



Payame Noor University

School of Letters and Humanities

Department of Linguistics and Foreign Languages

**THE RELATIONSHIP BETWEEN TEST-TAKERS' AND EXPERTS'  
PERCEPTION OF PROMPT DIFFICULTY**

By: **Nastaran Safi**

Advisor: **Dr. G. Hessamy**

Reader: **Dr. F. Hemmati**

A Thesis Submitted to the Graduate Studies Office in Partial Fulfillment of the Requirements

for M.A in English Language Teaching

**Tehran- Iran**

**October, 2010**

**Dedicated to**

**MY BELOVED FATHER, MOTHER,  
AND MY HUSBAND**

## **ABSTRACT**

Writing performance is greatly affected by mostly three broad factors: writer characteristics, rater characteristics, and characteristics of the prompt. Among them, prompt variables have been reported to have particular effects on test-takers' performance. Exploring validity in writing assessment, an aspect which has been largely neglected is test-takers' perception. Following, the current research was conducted to see prompt difficulty from test-takers' point of view in the case of IELTS examination task 2. The purpose of the study was to investigate a) the difference between test-takers' and experts' perception of prompt difficulty, b) the difference between male and female test-takers and their perception of prompt difficulty, c) the difference between test-takers at different age groups and their perception of prompt difficulty. The research was conducted in two phases, quantitative and qualitative. In the quantitative phase - using ex post facto as its research method - 120 IELTS test-takers and 10 IELTS experts participated in the survey questionnaire. In the qualitative phase, ten interviewees from among the questionnaire respondents participated in twenty- minute semi-structured individual interviews. The data analysis also had two phases. The researcher utilized SPSS 16 to analyze the quantitative data with descriptive statistics, t- test, and ANOVA. According to findings of this study, there was a difference between test-takers' and experts' perception of prompt difficulty. Conversely, no difference was found between the test-takers at different age groups and their perception of prompt difficulty, and between the test-takers' gender and their perception of prompt difficulty either. Findings of interviews responses were content analyzed and revealed some difficulties Iranian EFL test-takers have encountered in terms of prompt variables. Based on the findings, the researcher provided some implications for teachers and IELTS course designers. It is hoped that

test developers design prompts in a way that test-takers have the equal chance to show their writing ability.

## **ACKNOWLEDGEMENTS**

Learning is a journey; this journey would not have been possible for me if it was not for my companions who made this process as rewarding as it could be. As Newton said, 'If I have seen a little further, it is by standing on the shoulders of Giants! I am grateful to these Giants in my life, who have made my journey possible. I would like to express my sincere gratitude to my supervisor, Dr Hesammy, for his patient support, invaluable and insightful suggestions, and expert guidance throughout all stages of this project. I truly appreciate his mentorship and friendship. My heartfelt thanks go to Dr. Hemmaty for helpful advice and valuable suggestions during the proof-reading.

My sincere thanks go to Mr. Ashrafi for his useful advice and great support in statistical section of this project. I also wish to thank Mrs. Mehran who spared no effort to provide access to the references of this study. Lastly, I wish to express my heartfelt gratitude to my beloved father, mother, and my husband for their loyal support.

## TABLE OF CONTENTS

	Page
Dedication .....	iii
Acknowledgements .....	vii
List of Figures.....	viii
List of Tables.....	ix
Abstract.....	x

### Chapter One: Introduction

1.1 Background .....	2
1.2 Statement of the Problem and Purpose of the Study .....	3
1.3 Significance of the Study .....	4
1.4 Research Questions and Hypotheses .....	5
1.5 Limitations of the Study .....	6
1.6 Definition of key Terms.....	7

### Chapter Two: Review of Literature

2.1 Introduction .....	
2.2 The process of writing .....	
2.3. Writing assessment .....	
2.3.1. Validity and reliability issues .....	

