

5-2014

# Students' Perceptions of International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL) Tests

Hartati Suryaningsih  
*Indiana University of Pennsylvania*

Follow this and additional works at: <http://knowledge.library.iup.edu/etd>

---

## Recommended Citation

Suryaningsih, Hartati, "Students' Perceptions of International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL) Tests" (2014). *Theses and Dissertations (All)*. 1235.  
<http://knowledge.library.iup.edu/etd/1235>

This Thesis is brought to you for free and open access by Knowledge Repository @ IUP. It has been accepted for inclusion in Theses and Dissertations (All) by an authorized administrator of Knowledge Repository @ IUP. For more information, please contact [cclouser@iup.edu](mailto:cclouser@iup.edu), [sara.parme@iup.edu](mailto:sara.parme@iup.edu).

STUDENTS' PERCEPTIONS OF INTERNATIONAL ENGLISH LANGUAGE  
TESTING SYSTEM (IELTS) AND TEST OF ENGLISH AS A FOREIGN LANGUAGE  
(TOEFL) TESTS

A Thesis

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Master of Arts

Hartati Suryaningsih

Indiana University of Pennsylvania

May 2014

© 2014 Hartati Suryaningsih

All Rights Reserved

Indiana University of Pennsylvania  
School of Graduate Studies and Research  
Department of English

We hereby approve the thesis of

Hartati Suryaningsih

Candidate for the degree of Master of Arts

---

David I. Hanauer, Ph.D.  
Professor of English, Advisor

---

Gloria Park, Ph.D.  
Associate Professor of English

---

Sharon K. Deckert, Ph.D.  
Associate Professor of English

ACCEPTED

---

Timothy P. Mack, Ph.D.  
Dean  
School of Graduate Studies and Research

Title: Students' Perceptions of International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL) Tests

Author: Hartati Suryaningsih

Thesis Chair: Dr. David I. Hanauer

Thesis Committee Members: Dr. Gloria Park  
Dr. Sharon K. Deckert

This study aims to explicate the differences between test takers' IELTS and TOEFL testing experiences. The basis of this research is the idea that test takers' testing experiences are important because they can provide information that might not be recognized by testers (Shohamy, 2001a). The data of this study were collected by interviewing six graduate level students who have taken both IELTS and TOEFL tests. The results are presented in three different tables with their descriptions and explanations. The analysis of the results shows that the participants in this study perceived IELTS and TOEFL tests differently in three aspects: on the tests, in the tests, and effects of the tests. Furthermore, this study also presents information of high-stakes tests testing experience, concept of ideal tests, and critiques of the effects of the tests. Finally, this study provides some implications and recommendations for future research.

## ACKNOWLEDGEMENTS

This thesis is the representation of other people's supports and help. I would like to thank my advisor, Dr. David I. Hanauer, for guiding, advising and supporting me to conduct this research. The lessons I learned from you throughout the process of writing this thesis helped me to know more about research. I also would like to thank my committee members, Dr. Gloria Park and Dr. Sharon K. Deckert, for the supports and understanding when I was writing this thesis.

I also would like to express my gratitude to the faculty members of IUP MA TESOL program. Without their helps, I would not be able to complete my study and gain insight about theories and practices of issues around teaching and learning English. I especially would like to thank Dr. Gloria Park as my academic advisor. Without your guidance, advices, and encouragements during my two years study, I would not be able to be where I am now.

I thank all my participants in this study. Without their helps, this thesis would not be able to be conducted. Their helps are very meaningful for me, not only because their participation in the research, but their ideas, opinions and perceptions enrich my knowledge about different things in the world. My MA TESOL friends who always supported me Adelay, Berry eonni, Jocelyn, Meghan, Mohammed, Muna, Najah, Rossella, Sherry, Suad, Wenxi, and as well as other friends. I especially want to thank my comrades in writing thesis: Muna, Rossella, and Wenxi, who were always with me during the hardest time.

My family: Bapak Wasis Mujiono, Mamak Siti Sanimah, Tari and Tami, thank you for listening to my complaints, grudges and nags all the time during my two years

study. My Indonesian family in Indiana, Dr. Kustim Wibowo, Suwarni Wibowo, Dr. Nurhaya Muchtar, Mas Deddy, Mba Rini, Ghibran, Ghina and Jessie, I would like to thank you because you make me feel loved. I also would like to thank Rebeka, Tisya, Jinhee, Odai, Ali, Nes Eonni, Tebel, Erika, Gobe, Isna and Bela Pertiwi who always make me laugh of their jokes. Also, I want to express my gratitude to Mba Indah Puspawati, because she opened my eyes to this topic. I thank the Fulbright Program which gives me chances to pursue my Master degree in Indiana University of Pennsylvania.

I dedicate this thesis to the all people I mentioned above. Without their caring and love, I would not be able to finish it. At the end, I would like to say that this thesis is my first step to enter research world in which I can share my knowledge and information about issues around English teaching and learning practices.

## TABLE OF CONTENTS

Chapter		Page
One	INTRODUCTION.....	1
	Why IELTS and TOEFL Tests? .....	1
	Researcher’s Personal Experiences.....	2
	Studies on Language Tests.....	4
	Purpose of the Study .....	6
	Research Question .....	6
	Significance of the Study .....	6
	Overview of Upcoming Chapters .....	7
Two	LITERATURE REVIEW.....	8
	High-Stakes Language Testing.....	8
	The IELTS and TOEFL as High-Stakes Tests.....	11
	Introduction to IELTS and TOEFL Tests.....	11
	Effects of High-Stakes Tests .....	17
	Studies on Students’ Perceptions of High-Stakes Tests .....	23
	Impacts of Tests.....	24
	Problems in the Test .....	29
	Factors Affecting Students’ Performance on the Test.....	32
	Democratic Assessment .....	35
	Conclusion .....	39
Three	METHODOLOGY .....	41
	Research Design.....	41
	Participants.....	42
	Data Collection Method.....	43
	Semi-structured Interview .....	43
	Individual Interview .....	44
	Potential Drawbacks and Countermeasures .....	46
	Data Collection Procedure .....	46
	Recruiting Participants .....	46
	Interview.....	47
	Data Analysis .....	48
	Transcription .....	48
	Coding and Categorizing.....	50
	Summary .....	53
Four	RESULT .....	55
	Participants’ Perceptions of the TOEFL Test .....	55



Chapter	Page
Testing Experience .....	55
Topic in the Test.....	58
Tasks in the Test.....	60
Time in the Test.....	63
Score of the Test.....	65
Effects of the Tests .....	66
Participants' Perceptions of the IELTS test .....	69
Testing Experience .....	69
Topic in the Test.....	74
Tasks in the Test.....	75
Time in the Test.....	77
Score of the Test.....	79
Effects of the Test.....	80
The Comparison of Participants' Perceptions of the IELTS and TOEFL Tests .....	82
Conclusion .....	85
Five DISCUSSIONS AND CONCLUSIONS .....	86
Different Perceptions of IELTS and TOEFL Tests .....	86
On the Tests Aspects .....	87
In the Tests Aspects.....	88
Effects of the Tests Aspects .....	92
Conclusion .....	93
Implications of the Study .....	95
Limitation of the Study .....	96
Recommendations for Future Studies .....	96
Final Reflections .....	96
REFERENCES .....	98
APPENDICES .....	111
APPENDIX A- IRB Approval.....	111
APPENDIX B- RTAF Approval .....	112
APPENDIX C- Informed Consent Form .....	113
APPENDIX D- E-mail Protocol .....	115
APPENDIX E- Interview Protocol .....	116

## LIST OF TABLES

Table	Page
1 TOEFL Test Development .....	13
2 Introduction to IELTS and TOEFL Tests.....	16
3 Participants' Demographic Information.....	43
4 Transcription and Excerpts Features .....	49
5 Participants' Perceptions of the Tests .....	53
6 Frequency of Participants' Perceptions of the TOEFL Test .....	69
7 Frequency of Participants' Perceptions of the IELTS Test.....	82
8 Comparison of Participants' Perceptions of the IELTS and TOEFL Tests .....	83

## CHAPTER ONE

### INTRODUCTION

“It is difficult to find a person who does not have a testing story that relates to how a single test affected and changed his or her life” (Shohamy, 2001a, p. 7). The language test is one of the examples of this phenomenon. However, there are not many studies focusing on test takers’ experiences even though language tests play important roles in their lives. Thus, this research tries to convey test takers’ perceptions of two high-stakes tests, the IELTS and the TOEFL tests.

This chapter has five main sections. The first section is related to the background of the study followed by sections on the purpose, research question, and significance of the study. At the end of this chapter, an overview of upcoming chapters is presented.

#### **Why IELTS and TOEFL Tests?**

Language tests are used for many purposes in many contexts and imposed as one of the important aspects in professional life. A discussion regarding language tests cannot be separated from International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL) tests. Both tests are “almost the household of the professional circle” (Leung & Lewkowics, 2006, p. 222). The IELTS and TOEFL are two tests for English as Foreign Language (EFL)/English as Second Language (ESL) learners to measure their English proficiency. The TOEFL test aims at measuring English proficiency of EFL/ESL students intending to continue their education at higher institutions. It evaluates the test takers’ English skills: listening, reading, speaking, and writing regarding how they perform on academic tasks (ETS, 2013). The IELTS test measures academic and general English language proficiency. The academic test is

addressed for those willing to pursue their study at a post-secondary institution in an English-speaking country. Meanwhile, the general training test is for people willing to work, go to secondary school or migrate to English-speaking countries (IELTS, 2013).

