



UNIVERSITI PUTRA MALAYSIA

**EFFECT OF PROMPT LENGTH ON STUDENT'S SPEAKING
PERFORMANCE AND USE OF TARGET VOCABULARY**

MASOUMEH AKHONDI

FPP 2007 3



**EFFECT OF PROMPT LENGTH ON STUDENT'S SPEAKING
PERFORMANCE AND USE OF TARGET VOCABULARY**

By

MASOUMEH AKHONDI

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,
in Fulfillment of the Requirements for the Degree of Master of Science**

June 2007



To my late father, my loving mother, and my nice husband Faramarz



Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Master of Science

THE EFFECT OF PROMPT LENGTH ON STUDENT'S SPEAKING PERFORMANCE AND USE OF TARGET VOCABULARY

By

MASOUMEH AKHONDI

June 2007

Chairman: Arshad Abd. Samad, PhD

Faculty: Educational Studies

The purpose of this study is to investigate the impact of the length of prompt (written stimulus) on the interviewee's performance in language proficiency interview. The main quantitative objective is to determine the extent to which the interviewee who receives long/short prompt uses the target vocabulary from the prompt in his speech. Qualitatively, this research was attempted to determine the effectiveness of the long/short prompts on the quality of the student's overall performance. Here, quality refers to the student's ability to produce extended, coherent, as well as cohesive chunks of speech. It was hypothesized that, "there is no significant difference in using target vocabulary among the groups receiving long/short and no prompt." In an experimental randomized



subjects posttest-only control group design, 45 participants, comprising the three groups, attended the interviews. The participants received the long (600 words) and the short (300 words) prompts pertaining to the topic prior to the interview, while the control group experienced a prompt-less interview. The following research questions were answered: 1. Will the long prompt group interviewees, receiving heavier target vocabulary load prompts, compared to the short prompt group participants have less difficulty in finding the most appropriate vocabulary to come up with the posed questions? 2. Will the long prompt group perform qualitatively better? After an Analysis of Variance (ANOVA), quantitative findings of the post hoc comparison among the three groups revealed that the difference between the two experimental groups is not statistically significant ($p > .05$), yet there were differences observed between each of the experimental group and the control group ($p < .05$). For the qualitative analysis of the data, verbal protocol analysis applied. The findings revealed that, giving interviewee prompt with higher lexical density and higher vocabulary load has noticeably affected his performance. The content of the prompts has provided the fundamental floor for the interviewees to ground his speech on which led to producing more coherent and more cohesive chunks of speech.



Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains

**KESAN PANJANG RANSANGAN TERHADAP PRESTASI PERBUALAN
DAN PENGGUNAAN PERBENDAHARAAN KATA PELAJAR**

Oleh

MASOUMEH AKHONDI

Jun 2007

Pengerusi: Arshad Abd. Samad, PhD

Fakulti: Pengajian Pendidikan

Tujuan kajian ini adalah untuk meneliti impak panjang ransangan (ransangan bertulis) ke atas peserta semasa temuduga kemahiran bahasa. Objektif kuantitatif utama adalah untuk menentukan tahap penggunaan kosa kata sasaran dari ransangan oleh “prompt”. Secara kualitatif, penyelidikan ini berusaha menentukan keberkesanan ransangan pendek/panjang terhadap kualiti kemahiran keseluruhan pelajar. Kualiti merujuk pada kebolehan pelajar menghasilkan “chunks of speech” yang “extended” dan “coherent”. Di “hypothesized” bahawa tiada perbezaan signifikan antara penggunaan kosa kata sasaran di kalangan kumpulan yang menerima ransangan panjang/pendek dan tiada ransangan.

Dalam rekabentuk subjek rawak, pasca ujian sahaja kumpulan kawalan, 45 peserta



yang terdiri dari tiga kumpulan menghadiri temuduga. Peserta menerima ransangan panjang (600 perkataan) dan ransangan pendek (300 perkataan) berkaitan tajuk tertentu sebelum temuduga manakala kumpulan kawalan ditemuduga tanpa ransangan.

