Assessment in an Era of Rapid Change: Innovations and Best Practices

Presenter: Arzami Bin Salim
Education Officer of Innova Junior College
Ministry of Education, Singapore

Title of Paper: The Role of Questioning Prompts in Conversation: A Study of
Pupils' Performance in a Secondary School EL Oral
Assessment.

Abstract

Pupils' performance in the English Language oral examination is never an isolated outcome. The examiner's role would in some ways, affect how well a student performs. A detailed analysis of the discourse characteristics of the oral interactions could help to highlight this relationship. This study considers the effect of examiners' prompts in the conversation session of the examination. Based on conversational analyses of 17 full-length oral interactions, it is argued that assessments of students' performance are reflected in the extent of promptings by the examiners and that the variation in results may provide a useful understanding in validating proficiency and providing a parameter for examiners' conduct. In general, the findings of this study suggest that the examiner plays an important role in students' display of oral proficiency. Apart from the questionings, factors like linguistics, pedagogical and social, interact in influencing the performance of test takers.

1   Introduction

It seems that students still face difficulty in talking freely about a topic in an examination context in spite of spending at least ten years of education in English. Although there are many possible reasons for this difficulty (e.g., social and cultural; pedagogical; personal) that go beyond the immediate assessment situation, there may also be factors within the assessment procedures themselves that could contribute to the disappointing performance. In the context of this study, `performance' refers to the marks students achieved and the linguistic skills (like turn-taking, pauses, contribution of spoken words) the students exhibited in the interactions.

2   Aims of Research

This research examines the role of the examiners' questioning in the conversation session of the English Language oral assessment of secondary school students in Singapore. It aims to achieve a better and more critical understanding of how and why elicitation questions play a role in determining a student's oral language performance, in the assessment process. The knowledge derived from this research could provide new insights for language practitioners like teachers, language academics and education policy makers into considering the validity of the conversation assessment in the oral examination.

3   Statement of Problem

In the English Language (EL) oral exam, examiners evaluate among other skills, the students' ability to sustain a meaningful conversation. However, the different amount of attention and response that examiners themselves give the students, can have an effect on the students' overall oral performance.

4   Research Questions

General Question

Does the amount and nature of prompts given by the examiner affect the oral performance of students in the conversation part of the oral exam?

Specific Questions

What is the impact on the overall interaction and the grading of the student's performance if conditions in the examiner's prompting are varied in the following manner:

1. **Condition A -** No verbal prompts are provided by the examiner;
2. **Condition B -** Minimum prompts are provided as suggested in the examination paper; and
3. **Condition C -** Extended (maximum) promptings are supplied by the examiner, including prompts that build upon what is said by the candidate.

The analysis under the three different conditions focuses on the following questions:

a.   What forms of prompts are used in an EL conversation oral assessment?

b.   How much will the student response vary (in terms of length of time and utterances)?

c.   What forms of prompting encourage students to perform better in the oral interaction?

5   Research Background

Many language studies have pointed to the importance of feedback in oral interaction and how it affects the overall performance of the test takers. Although the studies reviewed have highlighted the presence of examiner's support (Lazaraton, 1999) and candidate's difficulty under different task conditions (as in Fulcher and Reiter, 2003), there are no observations made that critically discuss the significance of rating in the absence or presence of examiner's support under varying task conditions.

This study could be useful in providing further understanding of the oral assessment process especially in accommodating candidates' utterances to help them perform better. Finally, it would be useful to note from the results of this study, the best form of (if any) and the optimal extent of examiner's support that would help candidates to perform to the best of their ability.