These two tests have been used for many purposes which give them a powerful position in society. Due to their position as high-stakes tests, test takers strive to perform well on the tests because it can guarantee them a better opportunity to apply for jobs, schools and scholarships. Besides, the detrimental effects of tests also make test takers try their very best to obtain a high score. Tests can create a gap between people where the failed people will be seen as losers, and the successful people will be seen as the winners. Furthermore, tests also have power to control what should be taught and what should be learned. In other words, they have power to control the knowledge of people. This is something critical about testing where it is used as a disciplinary tool, meaning the test takers, whether they are willing or not, have to change their behavior in order to meet the tests demands (Shohamy, 2001a).

### **Researcher's Personal Experiences**

My experiences as an English student and as an English teacher portrayed the powerful position of the tests use and their effects. As an English student, I witnessed how my colleagues struggled to get the required score needed in order to apply for scholarships. They got rejected because of their low score, which in my opinion did not represent their real proficiency. Also, there are many people that I know who are forced to delay their intention to continue their study at higher institutions due to a low score on the tests.

As an English teacher, I saw how tests have been used to group students based on their language proficiency. In addition, I witnessed that some people were willing to spend their money in the courses to achieve the score required by the institutions in which they want to register. They bought more books and asked English teachers to teach them privately at home. Zhengdong (2009) stated, “the higher the stakes of the test, the stronger the urge to engage in specific test preparation practices” (p. 24). Furthermore, some people also paid other people to take tests on their behalf. This shows students’ fear of not meeting the required score. How they obtain the score is less important than getting the required score. This phenomenon shows that the issue of the power of the tests has led some people to engage in “unethical” practices.

Based on these phenomena, I acknowledged Shohamy’s idea (2001a) as she stated, “when a language is tested in a higher institution as criteria of acceptance, the symbolic power of the test and the language are enhanced significantly presenting the society with an extremely powerful tool” (p. 123). This power gives tests a strong position to determine students’ future lives. Green (2006) also stated that performance on the test may have a big impact on test-takers’ lives.

Furthermore, these experiences show me that listening to test takers is important, because they are the people who are most affected by the test (Shohamy, 2001a). However, there is not much heard about their experiences. I only could listen to their stories telling that the TOEFL test did not measure their English, thus it did not represent their real proficiency. They mentioned problems and reasons why they claimed that the TOEFL test was problematic. Therefore, this issue raises a question, are these complaints

about the TOEFL or complaints about language test? The way to answer this question is by investigating people who have taken more than one language test.

### **Studies on Language Tests**

Since IELTS and TOEFL tests are imposed as important subjects in society, there are many studies focusing on the validity and reliability of these two tests. Winke, Gass, and Myford (2012) investigated L2 (Second Language) speech raters' bias in the TOEFL test. They found that L2 raters were more tolerant with the test takers coming from the same country as them. In other words, if the L2 raters are coming from China, they will be more lenient to the L1 Chinese test takers. This poses the reliability problems of the test.

Furthermore, a study conducted by Bridgeman, Powers, Elizabeth, and Mollaun (2011) investigated the speaking test score rated by professional test raters and SpeechRater™ - an automated speech scoring system developed by Educational Testing Service (ETS). The result showed there is a difference in scores given by these two raters. The professional TOEFL iBT raters gave a higher score for the test takers than the SpeechRater™. The SpeechRater is able to measure some aspects of communicative competence; however, it missed some important components such as: "coherence, progression of ideas, and content relevance" (p. 100) which could be identified by the expert test raters.

Regarding the IELTS test, some studies showed that this test poses some problems. Brown (2003) found that the problem in the IELTS test was on the interviewers. She found that interviewers could affect the performance of test takers by the way they elicited information from them. Furthermore, Uysal (2010) noted and

criticized that the writing test in the IELTS more focus on the “form rather than the meaning and intelligibility” (p. 317). These studies on IELTS and TOEFL tests show that even though these two tests are powerful, they possess some problems picturing that the tests might not measure test takers’ English proficiency effectively.

As explained above, most of the studies focus on investigating the tests regarding their validity, the content of the tests, and the students’ academic performance (Hill, Storch, & Lynch, 1999; Zahedi & Shamsaee, 2012). The research regarding test takers’ perceptions of tests is limited. Listening to test takers’ voices is important because it provides information that might not be addressed by other stakeholders (Shohamy, 2001a). Therefore, besides focusing on investigating the tests, it is essential to explore the consequences of tests as a matter of validity (Messick, 1998). As the people who have been through the tests and experienced the consequences, test takers will be able to provide information from different perspective which might help practitioners and stakeholders to gain more understanding about language tests. The problem might be either in language tests or the test itself based on their perspectives.

Therefore, one reason for looking at people who have taken both tests is because there is an opportunity to compare students’ perceptions between the tests. Another reason is related to language tests because it can give a sense of what large tests are like in themselves. Furthermore, test takers’ perceptions are important because they provide evidence of tests’ effectiveness (Tsai & Tsou, 2009). As Shohamy (2001a) argued that the investigation of the tests based on test takers’ experiences is significant because it provides stakeholders, especially testers, with new understanding about tests and their meanings.

### **Purpose of the Study**

The main purpose of this study is to provide deeper understanding of language tests based on test takers' perceptions. In other words, this research attempts to emphasize the importance of listening to test takers' voices as a way to understand language tests. In order to achieve this purpose, this research explicates students' different perceptions of IELTS and TOEFL tests based on their testing experiences. Furthermore, this study can be a way to find out whether the problems that appear in the high-stakes tests are related to languages test in general or the specific test itself. The participants of this research were six graduate students who have taken both the IELTS and TOEFL tests. All of the participants are currently enrolled as graduate students in the universities in the United States. These participants shared their experiences of taking both tests in an individual interview.

### **Research Question**

This study investigates students' perceptions of IELTS and TOEFL tests. This study compares their perceptions of whether they see and experience the two tests equally or differently. In order to find the answer of this question, the research question is:

Do master level students who have taken both IELTS and TOEFL tests perceive the two tests differently? If so, in what ways?

### **Significance of the Study**

As this research is an endeavor to emphasize the importance of listening to test takers' experiences, there are five significant contributions of this study. First, the findings of this study might contribute to increase other stakeholders' awareness that test takers' voices can provide information about language tests from different perspectives.



Second, the findings might possibly help practitioners and stakeholders to understand the way participants in this study perceived high-stakes tests and experienced them. Third, this study will provide participants' different perceptions which will deepen our understanding of language tests and the impact of the tests on test takers' lives. In addition, this research gives the participants chances to share their thoughts and experiences of taking both tests, and offers ideas about how language tests should be. Finally, language assessment has become a growing industry which drives the development of many language teaching courses. This research may be useful in increasing testers' awareness about the consequences of tests.

### **Overview of Upcoming Chapters**

This thesis has five chapters. Following this chapter, Chapter Two will be a literature review providing information about high stakes language testing, the IELTS and TOEFL tests, studies on students' perceptions of high stakes tests and the notion of democratic assessment. After that, Chapter Three, the methodology, presents the way this study was conducted such as the research design, participants, data collection method, data collection procedure, and data analysis. Next, Chapter Four provides the result of this study. Finally, Chapter Five presents the discussion and conclusions of this study. This chapter also provides information about the implications and the limitations of this study, recommendations for future studies and final reflections.

## CHAPTER TWO

### LITERATURE REVIEW

The aim of this study is to investigate the following research question: do master level students who have taken both IELTS and TOEFL tests perceive the two tests differently? If so, in what ways? Based on this purpose, the topic of this study covers the knowledge of IELTS and TOEFL tests and studies about tests. To have an understanding and current development of this issue, the literature review is presented in four main sections. The sections in this literature review are adapted from Puspawati's (2012) thesis entitled EFL (English as Foreign Language)/ESL (English as Second Language) students' perceptions towards TOEFL test. Some adaptations and modifications from her thesis literature review were done in order to meet the purpose of this research. I first introduce high-stakes language testing to raise critical and ethical issues of language testing. Afterwards, I discuss IELTS and TOEFL tests as high-stakes tests: the objectives, their development, the form of the tests, and the effects of high-stakes tests. Then, I highlight and review studies focusing on students' perceptions on tests. Finally, this chapter discusses democratic assessment principles and the practices.

#### **High-Stakes Language Testing**

Many studies have found that language tests have become powerful in society. They have played crucial roles in the lives of the test takers (Uysal, 2010). The use of language tests can affect both educational and social lives of students (Shohamy, 2001a; Puspawati, 2012). Regarding students' educational lives, tests dictate students' learning. This means that their educations are based on tests, for examples: students are taught materials that will appear in the tests, and their learning processes are purposively

designed to deal with tests. Shohamy (2001a) stated that “tests are capable of dictating to test takers what they need to know, what they need to learn, and what they will be taught” (p. 17). In terms of students’ social lives, tests influence students’ behavior and attitudes. Students change their behavior in order to fit the tests demands (Shohamy, 2001a). In other words, they perceive tests not as a tool to measure their abilities and knowledge, but rather as a judgment time whether or not they can move forward in their lives. Students strive to perform well in tests due to the importance of tests result towards their future. Dooley (2008) asserted that language test results have become one of the important aspects in creating opportunities for test takers. In many countries, language tests have become one of the requirements in order to enter the next level of education or professional life. In other words, language tests have been used as the gatekeeper (Bridgeman & Cho, 2012). The language tests score is used as a standard to screen people to enter another stage in their lives. Bachman and Purpura (2008) argued that language tests have been used in many cases such as selecting students, giving certifications, and giving promotion. Reflecting on these uses, tests have marked their powerful position in society.

