THE AFFECTION OF INTERVIEWER SUPPORT ON ONE'S PERCEPTION AND PERFORMANCE IN AN ORAL PROFICIENCY INTERVIEW

Hendi Pratama
Semarang State University

ABSTRACT

This study examines the relationship of the support provided by the interviewers on the test-taker perception and performance. There are two levels of support investigated: High Supporting Behavior (HSB) and Low Supporting Behavior (LSB). Ten participants sat in two interviews with two different interviewers applying LSB and HSB condition each. All participants completed a perception questionnaire at the end of the interviews. Twenty interview scripts were computed to determine the participants' actual performance in three categories: (1) fluency, (2) lexical resources and (3) grammar complexity. The results of the study reveals that there are effects on performance but not perception.

Keywords: Interviewer Support, Performance, Perception

INTRODUCTION

In late 1980s, van Lier (1989) exposed a latent problem attached to the speaking test in the form of interview. He argued that if the goal of the interview is to measure the ability of the candidate in a conversation, the test itself is supposed to resemble a conversation. However, a deeper discourse analysis shows that a speaking test does not meet the criteria of daily conversations in at least four aspects: (1) the control possessed by the interviewer, (2) the power inequality, (3) the purpose of the talk and (4) topic nomination. These aspects pose a potential threat to the validity of Oral Proficiency Interview (OPI).

When it comes to scoring, reliability of the test can be threatened. Indeed, a speaking test is rated or marked by raters. Those raters might have different standards and different experiences which may cause them to give different ratings to the same test-taker. The objectivity of the test is at stake. That is not all. In conducting the interview, interviewers often adopt their personal styles into the questions asked to the participants. It could be teacher-like, lawyer-like, interrogator-like or comedian-like discourse practiced in the interview. Brown (2003) reported that two different interviewers could completely build different types of discourse when they are assigned to interview a single test-taker. How do these different styles affect the test-taker perception and performance? This question becomes the point of attention in this study.

Students around the world including students from Indonesia take such tests to seek entrance to universities. Job seekers around the world, at the same time, rely on

this type of test to gain employment. More scientific studies conducted regarding the reliability of an OPI will make each penny they spend to take such test more worth it. Indeed, an OPI is a high stake test and it is the responsibility of the test maker and/or provider to pay more attention to its validity and reliability.

REVIEW OF THE RELATED LITERATURE

The Nature of the Oral Proficiency Interview

The term Oral Proficiency Interview (OPI) was introduced by the *American Council on the Teaching of Foreign Languages (ACTFL)* (O'Loughlin, 1997). OPI refers to an oral assessment process where the candidates are tested via a series of communicative exchanges with the examiner. This type of oral assessment test attempts to conjure a direct speaking test in which the test-taker and the interviewer are positioned in a face-to-face communication (Lorenzo-Dus, 2007).

In the later development, the need of modern second language assessment is demanded to be in line with the communicative competence pedagogy. The expectation is that the outcome of a test can predict the likeliness of the test-takers to succeed in a real life communication (Cheng et al., 2004, p. 16). Hence, there is a high demand for oral proficiency assessments especially in high stake tests, i.e. tests needed to apply for jobs, universities and other important affairs (Lazaraton, 2002, pp. 5-7). As the consequence of this trend, an OPI (or sometimes a SOPI) is adopted as an inevitable part of prominent standardized tests administered by a number of modern governments and large scale testing industries. The Australian Department of Immigration and Multicultural Affairs (DIMA) for some periods had been administering ACCESS test, in which the test-taker should sit in an interview, as a predictor of the migrants English proficiency which is regarded important for migration processes. Another instance, a test with larger scope of purposes, IELTS, is designed by University of Cambridge Local Examinations Syndicate (UCLES) and jointly administered by the British Council and IDP Australia (Lazaraton, 1996a). Individuals who are seeking entrance to universities in Australia and the UK are required to do IELTS. Again, a test like the IELTS requires the candidate of the test to sit in a face-to-face interview with a trained examiner (Brown, 2006).