6   Experiment One (Actual Oral Examination)
6.1   Method
6.1.2   Subjects

The subjects were 9 Secondary Three students. They had a year more before sitting for their GCE `O' and `N' Level examination. These Secondary Three students were 15 years old, comprising five girls and four boys. All these students were grouped (or streamed, as it is termed) by the school according to their different levels of learning abilities. In the school under study, the cohort of Express students (High Ability) had an average T-score of

201[1] whereas the Normal students' average T-score was 168 (Low Ability) and could rise to a maximum of 190[2]. Nevertheless, the grouping could show overlaps or minimal difference in T-scores, especially between the mid and high ability students. This is inevitable as other factors (like performance in other subjects) are taken into consideration when the students are streamed in Secondary One. This is shown in the data in Table 1. Students 3 and 4 have only a 1 mark difference even though they are in two different ability groups (high and mid ability).

A total of three groups of three students were studied in an actual oral examination setting. The three groups comprised one group of three students from the best Express class (High Ability), one group of three students from an average Express class (Mid Ability) and a final group of three Normal Academic students (Low Ability). The detailed description of each student profile is appended in Table 1 below:

Table 1: Student Profile of Actual Exam (Experiment One)

Student No.	Gender/Age	T-Score/PSLE EL Grade	Ability Level
1.	Male/15	230/B	High Ability
2.	Male/15	203/B	
3.	Female/15	199/B	
4.	Female/15	198/B	Average Ability
5.	Male/15	196/C	
6.	Female/15	191/B	
7.	Female/15	162/C	Low Ability
8.	Female/15	178/C	
9.	Male/15	162/C	

6.1.3 Task

The year-end oral examination for the Secondary Three students was conducted in similar manner to the oral exam that students take in their GCE

---

[1] The T-Score is a measurement derived from the results a student achieved in his or her PSLE examination. It is the total aggregate marks of four subjects; Mathematics, English, Science and Mother Tongue (1st native language), taken together. The total possible mark of 400 is computed to 300. Students with a T-Score of 230 -260 can be considered to have a high learning ability as they are mostly channelled to the Express and/or Special Stream.

[2] This figure could vary from school to school, depending on the cohorts of students that were enrolled. The figure for the school under study is provided by the Department Head of the school.

`O' and `N' Level examination. In the oral exam, the students are given a passage to read aloud to the examiner who would then engage in a short conversation with the latter on a topic which is thematically related to the passage the students had read.

The examiner-student interactions of this group were transcribed and analysed for discourse features like pauses and turn-taking.

6.1.4 Procedure

Students were given a passage to read silently for about 5-10 minutes. The purpose of this reading session was for the student to have an understanding of the content before being asked to read it aloud and talk about the issues that revolved around the theme of the passage. After the silent reading, the student read the passage aloud in the presence of the examiner. Three examiners were involved in the actual exam studied. Each examiner was assigned by the school to assess a particular ability group. Hence, in the three groups studied, one examiner was assigned to assess the low ability students while the remaining two examiners evaluated the mid and high ability students respectively. Once this was completed, the student had to describe in detail a photograph which was attached to the passage. Upon completing this task, the examiner prompted the candidate to talk about a topic. The passage, picture and the conversation topic are all thematically related.

At the end of the session, the examiner awarded the appropriate mark for each component of the assessment (the maximum mark the candidate could possibly score is 40, comprising 12 marks each for reading and picture description and 16 marks for the conversation component.) When doing so, the examiner referred to the scoring rubric (Oral Mark Scheme) provided for the examination. For the purpose of this study, the time taken in the conversation session was also noted.

As the voice recordings are the main source of data for analysis, transcriptions were made to allow for detailed discussion and analysis of the verbal interactions. Since accurate representation of speech sounds is not the focus of this study, phonological or phonetic transcriptions of data were not

used. Attempts were made to transcribe the data as accurately as possible since the analysis depends heavily on the transcription.

7   Discussion

Data in the actual exam shows that the academic ability of students does not necessarily correlate with actual performance. In the exam, the higher ability students scored lower marks than the lower ability students. As shown in Table 2 below, the high ability students scored between 8 and 9 marks whereas those who were of lower ability scored between 11 and 13 marks. This finding seems to suggest that there could be other possible factors involved in determining how well the students perform in the oral exam. One factor which is not explored in Experiment One (actual exam), is the amount of prompt that examiners provided to the students. The discussion of the data in the simulated exam (Experiment Two) will elaborate on this issue.

