. There were 10 students who majored in English and 10 students with other majors. The 10 English majors, who had an average TOEIC score of 840, were designated as learners of higher EFL proficiency, and the other 10 non-English majors, with an average TOEIC score of 530, were designated as EFL learners of lower proficiency. There were 13 female and 7 male participants. They ranged in age from 18 to 23 years old. They had a uniform first language background and had a relatively homogeneous cultural background. All of them had completed at least 8 years of EFL study prior to entering the university.

Evaluation Instruments

The main instruments used in the study consisted of an oral test and a rating scale. The oral test was mainly based on the role-play tasks adopted in Nakatani's (2005) study. The two test takers engaged in a simulated conversation derived from a situation described on a card. Cues for the role-play were taken from Shohamy (1985). Test takers were given 5 minutes to prepare the role-play. There were 40 role-play cards (see Appendix for examples). The other instrument was a rating scale used for scoring the role-play task. The scoring rubric adopted a scale of 1 to 5 in the subcategories of grammar and vocabulary, pronunciation, fluency, and content. Subcategory scores were averaged to produce a final overall raw score.

Procedures

Participants were randomly assigned to subgroups of four that contained two higher proficiency students and two lower proficiency students. Thus, there were five subgroups of four participants in each group. Within each subgroup, pairings were arranged so that each participant was tested once with a partner from the same proficiency group and once with a partner from the other proficiency level. To counterbalance the order effect, 10 participants took the oral test first with the interlocutor of the same proficiency level and then with the interlocutor of the different level. The other 10 participants took the test in the opposite order. Forty role-play cards were provided to each group, from which each pair randomly selected a card. To ensure each participant never used the same card twice, once a card was selected, it was removed from the set of available cards. At the beginning of the test, participants were told in detail what they were required to do. They were informed that the study was designed to obtain empirical information about the paired oral assessment. Each pair was given 5 minutes to prepare the role-play, and then was asked to begin speaking. Performances of the role-play were recorded using digital audio and video recorders for subsequent scoring and analysis.

Data Analysis

All examinees' test performances were scored by two raters who were native speakers of English with graduate degrees in English language teaching or a related field. They had experience teaching a similar population of students and of rating oral proficiency tests. Prior to starting the scoring work, the two raters reviewed the rating scale with the researcher. They were provided with the videotapes of three examples of role-play performances and rated the examples independently. Next, the two raters discussed the ratings and came to an agreement on the scoring. Test takers' scores were based on an average of the two raters' scores. The inter-rater reliability was 0.83 for the overall score. To examine the influence of interlocutor proficiency on the test scores, paired *t* tests were performed on the five dependent variables of overall score: grammar, vocabulary, pronunciation, fluency, and content.

In addition, the researcher examined the discourse produced in the paired oral test. All participants' performances in role-plays were transcribed. The transcriptions were made by English-major graduate students and checked by the researcher. Once the transcripts were completed, the quantity of language produced in each candidate's performance was measured by counting the number of words spoken by the individual participant.

In order to explore the elicited interactional patterns, candidate discourse was analyzed using Galaczi's (2008) framework (see Table 1), which characterizes discourse into collaborative, parallel, or asymmetric interaction types. Interactional type discourse was defined with two variables, equality and mutuality. Two raters who were familiar with discourse analysis were provided with explanatory information taken from Galaczi, written transcriptions, and examples. Following the initial coding, the two raters discussed any cases of disagreement until a consensus was reached regarding the final coding.

Table 1. Framework of Interaction Types*

Type	Dimension	Interactional Characteristics (Example)
Collaborative	High mutuality	Development of topics over several turns.
	High equality	Balanced quantity of talk between the two participants.
Parallel	Low mutuality	Lack of listener support, rare acknowledgement tokens.
	High equality	Balance between the two participants in the use of topic initiation and expansion moves.
Asymmetric	Moderate mutuality	Development of topic mainly by one person. Rare expansion of other-initiated topics
	Low equality	Lack of balance in terms of topic initiation and extension moves.

Collaborative interaction is characterized as having self-initiated and other-initiated topic expansion and a balanced quantity of talk between the two participants. Parallel interaction includes frequent initiation moves, a lack of listener support, and talk balanced between the two participants. In asymmetric interaction, there is topic development mainly by one person and an unbalanced quantity of talk.

Results

Table 2 shows the 20 participants' scores in the paired speaking test, and the results of paired *t* tests for oral test scores are shown in Table 3. Findings indicate that the participants per-

formed better in fluency when talking to the partners who had higher EFL proficiency than they themselves had.