Important Issues on Interviewer Support Variability

The interviewer's unique behavior in an oral interview has been attracting the researchers' attention since the early 1980's. Shohamy (1983) investigates an interesting phenomenon that the same test taker can be awarded different scores by different interviewers. This research was considered as one of the pioneers of inter-rater reliability studies in oral assessment

(Brown, 2003, p.4). As in the case of most OPIs, the interviewers also serve as the raters at the same time, meaning that while (or after) interviewing, the interviewer gives rating to the test-taker. This area was revisited in 1990's by other language testing researchers (Brown, 1995; Berwick and Ross, 1992; Lazaraton, 1996a).The variability of the raters or interviewers may include their scoring techniques; their professional and educational backgrounds; their testing experience and training; their gender; and their nativeness or non-nativeness status. Beyond those factors, there is a type of variability related to the different interactional behaviors or the discourse created by the examiners during the oral performance test (Brown, 2003). Brown (2003) claims that the variation in different interactional strategies utilized by different interviewers has a potential effect on test taker performance. In other words, interviewer style may affect the test-taker language. The study has successfully demonstrated through careful discourse analysis that a single test-taker tends to perform differently and be rated differently by two interviewers who posses different styles in interviewing. However, this study is too risky to generalize because it is based on the performance of a single subject which may have a sensitive issue on the interview structure (i.e. the interviewer structure is not counter-balanced). The study self-fulfils her prediction in her study with a colleague in the previous year. The study suggested that 'difficult' interviewers tend to elicit poor performance and 'easy' interviewers tend to elicit better performance (Brown and Hill, 1998).

The Roles of Interlocutor Frame and Potential Deviation

To control the improvisation done by the interviewers during the process of interview, the test makers are forced to make an effort to 'limit' the variability on the interviewer style to make sure that all test-takers can get equal treatments (Lazaraton, 2002, 21). As a direct response to this demand, test makers set up what is typically called an *interlocutor frame*. In their study, O'Sullivan and Lu (2006) describe the interlocutor frame as a script of questions which has to be followed by the interviewers. In addition, there is guidance from the test makers to avoid certain features of behavior. For instance, ACTFL OPI prohibits the interviewers from doing the following: slowing down, echoing or correcting responses, furnishing vocabulary, rushing response time, asking display questions, etc. (Buck, 1989). ACTFL OPI is not the only test adopting the limitation of behavior, the CASE test also possesses almost similar interlocutor frame and guidance (Lazaraton, 1996b).

Lazaraton (1996a) observed that regardless of how strict the interlocutor frame is, the interviewers from time to time deviates from the prescribed scripts with predictable manners. She picks out and

identifies a list of behaviors from 200-page transcript generated from 58 audiotapes of CASE interviews. At least eight types of deviation can be identified.

Table 1. Summary of Lazaraton's (1996b) listing of interviewer's behavior

Features of Behavior	Definition
1. Priming topics	Cueing candidates on the next topic
2. Supplying vocabulary	Completing the test-taker utterances
3. Giving evaluative responses	Giving judgmental comment on performance
4. Echoing and correcting responses	Repeating and then correcting part of utterances
5. Slowing Down & Over-articulation	Exaggerating utterances
6. Rhetoric Question	Questions which only need mere confirmation
7. Drawing conclusions	Interpreting the test-taker utterance
8. Rephrasing questions	Simplify complex question

The list has strengthened the view that the interviewers by nature tend to *support* and *accommodate* the candidates although such actions have been discouraged by the test maker. A more laborious study was then conducted in 2002. O'Sullivan and Lu transcribe 70 audio-taped interviews to take a deep look at the nature of these deviations. Another list containing the types of support provided by the interviewers can be displayed here.

Table 2. Summary of O'Sullivan and Lu listing of interviewer deviation (2006, p. 8)

Types of Deviation	Definition
1. Interrupting Questions	Question asked that stops the test-taker's answer
2. Hesitated Questions	Question asked hesitatingly
3. Paraphrased Questions	Paraphrased questions without request from test-taker
4. Paraphrased and Explained Qs	Repeating and then correcting part of utterances
5. Comments after Replies	Comment made after test-taker's reply
6. Improvised question	Asking questions out of the script based on previous utterances
7. Informal Chatting	Informal discussion
8. Loud Laughing	Laughing because of test-taker's reply
9. Offer of Clues	Offer a hint to facilitate candidate reply

The result of their study concludes that deviations demonstrated by those interviewers are reflections of the interviewer support to accommodate the test-takers.

INVESTIGATING PERCEPTION AND PERFORMANCE

There are two dependent variables involved in this study: (1) participant perception and (2) participant performance.

Regarding the first, eliciting participants' perception in a research is a common practice to achieve a deeper understanding of the investigated phenomenon from a more humanistic approach. Further extension of this conduct is to inquire the relation between perception and the outcome of the study. Brown (2003), for example, involves an investigation of the potential relation between the interviewers' perception of the test-taker and the rating awarded to the test-taker. To achieve this aim, she conducted a semi-structured interview. An excerpt from the first interviewer says 'She's expressing what she wants to say quite reasonably'; while the second interviewer puts forward a comment like, 'She's not being helpful, you know... there's no sort of purpose to what she's talking about'. Indeed, there is no definite convention to interpret this type of comments. However, the researcher is then able to construct a logical inference that, the first interviewer tends to have a positive perception on the test-taker and the second interviewer perceives it to the opposite direction.