The power of language tests cannot be separated from the features that they have. According to Shohamy (2001a), some features make tests become powerful in society. First, powerful institutions administer tests, meaning that the testers are in a powerful position where the test takers are powerless. Testers are the people who make decisions about what to be tested, and how the test score is interpreted. Second, tests use the language of science. Since testing is perceived as a scientific discipline, tests perpetuate their position in society. Third, tests use the language of numbers. Shohamy (2001a)

stated that the “language of numbers is considered to be a symbol of objectivity, scientism, and rationalism” (p. 21). In other words, language of numbers possesses all aspects or features that are related to “truth, trust, legitimacy and status” (p. 22). Finally, tests incorporate an objective format, meaning that the truth comes from the testers. Tests do not allow test takers to have their own interpretation for the truth.

By having all the features mentioned above, language tests have become powerful in society. Language tests can either open or close opportunities for the test takers’ future dreams (Shohamy, 2001a; Zahedi & Shamsaee, 2012). In addition, language tests are able to group people in a way that certain people will be categorized as successful or failures, winners or losers depending on the test results (Shohamy, 2001a). These conditions constitute the high-stakes of language tests. Luxia (2005) argued “high-stakes tests are those whose results are used to make important decisions that immediately and directly affect the test-takers and other stakeholders” (p. 142) such as teachers and parents. Language tests, especially English, perpetuate their powerful influence in society because of their high-stakes position.

In response to the high-stakes position of language testing, students strive to perform well on the test because it can facilitate them a better opportunity in their lives. In many studies, it is found that language tests have been used as admissions requirements, promotion requirements, and job application requirement (Cheng, Andrews, & Yu, 2011; Hirsh, 2007; Murray, Riazi, & Cross, 2012; Zhengdong, 2009). Due their high-stakes, tests are able to dictate test takers’ behavior. This is something critical about testing that needs to be addressed, so that their uses can be monitored and limited.

## **The IELTS and TOEFL as High-Stakes Tests**

The reason why IELTS and TOEFL became and remain powerful in society is a crucial question. These two tests were purposively developed for measuring test takers' English proficiencies. However, their uses and effects have gone beyond their original purpose. Their uses have placed them as high-stakes tests which affect individual(s)' future lives. Shohamy (2001a) argued that high-stakes test results are crucial for the test-takers' lives in order to make important decisions. These two tests' scores have been used and imposed as one of the essential factors to make important decisions in test takers' lives such as pursuing for higher education and applying for job. Zahedi and Shamsaee (2012) pointed out that IELTS and TOEFL tests are two examples of high-stake standardized tests which have "critical roles in determining the future life" (p. 246) of the test takers. There are three purposes of high-stake tests: "selecting people for an occupation, issuing a certificate, and permitting one to enter a higher education" (p. 264). In other words, as Shohamy (2001a) argued that tests have effects not only for education but also for society.

### **Introduction to IELTS and TOEFL Tests**

TOEFL and IELTS tests were developed in the 20<sup>th</sup> century to measure test takers' language proficiency. For more than 30 years, these two tests have been through many changes that were claimed to better measure test takers' proficiency. Since that time, many studies have been conducted to investigate the effects of the test towards society. Many scholars are concerned about the topic of TOEFL of IELTS tests because of their powerful position in society.

**TOEFL (Test of English as Foreign Language).** The TOEFL test was introduced by the National Council on the Testing English as a Foreign Language in 1963. The test was developed to assess non-native English speakers who intend to pursue their study in the medium English universities (ETS, 2013). Since 1963, the TOEFL test has evolved into three formats, from a paper-based test (PBT) to a computer-based test (CBT) and to an Internet-based test, TOEFL iBT (Hill & Liu, 2012).

Besides the TOEFL test, there are two separate tests developed by ETS (Educational Testing Service): Test of Spoken English (TSE) and TWE (Test of Written English). The TSE test was introduced in 1970s in order to assess the speaking ability of international graduate students applying for teaching assistants position. The TWE test was developed in 1986 where it measured students writing ability based on a topic. From their purposes, it can be inferred that these two tests were taken by smaller numbers of applicants in order to meet the requirements for institutions (ETS, 2013).

Currently, there are two formats of TOEFL test offered by ETS: Paper-Based Testing (PBT) and Internet-Based Testing (iBT) (Alderson, 2009). TOEFL iBT was launched in September 2005 to replace the TOEFL computer-based test (CBT) (Hill & Liu, 2012). The TOEFL iBT is different from the previous versions of TOEFL test – TOEFL CBT. It assesses four language skills to be needed in the academic lives of the applicants. This test assesses Listening, Reading, Speaking, and Writing skills where in the earlier version the speaking skill was not measured (Alderson, 2009). “The primary goal of this change is to better align the test design to the variety of language use tasks that examinees are expected to encounter in everyday academic life” (Sawaki, Stricker, & Oranje, 2009, p. 5). The difference is not only in the assessed skills, but also in the

content of the skills. The TOEFL iBT reading section measures test-takers ability to understand reading passages about a topic or a discipline. In the Listening section, test-takers are measured on their comprehension, pragmatic understanding, connecting and synthesizing information of listening conversation and lecture. The Speaking section measures test-takers’ ability to express their ideas and respond to many information from multiple sources. The last section is Writing, in which test-takers have to write two essays. One essay is based on a given topic, and the other essay is based on a lecture. This writing task measures test-takers’ quality of writing: organization, grammar, and vocabulary (Alderson, 2009). The evolution of the TOEFL test constructions and content over three stages of TOEFL test development is presented in Table 1.

Table 1

*TOEFL Test Development*

Stages	Construct	Contents
1. The first TOEFL test 1964 – 1979	Discrete components of language skills and knowledge	Multiple-choice items assessing vocabulary, reading comprehension, listening comprehension, knowledge of correct English structure and grammar
2. A suite of TOEFL tests 1979 – 2005	Original constructs (listening, reading, structure, and grammar) retained but two additional ones added – writing ability and speaking ability.	In addition to multiple-choice items assessing the original constructs, separate constructed-response tests of writing, the TWE test and speaking, the TSE test, were developed.
3. The TOEFL iBT test 2005 – Present	Communicative competence – the ability to put language knowledge to use in relevant contexts	Academic tasks were developed that require the interrogation of receptive and productive skills such as listening, reading , and writing or speaking, as well multiple choice items for listening and reading.

*Note.* Adapted from “TOEFL iBT Research, insight TOEFL program history,” 2006, p. 4. Retrieved from [http://www.ets.org/s/toefl/pdf/toefl\\_ibt\\_insight\\_s1v6.pdf](http://www.ets.org/s/toefl/pdf/toefl_ibt_insight_s1v6.pdf)

**IELTS (International English Language Testing System).** The IELTS (International English Language Testing System) score is accepted by medium English universities to prove that applicants have fair skills and knowledge of English in order to pursue their studies (Green, 2005). IELTS, which was previously known as ELTS (English Language Testing Service), was introduced for the first time in the early 1980s by United Kingdom (UK) test developers motivated by the possibility to create language test for specific purposes (O’Sullivan, 2012). It is also created to replace English Proficiency Test Battery (EPTB) - a traditional multiple-choice test used by the British council as required test for international applicants to universities and colleges in the UK- (IELTS, 2013). The original ELTS was a test for students planning to study particular disciplines. There were six modules generated from that purpose: Life Sciences, Social Studies, Physical Sciences, Technology, Medicine and General Academic (O’Sullivan, 2012).

According to Clapham and Alderson, due to the new idea prospecting “between practicality and maximum predictive power” (as cited in IELTS, 2013, para. 4), the test developers shrink the six modules into three modules: Physical Science and Technology, Life and Medical Sciences, and Business Studies and Social Sciences in 1989. Each had options in reading and writing, and all candidates had the same speaking and listening papers. In 1995, IELTS test started having no domain-specific modules, and all test takers had the same papers with just an academic and a general training (non-academic) option for the test (O’Sullivan, 2012).

Another revision was completed again to the IELTS test. The IELTS test developer launched a revision project for the speaking test in 1998. As a result, the



revision was introduced in 2001 with a more rigorous, structured interview format (Interlocutor Frame), and revised rating scale (Terry, 2003). A research conducted by Brown (2006) focusing on examining the use of revised rating scales showed a positive result. She found that the new rating scale minimized examiners' inconsistency in assessing and rating test takers' performances. Another research investigating the Interlocutor Frame (IF) conducted by O'Sullivan and Lu (2006) also demonstrated a positive result. It showed that the IF prevented interlocutors to treat each interview differently. The findings pointed out that even though there was a deviation of the IF, which likely occurred in the third phase of the speaking test, it did not necessarily affect test takers' performances. In 2005, IELTS launched new assessment criteria in the Writing section. The earlier version focusing on "communicative quality" was replaced by "coherence and cohesion" (Uysal, 2010, p. 317). Based on a research conducted to evaluate the new assessment criteria, it was found that these criteria were considered improved by examiners because of its clarity and detailed descriptions of what and how to assess (Shaw, 2006). Finally, a computerized version of the IELTS was held at some of the IELTS centers in 2005 (IELTS, 2013). The differences between General and Academic IELTS are in the content, context, and the purposes of the test. The similarities are in terms of time allocation, length of written purposes and score reporting (IELTS, 2013).