Table 2: Summary of Scores in Actual Exam

Student no.	Ability level	Conversation score (Total - 16)	Average score	Time* taken in conversation	Average time taken
1	High Ability	9	9	2.09	2.15
2		8		2.24	
3		9		1.31	
4	Mid Ability	13	11.3	1.56	1.63
5		9		2.33	
6		12		1.00	
7	Low Ability	12	12	1.54	1.37
8		11		1.23	
9		13		1.33	

* Time is recorded in minutes and seconds.

With a high turn number in an interaction, we would expect rich and meaningful interaction which in turn would possibly translate into a good perception of the candidates' performance. The actual exam data did not clearly conform to this expectation. Although the highest turn number was produced by the high ability students, they scored lower than the other groups and spent less time conversing in each turn. This group of students could have compromised their scoring ability by spending less time on their turns. This tendency could have given the examiners fewer opportunities to evaluate their ideas, thereby affecting the score.

There were two overlaps that occurred in the actual exam interactions which suggest that the verbal transactions were generally quite `smooth' and structured. These overlaps were also initiated by the examiners rather than the students. The `interruptions' caused by the overlaps suggest the presence of a clear boundary of social power between the interlocutors. The students did not at any time, attempt to interrupt the examiners but would instead wait for their turn to speak. This is normally signalled by the examiners' questions.

Apart from the minimal overlaps, the equal distribution of question and answer in the interactions also point to a structured form of interactions in the actual exam. The data also show that the students had contributed more (in terms of the amount of words spoken) compared to the examiners. Though it is tempting to conclude that the interaction format in the actual exam is actually an interview due to its consistent question-answer exchange structure, the high portion of students involvement in the interactions requires careful consideration. The need for control on the part of the examiners could have resulted in the question and answer exchange pattern but at the same time, there appears to be a parallel requirement for the students to contribute a large sample of their oral expressions in order for a reasonable evaluation of their communicative competence to be made. The linguistic evidence seems to suggest that the exchanges that took place in the exam were more of a blend of an interview and a conversation.

In the study on conversation and oral interview, Lazaraton (2002) seems to suggest that there may not be a clear distinction between these two genres. The evidences discussed here suggest that the oral exam may be regarded as a blend of an interview and a conversation. The combined presence of an institutional goal in mind (evaluating students' communicative competence), which is a norm in an interview, and the locally managed turn-taking patterns commonly observed in casual conversation, support this conclusion.

7 Experiment Two (Simulated Oral Exam)

7.1 Method

7.1.1 Subjects

For this simulated exam, the three groups of students selected were different from those in Experiment One. The first two groups were from the high and low ability groups respectively. The sole examiner of each group in Experiment One played the same role as examiner in Experiment Two. The rationale for doing this was to ensure (as far as possible) that other factors remain constant except for the prompting conditions. The third group of three students was selected from a Secondary One class to represent the mid ability group (some difficulty in getting the mid ability students from the same Secondary Three class studied in Experiment One due to school schedule). This situation would be taken into consideration when the analysis of the study results is made.

The three groups comprise four 15 year old girls, two 15 year old boys, two 13 year old girls and one 13 year old boy. The detailed description of each student profile is shown in Table 3 below.

Table 3: Student Profile of Simulated Exam (Experiment Two)

Student no.	Gender/Age	T-Score/PSLE** EL grade	Ability level
10	Female/15	233/A	High Ability
11	Female/15	220/A*	
12	Male/15	223/A	
13	Female/13	190/B	Mid Ability
14	Male/13	186/B	
15	Female/13	188/B	
16	Female/15	179/C	Low Ability
17	Female/15	166/C	
18	Male/15	182/C	

Note: * Read as `A star' (a grade above distinction)

** Primary School Leaving Examination (PSLE) is the English grade obtained by students in the PSLE .

7.1.2 Task

The same set of exam papers used in Experiment One was tested on these students so as to maintain a sense of consistency in the test items. At the same time, the data gathered in Experiment Two would enable valid comparisons to be made with the students in Experiment One. A new group of students were studied as it is felt that familiarity with the questions would compromise the authenticity of responses (Fulcher and Reiter, 2003).