As for the second variable, researchers are required to use certain methods, software or technique in their research (Read and Nation, 2006). Each sub-skill under the speaking test needs different method to be assessed quantitatively. For example, to measure *lexical density*, type token ratio (TTR) technique can be used (Lorenzo Dus and Meara, 2005). To quantify the grammar complexity of the test-taker language, *number of words per utterance* and/or *number of clauses per utterance* can be analysed (Brown, 2002). Such quantitative analysis can be done after the interviews are recorded and transcribed using certain convention.

Interviewer Support, Test-taker Perception and Test-taker Performance

There are a number of studies which investigate the relationship between the interviewer support and the test-taker language performance. Lorenzo Dus and Meara (2005) demonstrated that an interviewer tend to give more features of support to the test-takers whom he/she assumes to have low proficiency and less features of support to the higher level test-takers. Hence, an abundant amount of support displayed in an interview is closely related to mediocre test-takers and little support from the interviewer is related to higher level test takers. Brown and Hill (1998) examined the possibility of different styles of interviewers in affecting the test-taker language. The result of the study strengthens the claim that there are particular advantages and disadvantages faced by the test-takers related to different interviewer styles. Brown (2003) confirms the finding by reassigning two of the interviewers used in her previous study, to interview, this time, a single test-taker. In terms of rating given by both interviewers, a significant difference can be noted; the 'easy' interviewer gives higher rating than the rating given by the 'difficult' interviewer.

Further discourse analysis shows that, the 'easy' interviewer gives more support by integrating the test-takers utterances into the next questions and reformulating misunderstood or unintelligible questions. The 'difficult' interviewer gives less support. It was revealed in the study that the 'difficult' interviewer thought that the test-taker so often failed to anticipate communication breakdowns. Both Lorenzo Dus and Meara's (2005) study and Brown's (2003) study lay their finding on the interviewer perception on how well the interview goes. This study will shift the point of view to the perception of the test-taker.

Other research on interviewer support and the test-taker language performance has been conducted by O'Sullivan and Lu (2006). Features of support provided by the interviewer to the test-takers in their study are treated as 'deviations of the *interlocutor frame* provided by the test-maker'. Their study investigates the nature of those *interviewer deviations* (in my study, they refer to *the features of interviewer support*) in particular test (in their case IELTS speaking test). By comparing the quality of the test-takers utterance 30 seconds before and after the deviation occurs, the study aims to find out if there were any effects of the deviation to the test-taker language by (O'Sullivan and Lu, 2006, p.10). The finding shows that no significant effects were found, other than a slight performance, related to the support provided by the interviewers. The study was comprehensive but there is an assumption in the method of data collection which may have possible threat to the finding. In a 10-to-15-minute interview, a quite number of deviations may occur closely to each other. The utterances made by the test-taker 30 seconds before and after a deviation is not necessarily related to the intended deviation but they may be related to the other deviations, which occur before and after the utterance. To avoid this problem, this research will conduct a series of manipulated interviewing processes which controls the amount and types of support. The first research question (RQ) of this study will be:

> *(RQ1) How does the interviewer support affect the test-taker perception?*
> The second question of this research is:
> *(RQ2)What are the effects of the interviewer support on the test-taker language performance?*

The Participants and the Interviewers

The participants of this study are ten English as a Second Language (ESL) learners who are currently studying at the University of Queensland Australia. All are postgraduate students. As part of their admission requirements to the university, they all have an IELTS score equal or greater than 6.5. Differences in cultural backgrounds are minimized because all of the participants are from Indonesia. The participants never lived in an English-

speaking country for more than 6 months prior to their stay in Australia. If this study is compared to those of Lorenzo-Duz and Meara (2005) and O'Sullivan and Lu (2006), the number of participants in this research is relatively small. However, if this study is regarded as the extension of Brown's (2003) study which only deals with a single test-taker, ten test-takers provide a more complete picture of the effects of the supporting behavior to the test-takers performance.

There are two interviewers participated in this study. The interviewers are English native speakers who were experienced ESL teachers. The interviewers are trained to act out two types of behavior in the interview: (1) *Low support behavior* (LSB) and (2) *High support behavior* (HSB). In an LSB interview, the interviewers were trained to give a very minimum amount of support provided to the test-takers. In contrast, in an HSB interview, the interviewers are encouraged to give maximum support, as prescribed by the researcher, to the test-takers.