Table 2. Descriptive Statistics of Paired Speaking Test Scores

Performance Category	Speaker							
	Lower Proficiency ($N = 10$)				Higher Proficiency ($N = 10$)			
	Lower int		Higher int		Lower int		Higher int	
	M	SD	M	SD	M	SD	M	SD
Grammar	3.27	0.34	3.50	0.42	4.10	0.32	4.17	0.24
Pronunciation	3.60	0.34	3.80	0.39	4.27	0.41	4.23	0.27
Fluency	3.30	0.37	3.70	0.40	4.20	0.32	4.33	0.31
Content	3.53	0.42	3.87	0.55	4.27	0.47	4.27	0.14
Overall average	3.43	0.34	3.72	0.35	4.21	0.33	4.25	0.20

Note. int = interlocutor

Table 3. Speaking Test Scores: *t* Tests

Performance category	Speaker							
	Lower proficiency				Higher proficiency			
	df	D	t	p-value	df	D	t	p-value
Grammar	18	0.23	1.35	0.19	18	0.07	0.53	0.60
Pronunciation	18	0.20	1.21	0.24	18	-0.04	-0.21	0.83
Fluency	18	0.40	2.33	0.03*	18	0.13	0.94	0.36
Content	18	0.33	1.52	0.15	18	0.00	0.00	1.00
Overall average	18	0.29	1.89	0.07	18	0.04	0.34	0.74

Note. D = mean difference; *$p < 0.05$.

Results show that higher proficiency participants paired with higher proficiency interlocutors had the highest scores in the three subcategories of grammar and vocabulary, fluency,

content, and total score. The second highest test scores were produced by higher proficiency speakers when paired with lower proficiency interlocutors. Moreover, lower proficiency participants paired with lower proficiency interlocutors got the lowest scores in all categories. Among the four subcategories and total score, there was significant difference in the subcategory score of fluency for lower proficiency participants.

Table 4. Words Produced in Speaking Test

Measure	Speaker			
	Lower proficiency ($N = 10$)		Higher proficiency ($N = 10$)	
	Lower int	Higher int	Lower int	Higher int
M	47.00	65.90	103.30	106.90
SD	14.04	29.33	56.84	29.79
Rank	4	3	2	1

Note. int = interlocutor.

In addition, the quantity of language produced in each candidate's performance was measured by counting the number of words spoken by the individual participant. Table 4 shows that higher proficiency participants paired with higher proficiency interlocutors produced the largest amount of discourse. Lower proficiency participants paired with lower proficiency interlocutors produced the least amount of discourse.

In order to explore the elicited interactional patterns, candidate discourse was analyzed with Galaczi's (2008) framework (see Table 1), which characterized discourse in terms of collaborative, asymmetric, or parallel interaction types. Table 5 indicates that collaborative is the most frequent interaction type identified in the discourse of the paired speaking test for both higher proficiency and lower proficiency participants. Moreover,

the asymmetric type is also identified in the interaction between two candidates of different EFL proficiency.

Table 5. Frequency of Interaction Type

Interaction type	Speaker							
	Lower proficiency (N = 10)				Higher proficiency (N = 10)			
	Lower int		Higher int		Lower int		Higher int	
	f	Rank	f	Rank	f	Rank	f	Rank
Collaborative	9	1	7	1	7	1	10	1
Parallel	1	2	1	3	1	3	0	2
Asymmetric	0	3	2	2	2	2	0	2

Note. int = interlocutor ; f = frequency; Rank = ranking of the three interaction types.

Discussion

The first research question was to examine whether the proficiency level of one speaker influences the performance of the other speaker in a paired speaking test. Results indicate that lower proficiency participants performed significantly better in fluency when talking to partners who had higher EFL proficiency than they had. The findings of the study support those of previous research (Cao, 2011; Davis, 2009; Iwashita, 1996). Test takers in low proficiency groups also produced more language when working with a higher level interlocutor. Moreover, the EFL students seemed to prefer to talk to interlocutors who were more competent than they were. Thus, it is proposed that interlocutor proficiency of one partner may influence the fluency subscale of test scores received and the amount of talk produced by the other partner in paired oral assessment.

However, mixed findings were presented in previous studies regarding the impact of interlocutor proficiency on oral test performance. It is likely that proficiency is fluid, based on who we are talking to in L2 and what sorts of identities are mediated in the interaction (Lazaraton & Davis, 2008). Besides proficiency level, other interlocutor variables might affect L2 learners' performances in paired speaking tests. Since the influence of interlocutor proficiency has not been verified or concluded yet, it is suggested that paired oral tests can be conducted in L2 classrooms regardless of examinees' proficiency differences.