<b>2.3.2. Task variables</b> .....	
<b>2.3.2.1. Prompt</b> .....	
<b>2.3.2.2. Prompt difficulty</b> .....	
<b>2.3.2.2.1. Test takers' perception of prompt difficulty</b> .....	
<b>2.3.2.2.2. Experts' feedback</b> .....	
<b>2.3.3. Test taker variables</b> .....	
<b>2.3.4. Rater variables</b> .....	
<b>2.3.5. Scoring procedures</b> .....	
<b>2.3.6. Text variables</b> .....	
<b>2.4. IELTS writing module</b> .....	
<b>2.4.1. Test module difficulty</b> .....	
<b>2.4.2. Task type</b> .....	
<b>2.4.3. Test purpose</b> .....	
<b>2.4.4. Authenticity</b> .....	
<b>2.4.5. Scoring</b> .....	
<b>2.5. Prompts in IELTS task 2</b> .....	
<b>2.6. Conclusion</b> .....	

## **Chapter Three: Method**

<b>3.1 Introduction .....</b>	
<b>3.2 Participants .....</b>	
<b>3.3 Research Design .....</b>	
<b>3.4. Instrumentation .....</b>	
<b>3.4.1. Validity.....</b>	
<b>3.4.2. Pilot testing.....</b>	
<b>3.5. Procedure.....</b>	
<b>3.6. Data analysis.....</b>	
<b>3.6.1. Quantitative data analysis.....</b>	
<b>3.6.2. Qualitative data analysis .....</b>	

## **Chapter Four: Results and Discussion**

<b>4.1 Introduction .....</b>	
<b>4.2 Descriptive statistics.....</b>	
<b>4.3 Inferential statistics .....</b>	
<b>4.4 Discussion .....</b>	

## **Chapter Five: Conclusion**

<b>5.1 Introduction .....</b>	
-------------------------------	--



**5.2 Pedagogical implications .....**

**5.3 Suggestions for Further Research .....**

**Appendices.....**

**References .....**

## **TABLE OF APPENDICES**

Appendix A: Sample of IELTS writing task 1 & 2

Appendix B: IELTS band scores

Appendix C: Oxford Placement Test (OPT)

Appendix D: Items of prompt difficulty

Appendix E: checklist for preparing writing tasks

Appendix F: A sample of questionnaire

Appendix G: Interview questions

## **LIST OF FIGURES**

- Figure 2.1 The Hayes-Flower writing model
- Figure 2.2 structure of the knowledge-telling model
- Figure 2.3 structure of the knowledge-transforming model
- Figure 2.4 Factors in writing assessment
- Figure 2.5 A multiple-trait scoring rubric
- Figure 4.1 Gender of the participants
- Figure 4.2 Distributions of scores on prompt difficulty

## **LIST OF TABLES**

- Table 2.1 A comparison of genre and process orientations
- Table 2.2 Features of direct and indirect assessment
- Table 2.3 Main types of validity
- Table 2.4 Dimensions of tasks for direct writing assessment
- Table 2.5 characteristics of the test taker
- Table 2.6 Results of Reid's study of learning style preferences
- Table 2.7 Advantages and disadvantages of holistic scoring
- Table 2.8 Advantages and disadvantages of analytic scoring
- Table 2.9 A comparison of holistic and analytic scales on six qualities of test usefulness
- Table 2.10 IELTS impact study student and teacher perceptions of IELTS module difficulty
- Table 3.1 Reliability statistics
- Table 4.1 Descriptive statistics for gender
- Table 4.2 Descriptive Statistics for Age groups

Table 4.3 Independent t-test for the Test-Takers' and Experts' Perception of Prompt

Difficulty in IELTS Writing Examination

Table 4.4 Descriptive Statistics for Test Takers' and Experts' Perception of Prompt Difficulty in

IELTS Writing Examination

Table 4.5 Independent t-test for Male and Female Test Takers' Perception of Prompt Difficulty

Table 4.6 Descriptive Statistics for Male and Female Test Takers' Perception of Prompt