The Interlocutor Frame

The interviewers' behavior is intentionally manipulated to achieve the desired result of this study. After studying, examining, combining and coining the terms and definitions of interviewers' behavior provided by Lazaraton (1996b), O'Sullivan and Lu (2002) and Ross and Berwick (1992), there are five features of support that will be investigated in this study.

Table 3. The list of supporting features used in the study

Features	Definition and Coverage	Code
1. Fronting	Interviewer gives additional information at the beginning of a question.	Fro
2. Suppliance	Interviewer gives options or new lexicon	Sup
3. Rephrase	Interviewer simplify the structure of the question	Rep
4. Back-channel	Interviewer makes utterance just for the sake of continuity It includes evaluative comments	Bac
5. Slowdown	Slowdown and/or over-articulate words	Slo

In an LSB interview, only *back-channel* and *slowdown* are allowed. As a result, the test-taker will only get as minimum support as possible from the interviewers. Back-channelling and slowing down are allowed to avoid total breakdown in the interview. In an HSB interview, all five features of support are allowed. *Back-channel* is allowed anytime during the interview while *slowdown* can only be practised when the participants ask for repetition or feel that the questions are not clear enough. However, the number of features in every exchange is carefully controlled. At this

stage, the interlocutor frame, i.e. a script, comes into play. The interviewers should obey strictly what has been prescribed by the interlocutor frame. Any modification is strictly prohibited. The script will distinguish the forms of questions used in the LSB and HSB interview.

Table 4. Sample of the interlocutor frame (complete scripts see Appendix A)

Features	LSB version	HSB version
Fronting	How important is the internet for the modern society?	*In the globalization era, the internet becomes an inevitable part of our society.* How important is the internet for the modern society?
Suppliance	Why do more and more people like online shopping these days?	Why do more and more people like online shopping these days? *Is it convenient, secure, practical or what?*
Rephrase	How would you explain the phenomenon that people do not have to attend classes to get formal education by means of the internet?	How would you explain the phenomenon that people do not have to attend classes to get formal education by means of the internet? *What do you think about distant learning using the internet?*
Back-channel	(very good, hmmmm, okay, I see)	(very good, hmmmm, okay, I see)
Slowdown	(allowed max twice for a question)	(allowed max twice for a question)

A single topic cannot be deployed in both interviews because the participants can be too familiar with the topic when they have to face the interview for the second time. Brown and Hill (1998) call this phenomenon as a practice effect. As the consequence, there are two topics displayed in the interview: the topic taken from the sample above is 'the internet' and the other one is 'the university life'.

There are two interviewers assigned in two separated rooms: *Interviewer A* and *Interviewer B*. The two rooms have been equipped with high-definition audio recorders and back-up recorders for the purpose of the study. The focus of this study is on the amount and types of accommodation given by the interviewers. Hence, both interviewers should be prepared to be able to act out two scenarios of support or accommodation provision, namely LSB and HSB. For example, if the interviewer A acts out an LSB in session 1 so he has to act out an HSB on the next session. The purpose of this design is to provide a counter-balance on the interview structure.

Post-test Questionnaire

The post-test questionnaire (Appendix B) is designed to elicit the participants' perception on both interviews that they have been through. The questionnaire was completed immediately after the participants have done with both interviews (LSB and HSB). The main point which is investigated is whether or not the participant noticed the difference between

Interviewer A's and Interviewer B's behaviors. And if they notice, which condition (either LSB or HSB) is perceived more positively by the participants. This part of the questionnaire uses 5-point Likert-scale measurement. There are fifteen statements to be responded by the participants and translation is provided to improve the understanding of the participants on the questions.

DATA ANALYSIS

As a reference, there are four common scoring criteria in an IELTS speaking test i.e. fluency and coherence; lexical resource; grammatical range and accuracy; and pronunciation. Brown (2006) has done a very good job in describing comprehensively the test-taker language in an IELTS speaking test. Her foci of investigation include: fluency and coherence; lexical resource; grammatical range and accuracy but not pronunciation. She argues that pronunciation is too subtle to describe and might be out of the scope of the research.

In this study three categories of performance are investigated: (1) fluency, (2) lexical resource and (3) grammatical complexity. There are three indicators of fluency and coherence: (1) total amount of speech, (2) speech rate, and (3) response length. There are four measurements of lexical resource to analyse in this study: (1) K2 words, (2) type token ratio, and (3) ratio of content non-content words. *Grammatical* complexity can be computed by counting: (1) words per *T-Unit* (words per sentence) and (2) *T-Unit* per response (sentences per response).