Persoalan kajian berikut dijawab: 1. Adakah kumpulan yang menerima ransangan panjang menghadapi kurang kesukaran untuk menggunakan kosa kata yang sesuai berbanding kumpulan yang menerima ransangan pendek? 2. Adakah kumpulan ransangan panjang menunjukkan prestasi yang lebih baik secara kualitatif? Setelah melakukan Analysis of Variance (ANOVA), dapatan kuantitatif berdasarkan perbandingan post hoc menunjukkan perbezaan antara dua kumpulan eksperimental adalah tidak signifikan ($p > .05$), tetapi terdapat perbezaan antara setiap kumpulan eksperimen dan kumpulan kawalan ($p < .05$). Untuk analisis kualitatif data, analisis protokol lisan digunakan. Dapatan menunjukkan bahawa dengan memberi “interviewee” ransangan yang mempunyai kepadatan lexicon dan beban kosa kata yang tinggi telah memberi kesan terhadap penggunaan bahasa. Kandungan ransangan telah memberikan “floor” asas untuk pelajar yang ditemuduga untuk mengasaskan ucapannya yang sesuai. Kumpulan ransangan pendek telah mendapat min tertinggi dalam penggunaan-penggunaan kosa kata ransangan “cohesive chunks”.

ACKNOWLEDGEMENTS

I would like to express my sincere gratefulness to my Supervisor, Dr. Arshad Abd. Samad, for infinite inspiration, insightful input, intellectual support and comments, cheering, and enthusiasm, which made this research possible. Without his faith, encouragement, help and patience, this research would still be in the evolutionary stage.

I also wish to thank Professor Dr. Ab. Rahim Bakar who offered valuable intellectual input. Without his help, I would not have come so far.

I am deeply grateful to my mother and to the rest of my family who kept encouraging me during my study. My thanks to them who believed in me, expressed their patience and provided me with moral support for the past year.

Finally, I am endlessly thankful to my nice husband, Faramarz who is also my best friend. He cheered and supported me daily, with his patience and extra generosity with his time to edit my stylistic errors and correct my mistakes. He continuously encouraged me through this one year of my study and research. My degree and this thesis would not have been accomplished without him.

The last but not the least, my special thanks to my son Farbod for the time we should have spent together. Thank you son, I love you.



I certify that an Examination Committee has met on 1st June 2007 to conduct the final examination of Masoumeh Akhondi on her Master of Science thesis entitled “Effect of Prompt Length on Student’s Speaking Performance and Use of Target Vocabulary” in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the student be awarded the Master of Science.

Members of the Examination Committee are as follows:

Maznah Baba, PhD

Lecturer
Faculty of Educational Studies
Universiti Putra Malaysia
(Chairman)

Ghazali Mustapha, PhD

Lecturer
Faculty of Educational Studies
Universiti Putra Malaysia
(Internal Examiner)

Nooreen Noordin, PhD

Lecturer
Faculty of Educational Studies
Universiti Putra Malaysia
(Internal Examiner)

Parilah Hj. Mohd Shah, PhD

Associate Professor
Faculty of Educational Studies
Universiti Kebangsaan Malaysia
(External Examiner)

HASANAH MOHD. GHAZALI, PhD

Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 3 AUGUST 2007



This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Arshad Abd. Samad, PhD

Lecturer
Faculty of Educational Studies
Universiti Putra Malaysia
(Chairman)

Fauziah Bt. Hassan, PhD

Lecturer
Faculty of Educational Studies
Universiti Putra Malaysia
(Member)

Ab. Rahim Bakar, PhD

Professor
Faculty of Educational Studies
Universiti Putra Malaysia
(Member)

AINI IDERIS, PhD

Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 9 AUGUST 2007



DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

MASOUMEH AKHONDI

Date: 22 JUNE 2007



TABLE OF CONTENTS

	Page
ABSTRACT	iii
ABSTRAK	v
ACKNOWLEDGEMENTS	vii
APPROVAL	viii
DECLARATION	x
LIST OF TABLES	xiv
LIST OF FIGURES	xv
LIST OF ABBREVIATIONS	xvi
CHAPTER	
1	INTRODUCTION 1
2	LITERATURE REVIEW 17
	Introduction 17
	Communicative language teaching and learning 20
	Communicative competence 20
	Communicative language teaching 22
	Communicative competence and test method 23
	Proficiency vs. communicative competence 27
	Psycholinguistic model of speech production 31
	Authenticity in language testing 36
	Discourse analysis 40
	Coherence and cohesion in spoken language 40
	Context-embedded and context-reduced 43
	Assessing speaking proficiency 45
	Testing methods and context 45
	Principles underlying oral testing methodology 48
	Elicit extended chunks of speech and control the input 49
	Elicit structured or organized speech 49
	Elicitation techniques used in the OPI 51
	Ensure a valid and reliable assessment 54