Table 2

*Introduction to IELTS and TOEFL Tests*

ITEMS	TOEFL iBT	IELTS
Section (in order)	Reading (60 – 80 minutes) Listening ( 60 – 90 minutes) Speaking (20 minutes) Writing (50 minutes)	Listening (30 minutes) Reading (60 minutes) Writing (60 minutes) Speaking ( 11 – 14 minutes)
Tasks and Topic	<p><b>Reading:</b> 3 texts of a topic or discipline. Exposition, argumentation and historical narrative. 36 - 56 questions</p> <p><b>Listening:</b> 4 - 6 lectures. Multiple choices with single, and more than one correct answer, matching and sequencing items. 34 - 51 questions.</p> <p><b>Speaking:</b> Six tasks, two of the tasks ask test takers to express their opinions toward familiar topics. The other 4 tasks ask test -takers to express based on what they heard or read.</p> <p><b>Writing:</b> two tasks: independent and integrated tasks. The independent task where the candidates write a response essay towards a topic. The integrated task meaning test takers response to a reading and lecture.</p>	<p><b>Reading.</b> Three long texts taken from books, journals, magazines and newspapers ranged from the descriptive, factual, discursive and analytical texts. 40 questions</p> <p><b>Listening.</b> Four recorded texts, monologues and conversations from a range of native speakers. Answer questions based on the questions.</p> <p><b>Speaking.</b> Consists of three parts: 1. Speak about daily conversation/topic 2. Speak about particular topic given by the examiner 3. Speak about further questions related to part 2.</p> <p><b>Writing.</b> Consists of two tasks: 1. Describing and explaining chart or diagram 2. Responding to a point of view, argument, or problem.</p>
Technology	All sections done in the computer by using Internet. In the speaking and listening sections, test-takers wear headphones.	All sections are paper based. On the listening section, all test takers listen from the same audio recording. In the speaking section, test takers are interviewed by test interviewer. All conversations are recorded by using recording device.
Time (in total)	3 hours 10 minutes	2 hours 44 minutes
Score	Each section range score: 0 – 30 points. 120 is the highest score	1 – 9 band of score is given, range from the non-user to expert user

*Note.* The data on TOEFL iBT are adapted from “Test of English as a foreign language: Internet-based test (TOEFL iBT),” by J.C. Alderson, 2009, *Language Testing*, 26(4), pp. 621 – 631. The data on IELTS are adapted from “IELTS.org,” 2013. Retrieved from [http://ielts.org/test\\_takers\\_information/what\\_is\\_ielts/test\\_format.aspx](http://ielts.org/test_takers_information/what_is_ielts/test_format.aspx)

## **Effects of High-Stakes Tests**

TOEFL and IELTS have perpetuated their position as high-stakes tests in which they can influence test takers' futures. Besides that, these two tests have broad effects to society. Shohamy (2001a) argued that tests have two effects: educational and societal effects. Educational effects are related to the impacts of tests toward education, such as pedagogy and learning materials. The other effect, the societal effects, refers to the change of attitude, behavior, and perspective of the stakeholders as a result of tests. These effects could be the bases for practitioners and experts to limit the use of the tests in society.

**Educational effects of high-stakes tests.** According to Shohamy (2001a) the educational effects of tests encompass “curriculum, teaching methods, learning strategies, materials, assessment practices and knowledge tested” (p. 46). High-stakes tests have changed the curriculum and teaching practices (pedagogy) in classroom (Au, 2007, 2009, 2011; Cizek, 2001; Horn, 2003; Kohn, 2000; Templer 2004; Triplett, Barksdale, & Lepswitch, 2003). Kohn (2000) argued that curricula is fundamentally developed based on tests because both the content and the format of instruction of tests affect the curriculum. Klinger and Luce-Kapler (2007) strengthened this idea by arguing that tests have affected the curriculum.

In relation to the influence of tests towards curricula, tests also have affected teaching and learning practices. Teachers tend to focus on materials and strategies that they predict will help students in tests (Triplett, Barksdale, & Lepswitch, 2003). Their major goal is to prepare students to pass the test. Shohamy (2001a) argued that teachers change their teaching strategies in order to “help” the students to get high scores, because

their success in teaching is viewed from the students' scores in tests. From this point of view, both teachers and students become the victim of high-stakes tests. Therefore, Au (2009) claimed, "when punitive consequences are attached to test scores, teachers do indeed match their pedagogy and content to the test norms" (p. 46). The structure and type of tests dictate teaching and learning activities. For instance, both TOEFL and IELTS tests reading section consists of multiple choice questions, therefore, instructors or teachers will teach their students to analyze the questions rather than reading and understanding the texts. Meanwhile, in academic life, students are asked to have an understanding about the content of texts, not merely answer a series of questions. In short, classes become test oriented where both the content and activities focus on the test (Teemant, 2010). Au (2009) supported this idea that pedagogy in the classroom has become test-oriented, where the type of the test really affects the activities in the classroom. Therefore, Kohn (2000) emphasized that high-stake test scores do not reflect the comprehension and understanding of students, but rather their skill in taking tests.

Furthermore, besides the consequences of test-based curricula and test-based pedagogy, which only focus on the test, tests also narrow the knowledge (Shohamy, 2001a). Au (2009) argued, "high-stakes testing narrows the instructional curriculum and aligns it to the tests" (p. 45). People only focus on what will appear on the tests, what problems need to be solved in the tests and create strategies to cope with those problems. In other words, tests narrow the taught topic. For example, in the TOEFL and the IELTS tests, the reading sections only focus on several genres. In the TOEFL test, the genres are exposition, argumentation, and historical narrative, and in the IELTS test the genres are descriptive, discursive, and factual. The students will only be taught these genres because

these genres appear in the test. The other genres are considered not important. To be clear, if there is a high probability that some materials will not appear in the test, the teacher will not teach them to the students, even though that material may be educationally important (Shohamy, 2001a). Therefore, if the students are asked to do the same activities outside the test-taking context then they will not be able to understand it. It is reported based on Amrein and Berliner (2002) research that:

High school teachers report that although practice tests and classroom drills have raised the rate of passing for the reading section of the TAAS at their school, many of their students are unable to use those same skills for actual reading. These students are passing the TAAS (Texas' Assessment of Academic Skills) reading section by being able to select among answers given. But they are not able to read assignments, to make meaning of literature, to complete reading assignments outside of class, or to connect reading assignments to other parts of the course such as discussion and writing. (p. 26)

In addition, high- stakes tests have power to redefine knowledge. Tests are able to make stakeholders (e.g., test takers and parents) believe that the tested knowledge is the most important knowledge to be mastered (Shohamy, 2001a). Furthermore, effect is related to what should be taught and what should not be taught. The policy-making drives idea about what knowledge should be taught and this knowledge becomes the focus of high-stakes test.

Based on the explanation above, it can be concluded that high-stakes tests have become 3 major controls: content control, formal control and pedagogic control (Au, 2007). Tests have become the curriculum, and affected teaching practices in classroom.

Shohamy (2001b) argued, “the content and the knowledge contained in tests represent the interests of those in power who are often interested in perpetuating their domination and in excluding unwanted group” (p. 375). The powerful people make policies which affects the test. They create the test based on their beliefs about what is important, and it is done in order to preserve their authority. Therefore, it can create disadvantages to those people who are not in power.

**Societal effects of high-stakes tests.** Societal effects “are concerned with the effects of tests on aspects such as gate-keeping, ideology, ethicality, morality, and fairness” (Shohamy, 2001a, p. 46). Shohamy (2001a) argued that tests have “economic value that estimates the worth of people” (p. 123). In other words, this means that tests can define people based on their achievement in the tests. For example, the economic value for the test-takers is to get into prestigious university. Tests also possess power to have people invest their money and time in tests preparation. For the TOEFL and the IELTS tests preparations, test takers have to pay a lot of money. It is not only related to the cost that they have to pay to the institution, the cost also comes from the books that they have to buy to “help” them to get the targeted score. Elaborated below are the examples of the societal effects of tests:

1. Tests and identity. As one of the consequences of tests is grouping people based on their test score (Shohamy, 2001a), tests have marked their position as important matters not only in educational level but also in society and identity of the test-takers. Some studies have found that people have different investments in language learning and many factors affecting those investments: gender (Gordon, 2004; Hruska, 2004; MacPherson, 2005;

Morita, 2004;), race (Kubota & Lin, 2006), and community practice (Gao, 2008; King & Ganuza, 2005; Morita, 2004; Rajadurai, 2010).

2. Tests and ethicality. The imposed power of tests has led stakeholders such as the test takers, parents, and teachers to believe that tests, IELTS and TOEFL tests, are everything. Templer (2004) argued, “the entry points into privileged career paths at the portal to university study in English-speaking countries—and within numerous job pyramids inside the corporate and professional world—are controlled by quantitatively measured performance on several international language proficiency tests” (p. 1). This means that the IELTS and the TOEFL tests have marked their position as the only instrument to enter such privileged career and university. This appealing construction of tests has made people invest their time and money for the tests including preparation and test taking.
3. Tests and fairness. Fairness is also another issue in the high-stakes tests. Xi (2010) argued about tests fairness focusing on “the comparison of testing practices and testing outcomes across different groups” (p. 154). In other words, she argued about the notion of equality in tests. Language tests are supposed to be designed equally relevant to each group. This means that tests should be designed without excluding or perpetuating power of any groups in society. Furthermore, some studies reveal that fairness is a prominent issue in language testing. In her study, Puspawati (2012) found that the students perceived the topic in the TOEFL test as unfair. They claimed that the topic mostly covers about American culture and scientific disciplines. Coming from

different backgrounds of knowledge and culture, they found the topic in the test unfamiliar. Wise et al. (2004) also found similar result. Based on their evaluation of the California High School Exit Examination (CAHSEE), they found out that Hispanic and African American students' passing percentage was lower than the American students were. These findings imply that test takers who are familiar with the knowledge and culture of the mainstream group will get advantage from tests.

Based on the explanation above, it can be derived that test takers try their best on tests in order to avoid punishments and to get rewards both in the educational and society level. In the educational level, prestigious university is seen as the reward for the successful test takers. In society level, by getting into prestigious university, these test takers have a special position in their communities. These effects of high-stakes tests have created many consequences: narrowing the knowledge, redefine knowledge, and legitimized the power of bureaucrats and other elite groups.