Like in Experiment One, the same examiner was assigned to one ability group of students. This was meant to ensure that all variables (except the prompting conditions) were kept constant so that the students' performance could be argued on the basis of the differences in prompting conditions. In each group of students, the examiner would behave in a manner that reflects the three prompting conditions. The examiner would behave in three different ways, one with each student of the group. In Condition A, the examiner would not prompt the student at all. In Condition B, the examiner would provide minimal prompts to the second student and in Condition C, the examiner would provide maximum response to the third student.

7.1.3 Procedure

Unlike the actual exam, students in the simulated exam were not asked to read the passage aloud and describe a picture as these tasks were not the issue of this study. Instead, the students were given a passage to read silently for about five minutes. This silent reading session was meant to get the students to tune in and prepare themselves to talk on the theme related to the passage. After reading it silently, each student would then have a conversation session with the examiner. The examiner used the questions given in the exam paper (refer to the actual exam papers in Appendices B and C) to start engaging the student in a conversation.

The first student of each group was not given any further verbal prompts except the initial question (Condition A). However, the examiner could respond in any non-verbal manner or provide acknowledgements like `mm'. By responding minimally through non-verbal gestures, it is still possible to consider the session as an interaction. Just listening impassively and not

giving any response at all to the student would be too extreme for the purpose of this present study, and might place students under excessive pressure to perform. The second student in each group was given the minimum verbal prompts that were provided in the question paper as the student talked about the topic (Condition B). Finally, the third student in each group received the maximum response from the examiner (Condition C). This was done not only through the prompts which were provided in the question paper, but also through the examiner's personal response to points raised by the student. At the end of the session, the examiner rated the student's performance and awarded a mark that reflects their competency. The marking scheme for the simulated exams is similar to the one used in Experiment One.

The rationale for subjecting each student in each group to different prompting conditions was to observe the effect these conditions exert on the students' oral performance in the conversation. The different groups of students would also give an insight into the difference in oral performance among students of different learning ability although we should keep in mind that individual case studies are not necessarily representative of larger populations.

8   Discussion

The findings observed in Experiment Two seem to provide mixed views about the effects of promptings in an oral examination setting. While some findings reinforce the importance of examiner's promptings and involvement in the interactions, others suggest the need for careful consideration and further investigations.

Data gathered in Experiment Two show that the absence of verbal promptings does not have a significant impact on high ability students. However, the effect it has on the mid and low ability students requires some attention. Results show that mid and low ability students in Condition A performed more poorly than students of the same ability level in Experiment One. This shows students could not perform well when examiners do not actively respond to the interactions. Interestingly, the better grades achieved by the mid and low ability students in Experiment One might indicate the presence of some involvements by the examiners in the actual exams

The amount of words spoken by students seems to be less clearly related to the prompting conditions. The amount of response varies with students under the three different conditions. The student ability seems to be a useful factor in explaining the differences in performance. While the high ability students were able to capitalise on the prompts given, the mid and low ability students were less responsive to the examiners' effort to encourage them to talk.

The results of interactions in Condition C reveal some interesting outcomes. With maximum prompting, the number of turns taken increases. This suggests that student-examiner interactions improve with promptings provided by the examiners. Students are also capable of using fillers to hold on to their turns when examiners are actively engaged with them. The ability to hold on to their turn enabled students to talk further and contribute more to the interactions.

Overlaps are prevalent in Condition C. With maximum promptings, students and teachers are highly involved and this has inevitably caused the interactions to `cross path'. Since all the overlaps noted were in interactions under C, it appears that this particular discourse feature is a characteristic of a highly interactive verbal encounter.