	Research on oral examinations	56
	Empirical studies on the OPI	57
	Theoretical framework	58
3	METHODOLOGY	62
	Introduction	62
	Research design	62
	Validity of the randomized subjects, posttest-only control group design	65
	Internal and external validity of the design	65
	Sampling	67
	Location of the study	68
	Procedures	68
	Instruments	71
	Scoring	71
	Variables	72
	Research framework	73
	Data collection	75
	Data analysis	76
4	RESULTS OF THE STUDY	77
	Introduction	77
	Demographic reports for participants	77
	Research objectives	80
	Qualitative description for the first objective	82
	Control group	82
	Experimental group 1	85
	Experimental group 2	86
	Descriptive and inferential statistics for the second objective	88
	Descriptive statistics and frequency distribution for the second objective	88
	Inferential statistics for the second objective	91
	Final remarks	95



5	DISCUSSION OF FINDINGS AND CONCLUSIONS	97
	Introduction	97
	Brief overview of the findings	97
	The target vocabulary used by the three groups of the interviewees	98
	Relating the study to communicative competence and language proficiency	106
	Relating the study to the authenticity in language testing	107
	Suggestions for further research	109
	REFERENCES	112
	APPENDICES	120
	BIODATA OF STUDENT	133



LIST OF TABLES

Table	Page
1 Frequency distribution of race for the three groups	78
2 Frequency distribution of gender for the three groups	79
3 Frequency distribution of the public speaking scores for the three groups	79
4 Frequency distribution of CGPA for the three groups	80
5 Frequency analysis of the target vocabulary used by the three groups	89
6 ANOVA Table	92
7 Multiple Comparisons Table (Post hoc table)	92



LIST OF FIGURES

Figure	Page
2.1 Speech production model	34
2.2 Theoretical framework	61
3.1 Research design model	64
3.2 Research framework	74
3.3 Procedural framework for the data collection	75



LIST OF ABBRIVATIONS

- ACTFL:** American Council on the Teaching of Foreign Languages
- ACTFL OPI TTM:** ACTFL Oral Proficiency Interview Tester Training Manual
- ANOVA:** Analysis Of Variance
- ASLPR:** Australian Second Language Proficiency Rating
- CLA:** Communicative Language Ability
- CLT:** Communicative Language Teaching
- EAP:** English for Academic Purposes
- FSI:** Foreign Service Institute
- IELTS:** International English Language Test System
- OPI:** Oral Proficiency Interview
- SPEAK:** Spoken English Assessment Kit
- TESL:** Teaching English as a Second Language
- TLU:** Target Language Use
- TOEFL:** Test Of English as a Foreign Language
- TSE:** Test of Spoken English
- UPM:** Universiti Putra Malaysia



CHAPTER 1

INTRODUCTION

Background of the Study

The testing of speaking has a long history, but it was not until the 1980s that the direct testing of L2 oral proficiency became commonplace, due, in no small measure, to the interest at the time in communicative language teaching. Oral interviews, of the sort developed by the Foreign Service Institute (FSI) and associated US government agencies (and now known as oral proficiency interviews-OPI), were long hailed as valid direct tests of speaking ability. Recently, however, there has been a spate of criticisms of oral interviews, which have in their turn generated a number of research studies. Discourse, conversation and content analyses show clearly that the oral proficiency interview is only one of the many possible genres of oral test tasks and the language elicited by OPIs is not the same as elicited by other types of tasks, which involve different sorts of power relations and social interaction among interactants. An influential set of guidelines for the assessment of oral language proficiency was published in 1986 by the American Council on the Teaching of Foreign Languages (the ACTFL guidelines--ACTFL, 1986).



The ACTFL proficiency Guidelines, the result of collaboration between U.S. government testing agencies, the American Council on the Teaching of Foreign Languages (ACTFL), and the Educational Testing Service, brought to academic foreign language professionals a framework for understanding and measuring oral language ability. It is uncontroversial to state that today, just 20 years later, the terms *oral proficiency*, *OPI*, and *ACTFL Guidelines* are common currency in the discourse of foreign language teachers and Pre-service teacher candidates.