The repeated measurement of only 10 participants cannot guarantee the normal distribution of the sample. A *t-test*, like any other parametric tests, requires strict prerequisites on the sample and data which may not be fulfilled by the sample provided in this study. Non-parametric tests are used due to the small number of participants (Field, 2005). The *Wilcoxon signed ranks test* is used here as a substitute of the *t-test* for related samples and it is still safely applied if the assumptions on the t-test are violated.

Participant Perception in HSB and LSB Conditions.

Since each participant sat in two different kinds of interviews (*LSB* and *HSB*), the questionnaire completed after the interviews is summarized to describe their attitudes on each condition. To simplify the concept of the Likert scale applied in this study, it can be generalized safely that *mean* values closer to *5* are regarded as positive perception. The bold figures indicate higher values than those of their counterparts.

Table 5. Summary of participants' perception on different interview conditions

		Perception			
		HSB		*LSB*	
Code	Questions	Mean	SD	Mean	SD
Q1	I feel that the interviewer listens to me	4.20	1.03	**4.40**	0.97
Q2	I feel appreciated	4.10	1.10	**4.40**	0.84
Q3	I feel comfortable	**4.10**	1.10	4.00	0.82
Q4	I feel that the interviewer was friendly	**4.20**	0.92	4.00	1.15
Q5	I feel that the interviewer has paid attention to my responses	4.10	0.88	**4.20**	1.23
Q6	The interviewer helped me to elaborate my responses	**2.40**	1.26	2.20	0.92
Q7	The interviewer helped me to understand his questions	**3.20**	1.40	2.90	1.20
Q8	The *Interviewer A*sked me easy questions	**3.70**	0.82	3.60	0.97
Q9	The interviewer was talkative	**2.00**	0.82	1.70	0.67
Q10	The interviewer in general, is easy to understand	4.00	0.47	4.00	0.67
Q11	I think that I was fluent	3.60	1.07	**3.70**	0.95
Q12	I think my vocabulary was good	**3.50**	0.97	3.20	0.92
Q13	I think my grammar was good	3.20	0.92	3.20	0.92
Q14	I think I responded completely	3.50	1.08	**3.80**	1.14
Q15	I think I would get a good score	3.30	0.95	**3.50**	0.85
	Overall Mean	**3.54**		3.52	

On the different supporting conditions of the interview, in this case the different wording of the questions, the overall mean shows that *HSB* was slightly well-perceived by the participants. If we apply *Wilcoxon signed ranks test* on this table, it shows that the *z* of the values born by each condition is only *-.389.* The *SPSS* software shows that *HSB* is a little bit more well-perceived than *LSB* but *z=.389* is below *1.96* to achieve the confidence level at $p < .05$. The supporting behavior has a very little impact to the test-taker perception in this study.

4.2. Participants' Performance in HSB and LSB Conditions.

After all interview recordings were transcribed, the scripts were cleaned, pruned and trimmed. The interviewer utterances were taken out for the sake of calculation. The first category of the performance examined in this study is *fluency*. There are three measurement categories under this heading. Hence, the final scripts for fluency analysis then analysed based on: (1) speech rate, (2) total words per session and (3) words per utterance. Speech rate is calculable via the

total words in that particular session and divided by the duration of the interview. The result value of this calculation would be words per minute. The second indicator of fluency is the total words produced by the participant in each session of the interview. Logically, the more fluent a participant, the more words he/she will produce in an interview. The third indicator of the fluency is the words per utterance. What is meant by words per utterance is the average of total words produced by the participants in answering each question asked by the interviewer. The result of this calculation might be contingent to the second indicator of fluency (total words). The following table shows the mean analysis of the three indicators representing the fluency analysis in this study. The values in HSB condition are compared with the values acquired in the LSB condition. The bold figures indicate higher values than those of their counterparts.

Table 6. Participant Fluency

	HSB		**LSB**	
	Mean	SD	Mean	SD
Speech Rate	74.74682258	16.67662	**78.61897**	7.913696
Total Words per Session	**1046.3**	293.9418	983.2	328.1154
Words per Utterance	**74.73571429**	20.99584	70.22857	23.43682

From the summary of the mean analysis of fluency above, there is a better performance reported in the total words produced per interview and the total words per utterance in HSB condition than the LSB condition. However, the speech rate value in HSB is less than that of LSB. It refers to the trend that the participants tend to speak slower in LSB condition.