Pragmatically, the concept of `only one person speaking' is only an idealisation. Although in some cultures (like Malay), one is not allowed to speak unless the other party (the current speaker) has finished talking so as not to be regarded as being rude, there could be overlaps and interruptions (for example, when one perceives wrongly a closure or when one is in an argument).

The data gathered in Experiment Two seem to counter the ideal notion of `conversation'. In the simulated exam, a total of 162 exchanges took place. Of these, there were 14 turns in which the responses were not of the question-and-answer structure. This accounts for only 9% of the overall number of turns in the exam. The majority of the exchanges showed the examiner as the one asking questions and the candidates answering. This structure of question-and-answer in interaction is an exchange system peculiar to interviews (Button, 1987).

Greatbatch's (1988) analysis of news interview discourse discusses turn-taking features that are also observable in this study. For instance, it is observed that both the examiners and students (in the actual and simulated exam) seem to systematically attempt to produce turns which are recognizable as questions and answers respectively. The attempt is obvious for both parties even when performed minimally. The following excerpt illustrates an example.

Excerpt

114	T2:	Have you ever thought of becoming a teacher?
115		
116	S7:	Got ,,
117		
118	T2:	Why?

In spite of the interesting observations made in Experiment Two, one has to be cautious when making any conclusion. The small sample used in the experiments and the unavoidable difficulties in obtaining data for this study are issues that require careful consideration in future study. Care should be taken in generalising the findings of this small-scale study to larger populations.

9   Implications

The results derived in this study are based on a very small sample. There is therefore a need to be careful so as not to make sweeping generalisations on the basis of findings made. The small sample may not be a reasonable representation of the larger population of the oral exam community. However, they provide us with a `close-up' view of what goes on in an oral exam, in particular, face-to-face interaction. The results also present important implications on the education system.

Educators need to reflect on their perceptions and practices in the conduct of the oral exam. The assumption that more able students (mid and high ability) are naturally able to perform and score better than their less able peers may not be true all the time. This study shows that there are other possible factors that need to be explored and taken into account in order to

achieve a reliable assessment of the students' oral proficiency. For example, the `interview' nature of the oral exam interactions may place students at a disadvantage if they are not equipped with the background information to talk on diverse topics when they answer questions from the examiners. Continuous training and sharing of experience will help educators reduce the gap in expectations and differences in assessment decisions. Unlike written exams where markers are given the opportunity to reflect their assessment in the form of a report, the oral exam lacks this procedure. Similar practice in the oral assessment may help highlight problems and raise awareness of the students' performance. These considerations would allow examiners to be more sympathetic and reasonable in their expectations of students' communicative competence.

The results of this study also show that students may lack the necessary skills to enable them to perform reasonably well in the exam. Continuous exposure and practice may help to relieve some of the anxiety and/or lack of confidence faced by students. In order to make the interaction more like a conversation, students need to be taught on how to introduce new ideas or topics into the interaction instead of simply answering the examiner's questions. This would allow students to break away from the question-answer interactions imposed by the examiners and talk on something else familiar but still related to the theme of the interactions.

At the institutional level, the awareness derived from this study could be useful to professionals who are concerned with testing and evaluation of student performance. The examination branch of the Ministry of Education (in this case, the Singapore Examination and Assessment Board) may be interested to note how a teacher's involvements in the examination affect the evaluation of a student's oral proficiency. The fact that examiners have different perceptions and levels of involvement in the exam requires serious consideration. Continuous effort should be made to bridge the gap between the expected and actual conduct of the exam by the ministry and the examiners respectively. The study also shows how examiners' inconsistent behaviour (in terms of promptings) may have an impact on students' performance. Trainings and workshops may help to foster a better understanding and commitment among educators in ensuring assessment

validity. A change in the exam procedure could also be considered to help ease students' anxiety. For example, a second examiner could be included as a scorer, sitting away from the candidate's view, which could help to reduce distractions or anxieties in the student. In view of the varied nature of students' ability, examiners' involvement in the exams could lead students to be more confident of their own ability and hence perform better.