This growing interest in the testing of speaking was accompanied by a corresponding expansion of research into how speaking might best be assessed. There have been a number of studies investigating the construct validity of the ACTFL OPI (e.g., Raffaldini, 1988; Valdes, 1989; Dandoloni & Henning, 1990; Henning, 1992; Alonso, 1997). The validity of the scores and the rating scale was also researched (e.g., Meredith, 1990; Halleck, 1992; Huebner & Jensen, 1992; Reed, 1992; Marisi, 1994; Glisan & Foltz, 1998). The rater's behavior and performance was also studied (Barnwell, 1989; Thompson, 1995). As the result of these studies, the most problematic theoretical as well as operational impediments of the OPI were recognized and pinpointed.

Despite the problems which exist with OPI rating scales and inequality between interviewer and interviewee proficiency, the OPI has flourished as an evaluation measure in both high-stakes decision making (e.g., hiring, promotion and employment)



and low-stakes proficiency evaluations (e.g., students' academic placement) (Thompson, 1995).

As for the ACTFL Guidelines, there are also some criticisms to the validity of these Guidelines. It was objected that the Guidelines had been constructed based on "intuitive judgments" (Fulcher, 1996) rather than on any documented collection and analysis of empirical evidence. Another criticism related to the validity of the ACTFL rating scale concerned positing the perception of the native speaker as a criterion against which the proficiency of non-native speakers would be measured (Liskin-Gasparro, 2003).

Despite these criticisms of the ACTFL Guidelines, these guidelines as many still regard are of the best frameworks for analyzing students' oral performance. Educators and employers who use the ACTFL OPI can expect reliable results and use the scores generated from the process with increased confidence (Surface & Dierdorff, 2003).

Testing oral proficiency has become one of the most important issues in language testing since the advent of communicative language teaching where speaking ability assured a more central role in language teaching (Nakamura, 1993). As Bostwick and Gakuen (1995) state, assessment can be used to improve instruction and help students take control of their own learning. Even though many students have mastered basic listening and speaking skills, some students are much more effective in their oral communication than others. In other words, some students have much more communicative effectiveness than others. "Communicative effectiveness" refers to the



test taker's ability to talk at length on a range of topics displaying a range of functional and discursal skills (description, argument, narration and speculation) (Brown, 2003).

The ability to use language communicatively involves both knowledge of and competence in the language, and the capacity for implementing, or using this competence (Widdowson, 1983; Candlin, 1986). The skills that can make the difference between minimal and effective communication can be taught, practiced, and improved. Communicative effectiveness could be directly influenced by the decision made over the test designs adopted by test takers. Test design affects the successful elicitation of the test-taker's discursal knowledge and language competence. It is necessary to be careful in selecting the appropriate design of test based on the purpose of the assessment, if a clear picture of the test-taker's effective communicative competence is intended.

There are many difficulties involved in the construction and administration of any speaking assessment. There is a great discrepancy between the predominance of the Communication Approach and the accurate measurement of communication ability (Hughes, 1989). The method used for assessing oral communication skills depends on the purpose of the assessment. A method that is appropriate for giving feedback (formative) to students who are learning a new skill is not appropriate for evaluating students at the end of the course (summative). However, any assessment method should adhere to the measurement principles of reliability, validity, and fairness. The instrument must be accurate and consistent, it must represent the abilities we wish to

measure, and it must operate in the same way with a wide range of students. The instrument and condition should not cause “criterion contamination” to the condition when an assessment produces scores or rating that measure other constructs or factors beyond the one of interest and therefore constitutes a major threat to validity (Surface & Dierdorff, 2003). The concerns of measurement, as they relate to the oral communication, are highlighted below.