Difficulty in IELTS Writing Examination

Table 4.7 Descriptive Statistics for Age Groups

Table 4.8 Levene Test of Homogeneity of Variances for Pretest

Table 4.9 One-Way ANOVA Prompt Difficulty by Age Groups

Table 4.10 Test-takers' and experts' opinions on Causes of difficulty in writing

Table 4.11 Test takers' and experts' preferences for the number of words in the prompt

Table 4.12 The effect of prompts out of cultural norms on prompt difficulty

Table 4.13 The effect of generality of the prompts in prompt difficulty

Table 4.14 The necessity of having specific knowledge in IELTS writing Task 2

# Chapter One

## *1.1. Introduction*

The ability to write has always been one of the essential skills in our global community whether it is used for communicative purposes or educational ones. Writing correlates highly, actually the highest of the four language skills, with overall language proficiency (Koda, 1993). Discussing the nature of writing, Hamp-lyons and Kroll (1997) see writing as an individual purposeful process which addresses intended audience and takes place within a context. However, it is important to view writing not just as a personal act involving some cognitive processes but in a sense a social and cultural phenomenon.

When it comes to assessing writing ability, defining the construct is one of the fundamental concerns which requires determining factors which are involved in real- world language use and the extent to which they influence what we want to measure (Weigle, 2002).

What does it mean to test writing ability? The simplest answer to this question could be “the best way to test people’s writing ability is to get them to write” (Hughes, 1989:75). However, designing a fair test is more than simply asking someone to write about a topic and then using our own criteria to judge about their writing performance. There are some factors influencing test performance which are not related to the construct being assessed, to which Messick (1989) refers as “construct-irrelevant-variance”. Construct irrelevant difficulty occurs when “aspects of

a task, extraneous to the focal construct make the test irrelevantly more difficult for some individuals or groups” (Mesick 1989, p.34). Generally there are three broad kinds of factors related to the difficulty in writing assessment: writer characteristics, rater characteristics, and characteristics of the prompt. It may not be possible to measure the effect of all of these factors exclusively since they interact with each other in complex ways which make the process of assessment even more difficult.

For instance, investigating factors which are specifically related to every individual test-taker, Willson (1982), and Zeinder & Bensussan (1988) sought for gender differences. Berry (2004) found that interest or motivation may influence the student’s performance. Hyland (2003) with reference to Grabe and Kaplan (1996) draws attention to cultural influences from L1 writing, though Zeinder (1988) found sociocultural background variable as a weak predictor of test-taker characteristics.

More specifically in writing assessment prompt variables have been reported to have particular effects on the test-takers’ final product and are considered as a potential threat to validity of the test (see Kroll and Reid, 1994; Reid 1990, Lee, 2008; Koda, 1993; Douglas, 2000) These variables include wording of the prompt, subject matter, topic familiarity, patterns of exposition, amount of specification, genre, information load, etc.

The wording of essay prompts has an effect on test scores but this has not been demonstrated empirically (Kroll and Reid, 1994). Ruth and Murphy (1988) summarized one study of “Information load” conducted by Brossell (1986) and concluded that longer prompts tend to introduce problems. Reid (1990) argues that it is a reasonable assumption that we would write better about a familiar topic than an unfamiliar one. However, Lee (2008) found no significant

difference between writing performance on a field-specific writing and a general topic test. Douglas (2000) found the content knowledge as a significant variable in test performance. The effect of discourse mode and patterns of exposition on writing performance was the concern of a study by Hamp-lyons and Mathias (1994). Carlson (1985), Spaan (1993), Hamp-lyons & Pochnow (1990) and Reid (1990) studied the interaction between topic and task type affecting the writers' performance. Concerning rhetorical function, Koda (1993) found that narrative discourse involves more demanding linguistic processing than descriptive discourse so narration is more difficult than description.