The second category of performance investigated in this study is *lexical resource*. There are three categories examined: (1) K2 words, (2) Type Token Ratio and (3) Content Non-Content Word Ratio. Assuming that K1 words are the 'unsophisticated' words and K2 Words are the 'more sophisticated' words, it can be said that the higher K2 words percentage shows that the vocabulary usage is deeper. Thus, the lexical choice is said to be more sophisticated. The other two measurements used represent the lexical density of the speech. Lexically dense speech can mean that the speaker repeating less words and tend to apply more unique words in his/her sentences; or the speaker produce more content words (noun, verbs, adjective, etc) than non-content words (grammatical markers, pronouns, etc.). The first is estimated by the use of *Type Token Ratio (TTR)* and the second is estimated by *Content Non-content Word Ratio*. Those three measurements of lexical density can be summarized through the following table.

Table 7. Participant Lexical Resource

	HSB		LSB	
	Mean	SD	Mean	SD
K2 percentage	**3.699**	0.868529	3.501	0.630334
Type Token Ratio	0.285	0.042492	**0.287**	0.026268
Content Non-Content Word Ratio	0.425	0.03504	**0.436**	0.030984

In the analysis of the K2 words production, the majority of the participants perform better in *HSB* condition (*z=-.255, p>.05*). In HSB condition, the TTR mean value is slightly lower than LSB. *Content Non-content Word Ratio* also measures the lexical density of the speech but it takes a slightly different approach than that of TTR. In Content Non-content word ratio analysis, participants actually perform better when they are interviewed in LSB condition. Overall, the analysis shows that participants display more sophisticated vocabulary in HSB condition but they show less dense vocabulary when they are interviewed in HSB condition. Both results show only marginal differences.

Moving on to the third category of performance examined in this study, grammatical accuracy is analysed. There are two indicators used to represent the category. The first one is words per *T-Unit* (words per sentence) and the second is *T-Unit* per response (sentences per response). A *T-Unit* in a spoken text is the substitute of a sentence in a written text. The assumption of the first indicator is that the more capable the participants the more words they use in a sentence. Meanwhile, the assumption of the second is that the more capable the participants the more sentences they use every time they answer a question.

Table 8. Participant Grammatical Complexity

	HSB		LSB	
	Mean	SD	Mean	SD
Words per *T-Unit*	**13.79**	1.137688	13.46	1.279931
T-Unit per Response	**5.88**	1.489071	5.43	1.866101

The grammatical complexity analysis shows consistent result on either words per *T-Unit* calculation and *T-Unit* per response calculation. The participants produce more words per *T-Unit* in HSB condition and they also produce more *T-Unit*s per Response in HSB condition.

CONCLUSIONS

Regarding the first research question, the study has revealed that there is no (or a very little) impact of different supporting behavior to the participant perception. However, when it comes to the performance analysis, mixed results can be reported. In fluency analysis, the high

supporting behavior consistently decreases the speech rate of the participants. Presumably, this related to the load factor involved: the more information provided by the interviewers increases the processing time for the participants to answer the questions. In the case of lexical resource, HSB increases lexical sophistication but not density. Meanwhile, in grammatical complexity analysis HSB is consistently related to more words per sentence and more sentences per utterance.

Based on the findings, this research supports the previous studies which asserted that different supporting behaviors can affect the participant performance. At the practical level, the findings from this study can be one of the references when the test-makers or the test-providers design their protocol for the interviewers. As there is a strong tendency shown in this study that the interviewer behavior may affect the test-takers performance, a set of guidelines should be set for the interviewers on how far they can modify the interview scripts or frames. Whether the modification of the interview scripts are incidental or affected by individual styles, the degree of freedom must be pre-determined by the protocol. Further research can assign larger number of test-takers and interviewers to improve the scope of investigation and the level of confidence. Discourse analysis on the interaction between the interviewers and the test-takers might also be included in the analysis to give a better understanding of how different types of interaction can result in different performance.

REFERENCES

Bachman, L. F., Lynch, B. K., & Mason, M. 1995. Investigating variability in tasks and rater judgements in a performance test of foreign language speaking. *Language Testing, 12*(2), 238-257.

Brown, A. 1995. The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing, 12*(1), 1-15.

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

Brown, A. 2006. Candidate discourse in the revised IELTS Speaking Test. *IELTS Research Report, 6*.

Brown, A., & Hill, K. 1998. Interviewer style and candidate performance in the IELTS oral interview. *IELTS research reports, 1*, 173-191.

Buck, K., Byrnes, H., & Thompson, I. 1989. *The ACTFL Oral Proficiency Interview Tester Training Manual*. ACTFL

Canale, M., & Swain, M. 1980. *Approaches to communicative competence*. Seameo Regional Language Centre.