Defining the domain of knowledge, skills, or attitudes to be measured is at the core of any assessment. Most people define oral communication narrowly, focusing on speaking and listening skills separately. Of course, it is something unfeasible to separate those two skills in assessing speaking. Since assessing speaking is an interaction between a sender and a receiver (reciprocal), in order to analyze the test-taker’s speech, examiner need to take into account the two skills as integrated skills not independent ones. Traditionally, when people described speaking skills; they did so in the context of public speaking. Recently, however, definitions of speaking have been expanded. One trend has been to focus on communication activities that reflect a variety of settings: one-to-many, small group, one-to-one, and mass media. Another approach has been to focus on using communication to achieve specific purposes: to inform, to persuade, and to solve problems. The purpose of the speakers will help to define the structure and focus of the interactions (Weir, 1993). A third trend has been to focus on basic competencies needed for everyday life -- for example, giving directions, asking for information, or providing basic information in an emergency situation. The latter approach has been taken in the Speech Communication Association’s guidelines for



elementary and secondary students. Many of these broader views stress that oral communication is an interactive process in which an individual alternately takes the role of speaker and listener, and which includes both verbal and nonverbal components.

There are a number of formats for testing spoken interaction in which some of them have been placed in the direct and the others in the indirect end of this assessment continuum. According to Weir (1993), examples of formats for indirect testing speaking are sentence repetitions, mini-situations on tape and information transfer. Weir puts different types of interviews, role-plays and information gaps in the direct form of testing speaking.

Although many English teachers are interested in CLT, communicative assessment has received little attention. If it is important to know if a person can speak a second language, then it should be important to test that person's speaking ability directly (Jones, 1977). Despite the interdependence of communicative teaching and communicative assessment (Bachman, 1990), speaking assessment does not assess students' oral proficiency from the perspective of language use and communication. For example, Nagata (1995) pointed out that, rote memorization of text dialogs has been a common practice for speaking assessment in Korea. It seems that Korean English teachers do not concern themselves much with matters of validity and reliability in relation to speaking assessment. A test should be comprehensive enough in attempting all major construct relevant components of oral language proficiency as set forward by the ACTFL Guidelines lately released. These Guidelines could be approached in high-

stakes tests formally and at the level of schools informally by language teachers practicing the observed standards.

However, the need for classroom teachers to be equipped with some measurement tools to evaluate students' oral proficiency is becoming more and more important (Nagata, 1995). Speaking assessment has become a vital part of all the examinations in schools. Every school is required to perform students' speaking assessment by the Malaysian Ministry of Education and Human Resources Development at least once a year. In order to inform the students of their strengths and weaknesses, performing formative assessment is an urgent decision to make. It will lead to the students' awareness of the suitability of their strategies employed during the processes of language learning. Needless to state that administering a summative assessment in a regular time interval locates students' positions among other classmates' ranking.

Speaking tasks used in tests of English for Academic Purposes (EAP) increasingly seek to replicate the roles and demands of students in academic contexts. An important but rather under researched development in EAP-test task design is that of *integrated* tasks (see Lewkowitz, 1997), in which test takers are required to process and transform a cognitively complex stimulus (e.g., a written text or a lecture) and integrate information from this source into the speaking performance. Such spoken performances are more complex and more demanding than the more traditional stand-alone or *independent* tasks, which test takers draw on their own knowledge or ideas to respond to a question or prompt. The absence of input in independent tasks means that these tasks are often

restricted to a fairly bland topic that draws on test takers' general knowledge. Consequently, these conventional speaking tasks arguably underrepresent the construct of speaking within academic contexts.

Developments within the Test of English as a Foreign Language (TOEFL) program are beginning to explore the use of independent *and* integrated tasks to assess test takers' readiness for academic study. (The rationale for the use of both independent and integrated tasks in the new TOEFL exam is articulated in some details by Enright, Bridgeman, & Cline, 2002.) With regard to these developments, the need for valid criteria for assessing performance on complex integrated tasks emerges as urgent. Given that, performance on such tasks involves the integration of cognitive skills (information selection and structuring) with more conventionally defined language proficiency skills.

It is of a paramount significance to have this clearly settled that, what is meant by the term prompt in this research project. The term 'prompt' has been in use by some researchers (Brown, 2003) to refer to the questions posed to the interviewee from the content of which he might have picked up a rough understanding of the topic from the linguistic content of the question. In this project, attempts are made to employ 'prompt' to refer to those 'written stimulus' employed to have triggered the participants' conscious and unconscious management of the language produced. The texts manipulated, so-called the prompts, would include some target vocabulary of the research interest related to the topic of the interview.