Having discussed the effects of the IELTS and the TOEFL as high-stakes tests, it is really important to know how the stakeholders', especially the test takers, perceptions of tests because they are the people affected by the test. However, the studies about the IELTS and the TOEFL were more focused on their validity and reliability. There are many studies focusing on the IELTS test's validity (Elder & Wigglesworth, 2003; Mickan, Slater, & Gibson, 2000; Sadeghi, 2010) and reliability (Merrylees, 1999; Sulaiman, 2012). As well, the TOEFL test studies focus on this topic: Validity (Attali, 2007; Bridgeman, Powers, Elizabeth, & Mollaun, 2011; Sawaki & Nissan, 2009; Weigle, 2010) and reliability (Hsu, 2012; Kokhan, 2012; Zhang, 2010).



Besides being studied solely, there are some studies comparing these two tests. Zahedi and Shamsaee (2012) focused on comparing the construct validity of IELTS and TOEFL iBT speaking module. ETS in 2010 also has conducted a study comparing IELTS and TOEFL iBT scores. Hill, Storch, and Lynch (1999) compared these two tests as academic success predictors; however, there is little research focusing on students' perceptions of these two tests. Meanwhile, as Shohamy (2001a) argued that listening to the test takers' voices about tests will lead us into a better understanding about tests especially for high-stakes tests. Reflecting on this argument, this study is the part of this process. As this current study focuses on students' perceptions towards the IELTS and the TOEFL test, the result of this study will give insight about how students perceive these high-stake tests. The next section provides information some studies related to students' perceptions of high-stakes tests.

### **Studies on Students' Perceptions of High-Stakes Tests**

Murray et al. (2012) pointed out test-takers' attitudes toward test may have significant implications to test validity. Dooley (2008) argued the importance to determine test validity related to the consequences of the test which is based on Messick's (1998) idea that validity is not just referring to "the accuracy of score inferences but also to evaluation of the appropriateness, meaningfulness, and usefulness of score inferences, which involves judgments not only of truth but of worth" (p. 41). Huhta, Kalaja, and Pitkanen-Huhta (2006) also asserted that social consequences and value of tests are two aspects of validity of tests. Listening to test takers' voices and perceptions of tests are important, because they are the people who have experienced the social and educational consequences of tests. Therefore, including students' perceptions of tests is one of the

aspects in construct validity (He & Shi, 2008). Shohamy (2001a) argued tests have its powerful effects not only in students' educational lives, but also in their social lives. Furthermore, she also emphasized the importance of knowing students' perceptions towards tests as it can lead into information about the importance, the use, and the meaning of the tests to the students' lives. Based on the literature review, there are some points related to the students' perceptions of tests.

### **Impacts of Tests**

Some studies found that tests affect students' emotion (Gan, Humpreys, & Hamp-Lyons, 2004; Huhta, Kalaja, & Pitnkanen-Huhta, 2006; Li, Zhong, & Suen, 2012; Murray, Riazi, & Cross, 2012; Triplett & Barksdale, 2005; Triplett, Barksdale, & Leftwich, 2003) and learning (Cheng, Andrews, & Yu, 2011; He & Shi, 2008; Huhta, Kalaja, & Pitnkanen-Huhta, 2006; Li, Zhong, & Suen, 2012; Puspawati, 2012; Xie & Andrews, 2013). These two impacts affect the way test takers in the study perceived tests and the way they dealt with the test.

**Impacts of tests toward students' emotion.** Triplett, Barksdale, and Leftwich (2003) found that students felt worried every time they talked about test. They worried before the test, while they were taking the test, and after the test. Another point found in the research was student's fear of being a failed student was in their mind when they were working on the test. One of the students even felt uncomfortable that he would make his parents worry. Even though he felt nervous and scared every day, he did not tell his parents because he worried that they would get nervous as well. The effect reported in this research is not only affecting students emotionally but also physically.

“it was hard and it made you nervous. And it made you tired because you got to get up early to study” (p. 19)

Triplett and Barksdale (2005) also did another research regarding students’ perceptions of test. The findings showed that the test had negative impact toward their emotions. They expressed their feeling and experiences negatively. Other students also expressed anger due to the possible consequences of failure. It shows that tests gave negative impact towards students’ emotions not only when they took the test, but they were aware about the consequences of the test. A student said, “I really don’t need dumb test telling if I get to go to another grade” (p. 251). It shows how powerful testing is, hence it can create hatred toward the test.

Triplett, Barksdale, and Leftwich (2003)’s finding is supported by Huhta, Kalaja, and Pitkanen-Huhta (2006). In their study, they found that test gave negative impact to test-takers’ feeling. Students felt stress in the preparation of the test.

Matti: “I’m that much stressed about English that (..) I’d like to get laudatur [the highest grade] because (..) well (..) that’s what (..) I’m doing here trying to improve my grade (..)” (p. 335)

The participant showed his stressfulness even before taking the test. This study shows that tests have marked their position as important factor, because even though the test-takers have not taken the test, they experienced the stress.

Another qualitative study conducted by Puspawati (2012) focusing on student-teachers’ perceptions towards TOELF test showed that the test created negative feeling towards their emotion. The students demonstrated that they experienced stressfulness, anxiety, and nervousness when they were taking the test.

“but you see like ugh...you feel very stressful because you always thinking about ‘I don’t have enough time, I don’t have enough time’ and you always...hmm....especially TOEFL for me is umm....the only way I can go to study abroad. I always thinking about, ‘ok, I made a mistake here, so what should I do for the next step. So, all the memory about TOEFL is kind of stressful” (p. 45)

One of the reasons that the participants in the study claimed as the factor of their negative feeling was related to the position of the TOEFL test as high stake test. Due to this factor, they were under pressure when they were taking the test.

A recent quantitative study done by Murray, Riazi, and Cross (2012) focused on test takers’ attitudes toward the test and interrelationship between test takers’ attitudes and their demographic found that tests gave negative impact to test-takers’ feeling. The result of the study shows that test-takers had negative responses towards the tests. 69.5% of participants said that they experienced anxiety because of PEAT (Professional English Assessment for Teachers), 59% of participants felt depressed. Furthermore, 54% of participants claimed that sometimes they wanted to give up, 42% of the test takers were angry and 39% of them were embarrassed. This study points out that tests can give negative feeling to the test-takers regardless their different backgrounds and origins.

Furthermore, Doe et al. (2011) investigated L1 and L2 test takers’ perceptions toward language test. Regardless of their native language, both groups responded the same about taking the test. They considered the test as unfair, bad, waste of time, and pointless because the test did not reflect their true ability in English.

However, these studies also found students' positive emotional responses toward the test indicating that each test taker had different attitudes towards the test especially if it was related to the result of the test. Murray, Riazi, and Cross (2012) found that the test-takers obtaining high score on PEAT had positive responses of the test. It aligns with a study done by Gan, Humpreys, and Hamp-Lyons (2004) showing that students who obtained satisfactory score of English proficiency test tended to have positive attitude towards the test, which was different from the students who obtain low score. The score itself gave an impact into students' confidence about their English. Students felt more positive toward their English proficiency when they got satisfactory score on the test.

**Impacts of tests towards students' learning.** Similar to the impacts of tests towards students' emotion in which students felt nervous before, during, and after taking tests, tests give impact students' learning in the same way. Xie and Andrews (2012) found that the students analyze the test to be able to deal with tests. The students expressed that by "knowing what is necessary could give them a sense of direction and control over the upcoming task" (p. 63). The students develop their own ways to deal with test. Puspawati (2012) found the similar result in her study investigating student-teachers' perceptions of TOEFL test. The result revealed that students tended to be more focus on dealing with the test instead of improving their English proficiency.

Cheng, Andrews, and Yu (2011) did another research related to the impact of the test on learning. The research intended to explore students and parents' perceptions of the impact of SBA (School Based Assessment). The result of the study showed that higher competence students responded positively to non-SBA compared to lower competence students. The lower competence students responded more positively to SBA. Regardless

what test they focused on, the test had an impact in their study. They studied and practiced not only while they were having the class, but also when they finished the class.

He and Shi (2008) also investigated how Chinese students perceived and compared standardized writing test of TWE (Test of Written English) and the essay test in LPI (English Language Proficiency Index) – a required English test to get in to post-secondary institutions in Canada-. Some students passed the TWE test claimed that they could get high score because they memorized sentences and structures. The impact of the test in this case is that the students developed their understandings about strategy in taking the test by memorizing pattern, sentences, and structures. Therefore, when they studied for the test they memorized all the information they obtained whether from their teacher or the Internet.

Huhta, Kalaja, and Pitkanen-Huhta (2006) also found that test could affect students' learning either positively or negatively. Some students tended to study more until the last minute before the test. They read all books related to English and borrowed books from their teachers as well in order to prepare themselves. However, some students already gave up before the test because they thought there was not enough time to study.

Suvi: "I haven't studied at all (.) and probably I won't because (...) there's not much time anyway (...) or I don't know I've got that grammar book there in front of me but I haven't studied it yet (..) let's see then tomorrow how it goes" (p. 335)

Doe et al. (2011) interviewed and observed three test takers during their preparation time, before the test, and after taking the test. The result implied that each test taker has their strategies based on the context of the test. In the test preparation, the

participants showed that they were trying to figure out what kind of strategies they would use on the test. The strategies changed over time as the test time became closer. It shows that the test takers had pointed out the importance of strategies in taking the test.

The recent research done by Li, Zhong, and Suen (2012) found that CET (College English test) had a big impact on the content of the material that they learned. The research done by Teemant (2010) also proved the same issue that tests have driven classroom activities which become more test oriented both in content and in learning.