Understanding the role of prompt difficulty will provide a base for reducing the effect of bias or extraneous influences unrelated to the quality of essays. "When a prompt is biased against some test-takers it could be considered a potential threat to validity; therefore, "Such unintended consequences would need to be evaluated, when establishing the consequential basis for assessment validity." (Messick 1989. p.21)

## ***1.2. Statement of the problem and purpose of the study***

In measuring learners' writing ability numerous factors other than writing ability affect their performance. In order to make a valid judgment based on a writing test, it is necessary to determine which factors affect test performance and to what extent. One such factor is the prompt that is the stimulus for the test-taker to respond to. Writing performance is greatly affected by task-related elements such as subject matter, topic familiarity, patterns of exposition, amount of specification and information load. Besides test specification, test performance can be affected by the test-taker's characteristics. These characteristics include age, gender, sociocultural background and level of proficiency. It should be mentioned that these factors do

not affect writers' performance each separately, but interact with one another to affect the writing performance (Chapelle & Roberts, 1986).

In exploring validity, an aspect which has been largely neglected is the test-taker's perception. Without considering test-taker's challenge when they tackle the prompt for writing, test developers are hindered in their attempt to improve validity. Beyond the test developers' perception and experts' judgment of the validity, therefore, we need to examine the issue from the test-taker's point of view as well. The purpose of this study was to investigate the relationship between test-takers' and experts' perception of prompt difficulty. For this purpose, we focused on IELTS examination which is one of the most widely-used tests of English proficiency. In spite of its importance, test-takers' perception of prompt difficulty in IELTS writing examination (task 2) specifically about Iranian EFL learners has been the concern of no study so far. This study attempts to elicit such information.

### ***1.3. Significance of the study***

Validation has always been one of the main concerns of language testing. Traditionally validity evidence has been gathered in three different categories: content-related, criterion-related and construct-related validity. Recent theories on validity view it as a unitary concept. Messick (1989) states that one of the threats to validity of any test is the potential for construct-irrelevant variance that exists "when test contains excess reliable variance that is irrelevant to the interpreted construct" (p.34); thus, test developers always attempt to minimize the sources of construct-irrelevant variance; they try to examine test specifications constantly to reduce the impacts of these threats on test validity.

Nowadays the IELTS exam is probably the most widely-used test of English proficiency in the world and hence, its validity should be studied from different perspectives for different groups of test-takers. Since the test-takers' writing ability is assessed based on only one prompt, without alternative options to choose from, it is important to ensure that all examinees are exposed to tests of equivalent difficulty. A major concern about IELTS writing is the impact of the prompt on performance. This issue has been looked at from different aspects, including familiarity with the topic, wording and length of the topic, genre in terms of form (e.g. letter, report, or essay) or function (communicative function or discourse mode), subject matter, and the amount of specification and information load. Some of related studies have reached particular conclusion, though many other studies have not led to a conclusive result (see e.g. Taylor & Falvey, 2007; Mayor et al., 2007; Shaw and Weir, 2007; Anthony, 2007).

It is generally maintained that test-takers' characteristics, raters' attributes and rating scales can be highly influential in test-takers' performance. Admittedly their impact and interactions need extensive investigation.

#### ***1.4. Research questions and Hypotheses:***

1. Is there any difference between test-takers' and experts' perception of prompt difficulty in IELTS writing examination (task 2)?
2. Is there any difference between male and female test-takers in terms of their perception of prompt difficulty?
3. Is there any difference between test-takers at different age groups and their perception of prompt difficulty?



On the basis of the research questions mentioned above, following hypotheses were formulated:

H 1: There is no difference between the test-takers' and experts' perception of prompt difficulty in the case of IELTS writing examination (task 2).

H 2: There is no difference between males and females in terms of their perception of prompt difficulty.

H 3: There is no difference between the test-takers at different age groups and their perception of prompt difficulty.