According to the present study, among the total 40 interactions in the paired speaking test, 33 were collaborative and only four were asymmetric. That is, most of the interactions in the paired speaking test were of high equality and mutuality, which can benefit the interaction. As a result, the paired speaking test can achieve the function of communicative testing in terms of collaborative interaction.

Conclusion

The present research found that most of the interactions in the paired oral test taken by EFL learners were collaborative. Also, interlocutor proficiency may influence their received test scores and amount of produced talk. Lower proficiency students performed better when paired with higher proficiency students while the students with higher proficiency performed worse when paired with students of lower proficiency level. By providing empirical evidence, it is hoped that this study has provided some pedagogical implications for EFL teachers. When evaluating students' speaking proficiency, teachers can adopt the paired oral assessment to make the test authentic and interactive. However, students should be paired with the partners with similar or higher EFL proficiency so that they can produce more talk and speak more fluently.

Bio Data

Huei-Chun Teng is currently a professor in the Department of Applied Foreign Languages at National Taiwan University of Science and Technology. Her research interests include L2 listening research, communication study, learning strategy, and language assessment. <tenghc@mail.ntust.edu.tw>

References

Bachman, L. F. (2001, February). *Speaking as a realization of communicative competence*. Paper presented at the meeting of the American Association of Applied Linguistics—International Language Testing Association (AAAL-ILTA) Symposium, St. Louis, MO.

Bonk, W. J., & Van Moere, A. (2004). *L2 group oral testing: The influence of shyness / outgoingness, match or interlocutors' proficiency level, and gender on individual scores*. Paper presented at the Language Testing Research Colloquium, Temecula, CA.

Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26, 341-366.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20, 1-25.

Brown, A., & McNamara, T. (2004). "The devil is in the detail": Researching gender issues in language assessment. *TESOL Quarterly*, 38, 524-538.

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly*, 3, 229-242.

Cao, Y. (2011). Investigating situational willingness to communicate within second language classrooms from an ecological perspective. *System*, 39, 468-479.

Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing*, 26, 367-396.

Együd, G., & Glover, P. (2001). Oral test in pairs: A secondary school perspective. *ELT Journal*, 55, 70-76.

Foot, M. C. (1999). Relaxing in pairs. *ELT Journal*, 53, 36-51.

Fulcher, G. (2003). *Testing second language speaking*. London: Pearson Longman.

Galaczi, E. D. (2008). Peer–peer interaction in a speaking test: The case of the first certificate in English examination. *Language Assessment Quarterly*, 5(2), 89-119.

Iwashita, N. (1996). The validity of the paired interview format in oral performance assessment. *Melbourne Papers in Language Testing*, 5(2), 51-66.

Kenyon, D. (1992, February). *Introductory remarks at a symposium on development and use of rating scales in language testing*. Paper presented at the Language Testing Research Colloquium, Vancouver, Canada.

Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing*, 13, 151-172.

Lazaraton, A, & Davis, L. (2008). A microanalytic perspective on discourse, proficiency, and identity in paired oral assessment. *Language Assessment Quarterly*, 5, 313-335.

Macqueen, S., & Harding, L. (2009). Review of the certificate of proficiency in English (CPE) speaking test. *Language Testing*, 26, 467-475.

McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.

Nakatani, Y. (2005). The effects of awareness-raising training on oral communication strategy use. *The Modern Language Journal, 89*, 76-91.

Nakatsuhara, F. (2004). *An investigation into conversation styles in paired speaking tests* (Unpublished master's thesis). University of Essex, UK.

Shohamy, E. (1985). *A practical handbook in language testing for the second language teacher*. Tel-Aviv, Israel: Tel-Aviv University.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Taylor, L. (2000). Investigating the paired speaking test format. *Cambridge ESOL Research Notes*, 2, 14-15.

Taylor, L., & Wigglesworth, G. (2009). Are two heads better than one? Pair work in L2 assessment contexts. *Language Testing*, 26, 325-339.

Appendix

Examples of Role-Play Cards

No.	Roles	Situation
1	You: A young girl Partner: A young boy	You meet someone you met at a party last week. He invites you to a movie. You can't accept his invitation. Make excuses.
2	You: Teenager Partner: Mother	You go shopping with your mother to buy some clothes for yourself but she wants to make all the choices for you. You disagree and explain why.
3	You: A client Partner: Car-agent manager	You signed a contract for a new car. Prices went up and you are asked to add the difference. You complain to the manager.
4	You: A pedestrian Partner: Policeman	You are on your way to an important meeting. You crossed the road on a red light. The policeman wants to give you a ticket. Try to persuade him to let you off.