From the explanations above, it can be concluded that tests tend to give negative impact emotionally to the students. Tests can make students feel uncomfortable even hatred because they have such powerful effects towards their future lives. Tests also give negative impact to the students physically. They had to get up early even though they could not sleep at night. In relation to the impact of test towards learning, tests do not only give negative but also positive impacts. Since tests have powerful effects to students' educational life, students tend to study more and created their own way of learning. However, their learning becomes more about how to get a good score rather than improve their knowledge. In addition, it gives negative impacts because the students felt like they did not have enough time to prepare, therefore they chose not to study. These impacts of tests towards students' emotion, body, and learning illustrate the power of tests in their lives.

### **Problems in the Test**

Besides giving many impacts towards students' emotion and learning, tests also pose some problems. He and Shi (2008) found that the writing topics in the LPI (English Language Proficiency Index) affected students' performance on the test. LPI essays

topics were related to Canadian culture. The students had problems in writing because they had no background information and had limited language vocabularies regarding the topic. The writing prompt gave advantages to the local students where they are familiar with the topic. Therefore, the Chinese students argued that the LPI test did not reflect their writing ability. He and Shi (2012) asserted their finding by doing another research about the relation between topic and ESL writing. They found that the students complained about the topic prompt that they were not familiar with and about the lack of vocabularies in the topic.

Bing: “Most of the students here [in this class] came here [to Canada] not long ago and haven’t settled down yet. It seems they only know the names of democracy party and conservative [party]. For other parties, they have no ideas. They are no familiar with the politics here yet. They don’t know the government system either. So, it is more difficult for us to write on this topic” (p. 458)

Another problem, which can be highlighted from the research done by He and Shi (2008), is related to the validity of the writing test result. Most of the students argued that they got a high score in the TWE test because they memorized patterns and materials from the test. They searched the patterns and materials online or found it in the bookstore, memorized it, and then applied it on the test. Therefore, there was a doubt that TWE is really measuring students’ writing ability. It is more about having good memory to pass the test rather than real proficiency in English. Thus, the test takers’ score from this test does not represent their real writing proficiency because this test does not measure their writing ability, but their memorization ability of patterns and materials.



Since this research focused on LPI and TWE (He & Shi, 2008), it allowed the researchers to compare the tests. All the participants in this study agreed that LPI was more difficult than TWE. The participants realized that the memorization strategy that they applied in TWE test did not work in the LPI test. However, the writing prompt in the test also made the students find difficulties because they were not familiar with the Canadian culture. By listening to students' perceptions who have taken two kinds of tests, information about each test can be obtained which enable the researchers to compare them and find out the sense of writing tests between themselves. One of the implications generated from the research is that there is a need to develop an appropriate writing prompt. The students claimed that the topic in TWE was too general, so they just could apply what they have remembered; meanwhile the topic in LPI was too specific. These contrasting ideas imply that there is a need to develop culturally fair writing prompt and general experiences which will not create advantage or disadvantage to certain group.

Another problem found in the test is related to the time allocation (Doe et al., 2011; Puspawati 2012; Triplett & Barksdale, 2005). They found out that students were concerned about the time allocation in the test. Some students argued that tests were too long; while the rest said they were too short (Triplett & Barksdale, 2005). Puspawati (2012) also supported this finding. The participants argued that the time given in the test was too short. It made them feel pressured and stressful.

“you feel very stressful because you always thinking about ‘I don’t have enough time, don’t have enough time [ to finish the test]’” (p. 53)

Another problem in the test is regarding the use of technology (Noubandegani, 2012; Sawaki, 2001). Sawaki (2001) found that the test-takers complained about the

intense exposure to the computer screen causing them to have eye fatigue while they were doing the test. The eye fatigue they got affected their concentration; therefore, they could not perform well in the test. Besides fatigue problems, Noubandegani (2012) also pointed out that most of the participants agreed that test takers' computer familiarity could affect their performances positively. This finding questions the intention of tests whether to assess students' English proficiency or their computer skills.

However, a study conducted by Rasti (2009) showed that test takers perceived IELTS test positively. They claimed that the place and the time given supported them to perform in the IELTS test. In addition, they reported that the test measure their English proficiency "quite accurately" (p. 145). They noted that test format familiarity and test rubric helped them to perform in the test.

The studies above illustrates that tests have some issues. The students argued that taking tests was more about having good strategies rather than having high proficiency. It raises question about the validity of the test whether it is really measuring students' proficiency or strategy in taking the tests (Puspawati, 2012). Also, the familiarity of test format and rubric affected their performance in the test. Furthermore, topics given in the tests can be advantage or disadvantage for certain test takers. It raises the question about the fairness issue in the test. Another problem that tests have is about time allocation and the use of technology in the test.

### **Factors Affecting Students' Performance on the Test**

Some studies also found some factors affecting students' performance on the test and test preparation. One of the influential factors affected students' performances on the test is the strategies that they use in the test (Klinger & Luce-Kapler, 2007; Puspawati,

2012). Puspawati (2012) found that test taking strategy was able to increase students' score. She found that preparation done by participants prior to the test was more about improving their strategy rather than improving English ability. Her finding aligns with Klinger and Luce-Kapler's (2007) finding reporting that students with strategies were able to pass the test. McNeil (2011) also found that the students applying strategies in the reading test had higher score in the reading comprehension score. McNeil (2011) proved that using strategies in taking test had significant impact towards students' score.

Another factor influencing students' performances on the test is students' personal negative experiences (Huhta, Kalaja, & Pitkanen-Huhta, 2006; Puspawati, 2012). The main sources of negative attitude towards test were test takers personal experience and feeling, and impact from other people (Murray, Riazi, & Cross, 2012). Huhta, Kalaja, and Pitkanen-Huhta (2006) found that one of the participants shared that he got a bad score in the last test, and he was not happy about it. Therefore, when he would take the test for the second time he felt nervous and stressed due to his failure in the past. This shows that failure on the past test also contributed to students' negative feeling. They wanted to get higher score because they did not want to have the same bad post-experience after taking the test; therefore, they felt nervous and stressed to take the second test. Also, the fact that their future would be determined by the test score increased their anxiety (Puspawati, 2012). Negative feelings and discomfort have been proven as factors that can affect students' performance on the test.

Students' familiarity with topic is another factor affecting students' performance on the test (He & Shi, 2008, 2012; Sasaki, 2000). He and Shi (2008, 2012) studies found that students' familiarity would affect their performance on the test significantly. The

students found difficulties in terms of vocabulary, cultural, and context. They also experienced lack of confidence in commenting on authorities. Their lack of vocabularies made them unable to express their ideas and thoughts. Related to the cultural contexts of the writing, the students were afraid that the words they used in the essay did not represent what they thought due to cultural barriers. Furthermore, students had lack of confidence in commenting on authorities because they did not know about Canadian politics. Therefore, they had no confidence in writing about that. Sasaki (2000) also revealed the same thing. These findings strengthen Fulcher and Reiter (2003), Alptekin (2006) and Lee (2007) findings about the significant influence of topic familiarity towards students' performance on test.

Place and environment also contributed to the students' performance on the tests. Puspawati (2012) found that the students felt anxious and could not focus on test because of the air conditioner and the noise from the key board and other test takers – on the speaking section. Since the context, in this case the place of the test, was not comfortable, the test takers could not concentrate and focus on the tests. They could hear what the other test takers said when they had their speaking section.

Computer familiarity (Noubandegani, 2012; Puspawati, 2012) becomes one of the factors that affect students' performance on tests. Students who were not familiar with computer and technology faced anxiety which affected their performances. Noubandegani (2012) asserted that computer familiarity affected students' test performance positively meaning that when the test takers were familiar with computer, it would help them in doing the test and vice versa.

Based on these studies on students' perceptions, tests impose impacts to the students and impose some problems. Tests affect students' learning and emotion. Regarding students' emotion, tests give negative feelings to the students. The students expressed their negative feelings prior to the test, on the test, and after the test. Also, tests gave negative impact physically, because they had to wake up early in the morning to study and could not sleep at night because they kept thinking about the test. In terms of learning, tests drive the learning to become more test-oriented. They learn everything that based on tests demand. Tests also pose some problems related to time allocation, topics covered in the test, and whether the test really measures students' English proficiency or test taking skills. These studies also identify several factors influencing students' performance on the test, such as: strategies, past experience, place and environment, computer familiarity and topic familiarity.

After the discussion how these two tests have marked their position in society as high-stakes tests which pose many consequences and impact not only in education but also in society. Students' perceptions also pointed many impacts and problems of tests; therefore there is a need to find an alternative perspective of language assessment. The section below will introduce about democratic assessment perspective and practices.

### **Democratic Assessment**

This study aims to investigate students' perceptions towards IELTS and TOEFL tests; therefore, there is a need to highlight about democratic assessment. This is due to the importance and the position of these two tests towards the test takers' lives. The notion of democratic assessment provides an insight about how to treat and create tests

that can be acceptable in a sense that tests' contents and uses do not discriminate any groups in society (Shohamy, 2001b).

As the discussion before, the uses of tests as instrument of power by those in authority have lead into the misuses of the tests. Tests are used as disciplinary tools, meaning test takers is pushed to change their behavior in order to comply the tests demand, so that they can improve their score (Shohamy 2001a, 2001b, 2004).

Furthermore, it is also used to redefine knowledge and impose the knowledge of those in authority. Such misuses of the tests narrow the knowledge, and can create contradiction with the existing knowledge. These misuses have violated the principles of democratic practices. Therefore, there is a need to have democratic assessment practices aiming at “monitoring and limiting the uses of tests, especially those that have the potential to exclude and discriminate against groups and individuals” (Shohamy, 2001b, p. 373).