### ***1.5. Definition of Key Terms***

#### **Prompt**

Prompt consists of the question or the statement students will address in their writing and the conditions under which they will write (O'Malley and Valdez, 1996).

#### **Assessment**

Assessment is appraising or estimating the level or magnitude of some attribute of a person. Assessment consists of taking a sample of what students do, making inferences and estimating the worth of their action. The assessment of a person's language abilities may include tests, interviews, questionnaires, observations, etc (Bachman 1990).

## **Writing Assessment**

An assessment task in which test-takers actually produce a sample of writing often referred to as a 'direct' test of writing, is probably the most common method for testing writing. Hamp-Lyons (1991) gives five characteristics of a so called 'direct' test of writing: Candidates must write at least one piece of continuous text, test-takers are given a set of instructions or 'prompts', each text is read by at least one or more trained raters, judgments are tied to a common yardstick, such as a set of sample responses or rating scales; and judgments are expressed as numbers rather than verbal descriptions (Weigle, 2002).

## **Test-taker's perception**

Perception refers to a reaction or response that is usually received by the learner or what they perceive (or pay attention to) in their environment. This response is the result of attitudes, values, opinions, judgments, and conceptions of the learner about objects, issues, and people.

# Chapter Two

## *Review of Literature*

### *2.1. Introduction*

Writing is a fundamental skill and it is assumed that it plays a central role in any language learning curriculum. Writing was neglected in the early years of studies of second language acquisition, at the rise of applied linguistics, in the late nineteenth century. The priority was given to the spoken language in the United States between the 1940s and 1960s through the work of Leonard Bloomfield and Charles Fries which ended up in audiolingual approach in which written script was mostly used to facilitate the learning of spoken language. However, the increasing number of ESL students in higher education and the need for college writing courses led to the emergence of L2 writing instruction in the U.S. higher education institutions. In terms of English as a second language, the debate for a privileged role for writing was made in the 1960s and 1970s to compensate for the relegated position of writing as the least important of language skill in audiolingual method of language teaching (Matsuda, 2001).

In response to the need to prepare students for writing composition in higher education, an approach that focused on sentence-level structure and consisted of combining and substitution exercises was designed to facilitate the learning of sentence structure. Sentence-level grammar exercises did not help students to produce original sentences, let alone free composition. Because of the limitation of controlled composition, the use of guided composition was devised in which

students were given some assistance such as a model to follow, an outline to expand from, or a partly-written text with some indication of how to complete it (Matsuda, 2003).

Both controlled and guided composition lacked “logical organization” judged by native speakers which led to this realization that writing should be viewed beyond the sentence level. Kaplan (1966) argued that the problem lied in the transfer of L1 structures and at the level of paragraph. Drawing on the principles of contrastive analysis and Sapir-Whorf hypothesis, he suggested that paragraph structures were language and culture specific which founded the principle of contrastive rhetoric (“rhetoric” was defined as the organizational structure). In the 1980s, researchers began to examine structures of written discourse in different languages and their influences on second language writing. In recent years, contrastive rhetoric research has evolved to a field of research which has provided various theoretical frameworks for investigating written discourse (Hamp-Lyons, 2003).

In the late 1970s and the 1980s, the focus shifted from features of text to the processes underlying the production of writing (Matsuda, 2003). The notion of process-based approach was pioneered by Zamel (1976) who emphasized on the process of developing organization as well as meaning (cited in Anthony, 2007). The view of writing as a process was coincided with the development of English for Specific Purposes (ESP) and English for Academic Purposes (EAP) which both viewed writing as language use in a specific context. In the late 1970s, since the ESL student population in the U.S. higher education continued to increase, the need for basic writing courses had to be met. Shagnessy (1976), a pioneer in the field of basic writing, tried to characterize the difficulties Native English Speaker (NES) writers faced in learning to produce the kind of formal writing required in higher education which led to the emergence of “writing as a second language” (cited in Matsuda, 2003). However, the challenge of ESL students to achieve