Cheng, L., Watanabe, Y., & Curtis, A. 2004. *Washback in language testing: Research contexts and methods*. New York: Lawrence Erlbaum Associates.

Douglas, D. (1994). Quantity and quality in speaking test performance. *Language Testing, 11*(2), 125-144.

Johnson, M. 2001. *The Art of Non-conversation*. New Haven: Yale University Press.

Koike, D. 1998. What Happens When There's No One to Talk to? Spanish Foreign Language Discourse in Simulated Oral Proficiency Interviews. *Talking and testing: discourse approaches to the assessment of oral proficiency*, 69-98.

Krashen, S. 1985. *The input hypothesis: Issues and implications*. New York: Longman Group United Kingdom.

Lazaraton, A. 1996a. *A qualitative approach to monitoring examiner conduct in the Cambridge assessment of spoken English (CASE)*.

Lazaraton, A. 1996b. Interlocutor support in oral proficiency interviews: the case of CASE. *Language Testing, 13*(2), 151.

Lazaraton, A. 2002. *A qualitative approach to the validation of oral language tests*. Cambridge: Cambridge Univ Press.

Lorenzo-Dus, N., & Meara, P. 2005. Examiner support strategies and test-taker vocabulary. *International Review of Applied Linguistics in Language Teaching, 43*(3), 239-258.

Lorenzo-Duz, N.2007. The best of both worlds? Combined methodological approaches to the assessment of vocabulary in oral proficiency interviews. In Daller, H., Milton, J., & Treffers-Daller, J (eds.). *Modelling and assessing vocabulary knowledge.* Cambridge: Cambridge University Press.

Lumley, T., & McNamara, T. F. 1995. Rater characteristics and rater bias: implications for training. *Language Testing, 12*(1), 54-71.

Luoma, S. 2004. *Assessing speaking*. Cambridge: Cambridge University Press.

Mackey, A., & Gass, S. 2005. *Second language research: Methodology and design*. New York: Lawrence Erlbaum Assoc Inc.

Messick, S. 1996. Validity and washback in language testing. *Language Testing, 13*(3), 241.

O'loughlin, K. 1997. *The comparability of direct and semi-direct speaking tests: a case study.* The University of Melbourne.

O'Loughlin, K. 2002. The impact of gender in oral proficiency testing. *Language Testing, 19*(2), 169-192.

O'Sullivan, B., & Lu, Y. 2006. The impact on candidate language of examiner deviation from a set interlocutor frame in the IELTS Speaking Test. *IELTS Research Report, 6*.

Read, J., & Hayes, B. (2003). The impact of IELTS on preparation for academic study in New Zealand. *IELTS research reports, 4*, 153–205.

Read, J., & Nation, P. 2006. An investigation of the lexical dimension of the IELTS speaking test. *IELTS Research Report, 6*.

Ross, S. 1992. Accommodative questions in oral proficiency interviews. *Language Testing, 9*(2), 173.

Ross, S., & Berwick, R. 1992. The Discourse of Accommodation in Oral

Proficiency Interviews. *Studies in Second Language Acquisition, 14*(2), 159-176.

Shohamy, E. 1983. The stability of oral proficiency assessment on the oral interview testing procedures. *Language Learning, 33*(4), 527-540.

Van Lier, L. 1989. Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOl Quarterly*, 489-508.

Weir, C. 1992. *Understanding and developing language tests*. New York: Prentice-Hall.

Weir, C. 2005. *Language testing and validation: An evidence-based approach*. New York: Palgrave Macmillan.

Young, R. 1995. Conversational styles in language proficiency interviews. *Language Learning, 45*(1), 3-42.

APPENDIX A
THE INTERNET (HSB)

1. In the globalisation era, the internet has become an unavoidable part of our society. How important is the internet to modern society?
2. It seems that the internet is very important to everybody. How about yourself? How do you use the internet as a part of modern society?
3. While you are studying in Australia, how does the internet help you in particular with your studies?
4. Nowadays, children and teenagers have relatively easy access to the internet. In your opinion, what are some of the negative effects of the internet on children and teenagers?
5. Dealing with the disadvantages of the internet on children and teenagers, how would you suggest these negative effects be overcome?
6. Can you tell me a bit about cyber crime? For example, hackers, carders, impersonators and so on.
7. Why do more and more people like online shopping these days? Is it because it is convenient, secure, practical or something else?
8. How do you or people in general expand personal networks and friendships over the internet? By email, facebook, chatting...
9. What is your opinion about downloading illegal materials from the internet? Music, video, software, pictures
10. What is an appropriate penalty for people who distribute illegal material? Imprisonment, fines, the death penalty?!
11. Why do you think mobile phone companies are investing in providing internet access in mobile phones? What do you think about the internet facility provided in mobile phones?
12. What do you think about people who fail to socialize in real life because they spend too much of their time in front of their computers? How do computers prevent some people from making friends?