The principles of democratic assessment are focused on cooperating stakeholders to monitor and limit the powerful use of tests. Based on the idea of the principles, five democratic assessment practices are required (Shohamy, 2004):

1. Incorporating diverse groups' knowledge on tests

This practice demands to involve different groups in the process of constructing tests. The testing policy is usually based on homogeneous knowledge coming from dominant group which ignore knowledge from other groups. Therefore, there is a need to collaborate knowledge from different group in constructing, developing, and designing tests. The need to consider knowledge from each group as equal and legitimate also becomes the essence of this practice. Each group needs to be recognized for the differences.

2. Cooperating test takers in conducting tests

Tests are usually top-bottom system where the test takers (bottom) only receive everything the tester give. The democratic assessment practice for this notion is a sharing power between the tester and the test takers. This new model of assessment asks these two stakeholders to collaborate then evaluate and experience the process of test together. This means that the tester and the test takers are in the same position in order to have democratic act of testing. For broader practices, it is not only between testers and test takers but also with other stakeholders, so that it can provide knowledge of the tested.

3. Applying CLT (Critical Language Testing) to examine the uses of tests

Critical Language Testing (CLT) refers to “the need to question the use of tests as tools of power, to examine their consequences in education and society, and eventually to monitor, control, and limit their powerful role” (p. 84). CLT addresses some issues related to the language testing which are: first, it examines the aim of the tests. It is related to the idea that tests are tools to impose knowledge from certain group in authority. Therefore, CLT points out a question whether tests assess the knowledge of diverse groups in society or preserve and maintain knowledge of those in power. Second, CLT asks for collaboration among stakeholders because the tester’s knowledge is incomplete. This act is needed in order to have more valid and accurate descriptions of knowledge. Third, CLT questions the use of tests as a single tool to assess knowledge, and asks for the use of other procedures which can give more valid interpretations of individual’s or group’s knowledge. Fourth, CLT asks the stakeholders to critically

look at the role of tests in society to raise the testers' awareness about the power of the tests. Fifth, CLT asks many questions regarding the test and the stakeholders, so that the misuse of the tests can be predicted. From these addressed issues, it can be concluded that CLT can defy the use of the test as a powerful instrument.

#### 4. Protecting test takers' rights

Test takers are often seen as powerless objects in testing. This notion is rooted in the practice where the test takers have to meet the demands of the tests. Such practices make the test takers seem to have no rights. However, by using democratic assessment principles, the test takers' rights have to be protected because they are the true victims of tests (Shohamy 2001a, 2001b). There are some ways to apply this democratic principle such as: questioning the uses of the tests, having the right not to be tested, suing litigation, protecting privacy and confidentiality, questioning test decisions: consequences and fairness, having alternative forms of assessment, and sharing discourse.

#### 5. Addressing test makers' and testers' responsibilities

This principle demands the testers to be responsible for the consequences of developing powerful instruments. They have to be aware about the consequences of the instrument that they have developed. Language testers have six responsibilities (Shohamy, 2001b):

- a. "Ethical responsibility" (p. 381)



- b. “Responsibility for making others aware” (p. 382): the tester has to provide the users and public about the “intentions, effects and consequences” (p. 382) of the test based on the dimensions and the facets use of tests.
- c. “Responsibility for all the language test consequences” (p. 382): the testers need to accept responsibility to test consequences.
- d. “Responsibility for imposing sanctions”: the testers have to be responsible for the product (test) and “therefore obliged to forbid its use when it is found to be defective” (p. 382).
- e. “Shared responsibility where the responsibilities for good conduct are in the hands of all those who are involved in the testing process (shared authority and shared responsibility)” (p. 382): the testers have to admit their knowledge is limited; therefore they need to collaborate and work together with other groups. This responsibility also means to be critical to the test and its use, to gather data about the effect and the consequences of tests, to against the misuse of tests, and to protect both testers and test takers.
- f. “Responsibility to abide by Codes of Practice (ILTA, 2000)”: it covers “professional’s role in society, integrity, conflicts of interest, diligence and due care, confidentiality, and communication with clients and the public” (p. 383).

### **Conclusion**

Based on the elaboration in this chapter, it is clear that listening to test takers’ voices is important. Test takers were able to point out issues that might not be noticed from the view of other stakeholders. Test takers highlighted the detrimental effects of the

tests such as the power of tests that can change their behavior to meet the tests demands. Furthermore, they also highlighted the societal and educational effects of tests. In a sense, test takers voices are essential in providing information about tests, their meanings, and their consequences. Therefore, I conduct this study in order to give further insight about test takers' perceptions of high- stakes tests.

## CHAPTER THREE

### METHODOLOGY

As elaborated in the previous chapters, this research is an endeavor to highlight the importance of listening to test takers' voices as a way to understand language tests. The aim of this study is to explicate test takers' different perceptions of the IELTS and the TOEFL tests. To meet t purpose, each participant in this study was interviewed through Skype for approximately 45 minutes.

This chapter describes the research design, the participants, the data collection method and the procedures applied in this study. Furthermore, it provides explanation on how the data was analyzed. At the end, the summary of the methodology chapter is presented.

#### **Research Design**

The purpose of the study is to explore and investigate the perceptions of six graduate students who have taken both IELTS and TOEFL tests, meaning the participants described and shared their experiences and thoughts about the tests. Based on this objective, the research approach chosen for this study was qualitative approach. Lichtman (2009) argued that qualitative research is a way to have knowledge about a specific issue by gathering, interpreting and organizing information from the participants. Reflecting on this definition, this means qualitative research provides rich information about individual(s)' experiences of a specific issue and it is not simply listening to their opinion and thought, but understanding their experiences (Scott & Garner, 2013). Merriam (2009) mentioned that "the nature of qualitative research: the focus is on process, understanding, and meaning; the researcher is the primary instrument of data collection of analysis; the

process is inductive; the product is richly descriptive” (p. 14). Based on these reasons, I justified the use of qualitative approach for my study.

### **Participants**

Based on to the purpose of the study in which I intended to explore and investigate students’ perceptions of the IELTS and the TOEFL testing experiences, the participants of this study have to have taken both of the tests. The participants were six international graduate students who have or pursue a degree in TESOL (Teaching English to Speakers of Other Languages)/TEFL (Teaching English as a Foreign Language) /TESL (Teaching English as a Second Language)/Applied Linguistics, and are second language learners of English. These criteria were selected because I sought to obtain rich information regarding the topic of the study. By having or pursuing a degree in the field of English, the participants were assumed to have deeper understanding about language testing which would be very important for the study. In terms of gender issue, there was no restriction of gender in this study. The age of the participants was 18 – 45 years old because this age range includes both undergraduate and graduate students’ age group. The demographic information of the participants is presented in Table 3.

Regarding the number of the participants, I decided to have six participants in this study. This decision was due to my intention to have rich information regarding the topic. Perry (2005) stated that information rich-paradigm emphasizes on the quality of the information, not the quantity of the sample. Thus, the number of participants in a qualitative research is not considered as a big issue, since the purpose of this research was to listen to individual(s)’ experiences of a particular issue, and was not going to be generalized (Lambert, 2012; Lichtman, 2009; Scott & Garner, 2013).

Table 3

*Participants' Demographic Information*

Participants (pseudonym)	Nationality	Gender	Academic Status	Reasons for Taking Test	
				IELTS	TOEFL
Aci	Indonesian	Female	Graduate Student (Master)	No requirements (as an experience)	Not identified
Shinta	Saudi	Female	Graduate Student (Master)	School requirement	Job requirement
Imam	Indonesian	Male	Graduate Student (Master)	No requirements (as an experience)	Not identified
Norah	Saudi	Female	Graduate Student (Master)	School requirement	School requirement
Andy	Saudi	Male	Graduate Student (Master)	Job requirement	School requirement
Erica	Indonesian	Female	Graduate Student (Master)	Not identified	Not identified

**Data Collection Method****Semi-structured Interview**

In order to collect data, I applied interview methodology. According to Lambert (2012) interview is an efficient way to elicit in-depth information of someone's experience, perspectives, and opinions. Grounded in this definition, interview is not merely: the interviewer asked questions and the interviewee answered them. In this respect, the researcher had to actively engage in the conversation focused on the study (Merriam, 2009; Lambert, 2012). The interview approach enabled me to obtain participants' thoughts, opinions, and ideas toward a specific topic.

According to Perry (2005) there are three types of interview: highly structured interview, semi-structured interview, and open structured interview. This research employed a semi-structured interview. Semi-structured interview is a type of interview where the researcher is allowed to ask additional question and discuss additional topics related to the issue discussed (Lambert, 2012). This type allowed me to “pursue new lines of enquiry in response to what interviewees say” (p. 104). Furthermore, Morse (2012) argued that semi-structured interview is applied when the researcher has sufficient knowledge related to the topic, but unable to anticipate the response of the participants. In this regard, I have reasonable amount of knowledge about students’ perceptions of language tests and issues in language assessment. The semi-structured interview questions were designed in order to meet the objectives of the study. I had interview questions guidelines (see Appendix E) assuring the flow of coverage topic and chance for the participants to explain their answers in detail (Scott & Garner, 2013). Beitin (2012) said, “interview question should help guide an interviewer but not so rigidly that an interviewee is not able to shift footings and perspectives. Interviewers should encourage participants to discuss how they constructed their narrative” (p. 251).

### **Individual Interview**

The kind of interview used in this research was an individual interview. Individual interview is the most common strategy used by researchers in order to collect data in qualitative research (Beitin, 2012). Lichtman (2009) argued that “individual interview is a process where the participants share what they know, have learned and can add dimensions to our understanding of the situation by using their own words, in his voice, with his languages and narrative” (p. 143). The reasons I applied this type of

interview is due to, first, individual interview would give the participants a comfort feeling when they described and expressed their thought to the researcher. Triplet, Leftwitch, and Barksdale (2003) acknowledged that individual interview give feeling of comfort to the participants. Second, arranging individual interview would be more manageable than arranging group interview (Beitin, 2012) relating to time constraints. Individual(s) have their own schedule, and matching their schedule would be difficult. Third, individual interview was “less susceptible to members holding back or altering information in the presence of another member for fear the information would negatively affect the relationship” (Beitin, 2012, p. 244). Another reason behind my decision of conducting individual interview was related to the use of Skype in the interview.