10   Conclusion

This study is a small step towards understanding the interaction that occurs in an oral exam, and the roles the examiner plays in the student's achievement. Although the study is mainly descriptive in nature, it has shed some light on some issues that arise in face-to-face testing. The techniques used in the data analysis provide us with unique opportunities to know the `inside story' of the assessment process.

Although there are suggestions made in this study, even if only tentative, many more questions remain. For example, it has been noted that student measured ability (based on their PSLE T-Score) did not always correlate with their oral proficiency scores reported here. This finding leaves us with more questions to explore.

Another critical question that remains to be answered is the question of how we can actually measure and determine the examiner's supportive behaviour in ensuring student's success in the assessment. How much support is necessary to allow students taking the exam, to produce a reasonable amount of talk for a reliable assessment to be made? The understanding of the nature of the oral interaction needs to be seriously considered too. The tendency for the oral exam to be more like an interview warrants a review of the evaluation process. Will we continue to assume that we are measuring the student's ability to interact in normal conversations when we are in fact evaluating their ability to answer questions in an interview context? I suggest that examiners be trained to change their expectations in order to accommodate the `interview' nature of the oral exam.

The study has opened up many possibilities for further research in the area of oral proficiency tests. Among other things, we have learned that giving support to students in oral assessment does not necessarily mean that we

have to bombard them with more questions to answer. In order to make the session more like a valid conversation assessment, examiners would need to provide the necessary interactional verbal cues like backchannel signals, acknowledgments through repetitions, and deal with the student's pauses in reducing their anxiety. Through these means, the examiner can then help set the necessary condition for meaningful conversation to take place.

References

1. Allison, D. 2002. *Approaching English Language Research.* Singapore: Singapore University Press.

2. Allison, D., Wee, L., Bao, Z., and Abraham, S.A (eds.). 1998. *Text in Education and Society.* Singapore: Singapore University Press.

3. Allison, D., Tauroza, S. 1995. "The Effect of Discourse Organisation on Lecture Comprehension." *English for Specific Purposes* 14. 157-173.

4. Bachman, L.F. 1990. *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press.

5. Braidi, S.M. 2002 "Reexamining the Role of Recasts in Native-Speaker/Nonnative-Speaker Interactions." *Language Learning* 52. 1-42.

6. Brown, A. 2003. "Interviewer Variation and the Co-Construction of Speaking Proficiency" *Language Testing* 20. 1-25.

7. Brown, G., and Yule, G. 1983. *Discourse Analysis.* New York: Cambridge University Press.

8. Button, Graham.1987. "Answers as Interactional Products: Two Sequential Practices in Interviews." *Social Psychology Quarterly* 50. 160-171.

9. Emanuel A. S., Jefferson, G., Sacks, H. 1997. "The Preference for Self-Correction in the Organization of Repair in Conversation." *Language* 53. 361-382.

10. Foley, J. 1988. *New Englishes: The Case of Singapore.* Singapore: Singapore University Press, National University Press.

11. Fulcher, G., Reiter, R.M. 2003. "Task Difficulty in Speaking Tests." *Language Testing* 20. 321-344.

12. Greatbatch, D. 1988. "A Turn-Taking System for British News Interviews." *Language in Society* 17. 401-430.

13. Have, P.T. 1999. *Doing Conversation Analysis: A Practical Guide.* London: Sage Publications.

14. Heritage, J. and Greatbatch, D. 1991. "On the Institutional Character of Institutional Talk: The Case of News Interviews", in D. Boden and D.H. Zimmerman (eds). *Talk and Social Structure.* pp. 93-137. Berkeley: University of California Press.