13. How would you explain the phenomenon that people do not need to attend classes but are able to receive a formal education over the internet? What do you think about distance education/learning using the internet?
14. How do you see the internet in the future? Perhaps 10 years from now? Will the internet play a different role in people's lives than it does now?

THE INTERNET (LSB)

1. How important is the internet to modern society?
2. How do you use the internet as a part of modern society?
3. How does the internet help you in particular with your studies?
4. What are some of the negative effects of the internet on children and teenagers?
5. How would you suggest these negative effects be overcome?
6. Can you tell me a bit about cyber crime?
7. Why do more and more people like online shopping these days?
8. How do you or people in general expand personal networks and friendships over the internet?
9. What is your opinion about downloading illegal materials from the internet?
10. What is an appropriate penalty for people who distribute illegal material?
11. Why do you think mobile phone companies are investing in providing internet access in mobile phones?
12. What do you think about people who fail to socialize in real life because they spend too much of their time in front of their computers?
13. How would you explain the phenomenon that people do not need to attend classes but are able to receive a formal education over the internet?
14. How do you see the internet in the future?

APPENDIX B.

Post-Questionnnaire

Tick the appropriate answer and fill in the blank when options are not provided.

A. Participant: □1 □2 □3 □4 □5 □6 □7 □8 □9 □10
B. Have you taken IELTS test before? □Yes □No
C. How many times? □1 □2 □3 □ >3
D. Best IELTS Score?: □6.0 □6.5 □7.0 □7.5 □8.0 □8.5 □9.0
E. Age: □20-24 □25-29 □30-34 □35-39 □40-45 □>45
F. Nationality: _____
G. How long have you been staying in Australia?
   □1-6 months □6 months – 1 year □1-2 years □more than 2 years
H. Program at UQ:_____

I. Recollect your experience on the first interview and the second interview. Respond to the following statement and circle the number which best reflecting your perception.

1 = Strongly Disagree(SD); 2 = Disagree (D); 3 = Undecided(U); 4 = Agree(A); 5 = Strongly Agree(SA)

During the first interview, I feel	**SD**	**D**	**U**	**A**	**SA**
1. that the interviewer listens to me	1	2	3	4	5
2. appreciated	1	2	3	4	5
3. comfortable	1	2	3	4	5
4. that the interviewer was friendly	1	2	3	4	5
5. that the interviewer has paid attention to my responses	1	2	3	4	5
During the first interview, the interviewer:	**SD**	**D**	**U**	**A**	**SA**
6. helped me to elaborate my responses	1	2	3	4	5
7. helped me to understand his questions	1	2	3	4	5
8. asked me easy questions	1	2	3	4	5
9. was talkative	1	2	3	4	5
10. in general, is easy to understand	1	2	3	4	5
Based on my performance in the first interview, I think	**SD**	**D**	**U**	**A**	**SA**
1. that I was fluent	1	2	3	4	5
2. my vocabulary was good	1	2	3	4	5
3. my grammar was good	1	2	3	4	5
4. I responded completely	1	2	3	4	5
5. I would get a good score	1	2	3	4	5
During the second interview, I feel	**SD**	**D**	**U**	**A**	**SA**
6. that the interviewer listens to me	1	2	3	4	5
7. appreciated	1	2	3	4	5
8. comfortable	1	2	3	4	5
9. that the interviewer was friendly	1	2	3	4	5
10. that the interviewer has paid attention to my s response	1	2	3	4	5
During the second interview, the interviewer:	**SD**	**D**	**U**	**A**	**SA**
11. helped me to elaborate my responses	1	2	3	4	5
12. helped me to understand his questions	1	2	3	4	5
13. asked me easy question	1	2	3	4	5
14. was talkative	1	2	3	4	5
15. in general, is easy to understand	1	2	3	4	5
Based on my performance in the second interview, I think	**SD**	**D**	**U**	**A**	**SA**
16. that I was fluent	1	2	3	4	5
17. my vocabulary was good	1	2	3	4	5
18. my grammar was good	1	2	3	4	5
19. I responded completely	1	2	3	4	5
20. I would get a good score	1	2	3	4	5

Additional comment please write here.

_____
_____
_____
_____