15. He, A. W. 1998. "Answering Questions in LPIs: A Case Study." In *Talking and Testing: Discourse Approaches to the Assessment of Oral Proficiency,* ed. by Richard Young and Agnes W. He, 101-115. Amsterdam and Philadelphia: Benjamins.

16. Johnson, M., and Tyler, A. 1998. "Re-Analyzing the OPI: How Much Does It Look like Natural Conversation?" In *Talking and Testing: Discourse Approaches to the Assessment of Oral Proficiency,* ed. by Richard Young and Agnes W. He, 27-49. Amsterdam and Philadelphia: Benjamins.

17. Jones, R.L. 1981, April. "Scoring Procedures in Oral Language Proficiency Tests." In *Selected Papers from RELC Seminar on Evaluation and Measurement of Language Competence and Performance, S' pore, April 1980,* ed. by John, A. S., and Read, Singapore: Singapore University Press.

18. Lazaraton, A. 2002. "A Qualitative Approach to the Validation of Oral Language Tests." *Studies in Language Testing* 14.

19. Lazaraton, A 1999. "Interlocutor Support in Oral Proficiency Interviews: the case of CASE." *Language Testing* 13. 151-172.

20. Lazaraton, A. 1992. "The Structural Organization of a Language Interview: A Conversation Analytic Perspective." *System* 20. 373-386.

21. Linell, P.1998. *Talk, Interaction and Context in Dialogical Perspectives.* Amsterdam/Philadelphia. John Benjamins Publishing Company.

22. McLaughlin, M., L. 1984. *Conversation: how talk is organized.* Beverly Hills: Sage Publications.

23. Nunan, David.1992. *Research Methods in Language Learning.* New York: Cambridge University Press.

24. Oller, J.W., 1983. *Issues in Language Testing Research.* Rowley, Mass: Newsbury House.

25. Riggenbach, H. (ed). 2000. *Perspectives on Fluency.* Michigan. University of Michigan Press.

26. Riggenbach, H. 1998."Evaluating Learner Interactional Skills: Conversation at the Micro Level." In *Talking and Testing: Discourse Approaches to the Assessment of Oral Proficiency,* ed. by Richard Young and Agnes W. He, 53-67. Amsterdam and Philadelphia: Benjamins.

27. Ross, S. 1992. "Accommodation Questions in Oral Proficiency Interviews." *Language Testing* 9. 173-186.

28. Shohamy, E.1983."The Stability of Oral Proficiency Assessment on the Oral Interview Testing Procedures" *Language Learning* 33. 527-540.

29. Sacks. H.1992. *Lectures on Conversation*, vols. 1 and 2. Cambridge, MA: Blackwell.

30. Schegloff, E., A., Jefferson, G., and Sacks, H. 1977."The Preference for Self-correction in the Organization of Repair In Conversation." *Language* 53.361-82.
31. Schegloff, E., and Sacks, H., 1973. "Opening up closings." *Semiotica* 8. 289-327.

32. Tan, J., Gopinathan, S., and Ho, W.K. (eds.) 1997. *Education in Singapore: A Book of Readings.* Singapore: Prentice Hall.

33. Tree, J., E. 2002. "Interpreting Pauses and *Ums* at Turn Exchanges." *Discourses Processes* 34. 37-55.

34. Underhill, Nic.1987. *Testing Spoken Language: A Handbook of Oral Testing* Techniques. Cambridge: Cambridge University Press.

35. Wennerstrom, A., & F. Siegel, A. 2003. "Keeping the Floor in Multiparty Conversations: Intonation, Syntax, and Pause." *Discourse Processes* 36. 77-107.

36. William, F., Lantolf, J. P. 1985. "Oral-Proficiency Testing: A Critical Analysis." *The Modern Language Journal* 69. 337-345.

37. Wilson, J., 1989. *On the Boundaries of Conversation.* Oxford. Pergamon Press.

38. Young, R. 1995. "Conversational Styles in Language Proficiency Interviews." *Language Learning* 45. 3-42.