The participants were interviewed by using Skype – an internet telephone service which allows video chatting. There were several reasons why I decided to use Skype (video conference) as the medium of interview. First, the use of Skype was due to the residence of the participants. The participants and the researcher were geographically dispersed. Therefore, the use of online media is possible (Lichtman, 2009). Second, I did not assigned participants to do any tasks during the interview. Furthermore, by using video conference, I would not miss non-verbal cues during the interview, which could be one way for me to know the state of participant’s comfort. Finally, using Skype would be convenient for the participants because they did not have to come to the assigned place. At this point, I also could assure the participants’ identity confidentiality because they would not be seen coming to certain places to meet me. In terms of the place of the interview, the participants decided the location they wanted to have a Skype interview. However, at the same time, I also asked them to have a comfortable and a quiet place.

## **Potential Drawbacks and Countermeasures**

As the principal investigator, I was aware that the participants might feel uncomfortable and embarrassed to talk about their personal experience of taking the IELTS and TOEFL tests. Another possible risk came from the fact that the participants might feel uncomfortable having approximately forty five minutes interview on Skype. Therefore, to handle the risk of comfortableness and embarrassment of the interview questions, the participants had been informed both in the consent form and before the interview that they had power not to answer certain questions. In addition, after the interview, I, as the researcher, gave my participants positive feedback telling them that their information was really helpful and valued. In order to handle risk related to discomfort with the length of the interview, the participants were given freedom to schedule the interview time. Also, if as, the interviewer, I felt that there was extensive discomfort to the participant; I would ask them if they wish to stop the interview.

## **Data Collection Procedure**

### **Recruiting Participants**

This study used purposive and snowball sampling method in order to have six participants. These two methods were applied due to the need for this research. The purposeful sampling method would give the researcher rich information about the topic because the samples were indicated to “supply most information possible regarding the research question” (Perry, 2005, p. 57). The other method was snowball sampling method. Johnston and Sabin (2010) defined snowball sampling as “chain referral sampling method that relies on referrals from initial subjects to generate additional subjects” (p. 38). In other words, the researcher obtained information about other



participants from the participants. This method was chosen because the population of this study has specific characteristics (Cohen & Arieli, 2011), and might only be recognized by certain people.

I started finding the participants by contacting colleagues and acquaintances who either fit the inclusion criteria or who might know someone fitting the criteria. After finding the participants, I sent them an email (see Appendix D) to invite them to be a participant in my study. Following email contact and agreement to meet by Skype, I asked each participant to accept my friend request on Skype. Afterwards, I discussed Skype meeting time with each participant either via email or Skype to explain the project and the consent form (see Appendix C) of this study. Only participants who agreed to participate after the explanation were asked to sign the consent form in front of me and to show me the signature. Following agreement and the signature of the consent form, they were asked to send the form to me by post.

### **Interview**

The interview was conducted at locations decided by the participants. The interview time was based on participants' convenience. The interview questions were constructed to meet the purpose of the research. The interview questions consisted of three segments where, first, the researcher asked about participants' perceptions of the TOEFL test. Second, the questions were about perceptions of the IELTS test. The last segment was related to comparison between IELTS and TOEFL tests. The questions for segment one and two were similar, only the subject of the questions was changed. At the end of the interview, I assured my participants that their identity would be kept strictly

confidential. In addition, I would offer whether or not they wanted to have the copy of the analysis.

### **Data Analysis**

According to Merriam (2009) “data analysis is a complex process that involves moving back and forth between concrete bits of data and abstract concept, between inductive and deductive reasoning, between description and interpretation” (p. 176). Due to its importance in the discussion part, the data was analyzed very carefully by transcribing, giving themes, and coding them. The analysis used in this research is thematic analysis (Gomm, 2004). The thematic analysis codes each conversation based on the theme. The steps in thematic analysis are (Gomm 2004, p. 189):

1. Deciding what themes will be
2. Deciding what will count as evidence of a theme
3. Coding a transcript to indicate this passage is an example of this theme, and that passage is an example of another theme;
4. From then on the analysis is in term of which kinds of people said what, which relates to a particular theme, and how saying something with regard to one theme relates to saying something with regard to another theme.

### **Transcription**

Transcribing process could be counted as the first step of analysis because it allowed me to become familiar with the data (King & Horrocks, 2010). In the process of transcribing the audio material into text, I focused more on “*what* is said rather than *how* it is said” (Poland, 2002, p. 630, italics in original). However, even though my focus was on what is said, a reliable transcript was needed in order to avoid ambiguity in analyzing

the content of the transcript (Poland, 2002). I decided to combine basic transcription taken from King and Horrocks (2010, pp. 145-155) and Kumagai (2013, p. 53) in the process of transcribing my digital audio-recorded data (Table 4).

The transcription process was started by converting audio data into text. To capture all words and features (e.g., pauses, emphasis, prolongation, etc.) in the interview, I used Windows Media Player program that enabled me to control the speed of the audio recording, and to see the length of the interview time. Following this process, I reviewed the transcript over time to ensure that my transcription represented what was said by my participants.

Table 4

*Transcription and Excerpts Features*

<b>Interview feature</b>	<b>Representation</b>	<b>Notes</b>
Prolongation	colons :, the number of colons indicates the length of prolongation	
Pauses	(.)very short pause, probably no more than one-tenth of a second (...) longer pauses.	If it is more than 5 seconds, I put number to represent how long the pause is, e.g. (.5)
Emphasis	Capital Letters	
Intonation	, low rise ? high rise . low fall ! exclamatory utterances	
Audibility problem	[inaudible]	
Laughing and/or similar features	round brackets {laughing }	
Concealed Information	square brackets [info]	
Utterances omission	<omit>	

## **Coding and Categorizing**

After transcribing all of the recorded material, the process of data analysis was followed by coding the data. I read all of the data and decided on the themes that emerged from the interview and coded them. Following this process, I coded the data relating to the themes.

Following the coding process, I then categorized them. I analyzed and compared the data of the six participants and identified themes that were discussed in all interviews. There are seven themes, which are prominently – mentioned by three participants - discussed: testing experiences, perceptions of topic, perceptions of task, perceptions of time, perceptions of score, and perceptions of effects of the test. Each theme in this study has its own subtheme. Described below is the description of each theme and subthemes in this research.

The first theme is testing experiences. Testing experience addresses the external factors, not including the content of the test, which is prominent during participants' test taking experiences. There are four subthemes for this theme: environment, exposure to the test, stakes of the tests and the technology. Environment refers to the condition of test takers' surroundings when they were taking the test including the other test takers, the seat, and the sound/noise in the room. Exposure to the test relates to test takers' familiarity with the test. Also, this theme covers the test takers' preparation before taking the test. The next subtheme is the stakes of the test. This subtheme indicates the reason tests takers take the tests, which can be related to their future lives or the position of the test towards their lives. The last subtheme is technology, which covers the equipment and tools used for the tests.

The second theme is topic in the test, referring to the area of discipline or content covered in the tests. Unfamiliarity, the only subtheme, addresses the degree of test takers' knowledge, awareness or insight into a topic. The third theme is the tasks indicating the four skills being measured in the tests. There are two subthemes for this theme: practicality and English proficiency measurement. Practicality is the applicability and connection of the tasks towards real life performance. The English proficiency measurement refers to whether or not the tasks in the test measure test takers' English proficiency based on participants' perceptions.

The fourth theme is the time in the test, which has two subthemes: the length or duration of the tests and the allocated time in the tests. The length or duration of the tests is the overall time for tests. The allocated time in the tests is the time given for each task in the tests.

The sixth theme is the score of the tests, referring to the score that the participants obtained on the test. Representativeness, the only subtheme, relates to whether or not the score that test takers obtained is significant compared to their English proficiency based on their perceptions.

The last theme is the effects of the tests, which are about the post-testing effects perceived by the test takers after they took the tests. There are two subthemes for this theme: perceived English proficiency and societal effect. Perceived English proficiency subtheme is related to how the test takers identify their English proficiency after getting the score of the test. The last subtheme is societal effect, referring to ethicality, morality, and gate-keeping issues of the tests based on the test takers' perceptions.

After analyzing the data based on the theme, then participants' ideas, arguments, and claims regarding the themes were determined into three categories: positive, neutral, and negative. Each category has its criteria. The criterion of positive perceptions is: the participants showed their fondness and favor towards tests by using positive expressions, such as: "it makes you feel comfortable", "it helps you to concentrate", and "it really measure your English ability". The criterion of neutral perceptions is: the participants used plain expressions that did not show whether they perceived the tests negatively or positively. For instance, when the participants said "the topic in both tests is academic", "I do not have any problem with the time", it was considered as neutral perceptions. Finally, the criterion of negative perception is: the participants showed their dislike, objection, and disagreement towards tests by using negative expressions, such as: "it is so stressful", "it doesn't measure your English ability", and "I don't like the topic in the test".

All the processes elaborated above were applied to each test: IELTS and TOEFL. This means that each test has its own table consisting of the seven themes which emerge from participants' data. Table 5 presents the summary of the themes and subthemes. Following the process of data analysis, a peer review was conducted to establish credibility and trustworthiness of my analysis. Creswell and Miller (2000) argued, "a peer reviewer provides support, plays devil's advocate, challenges the researchers' assumptions, pushes the researchers to the next step methodologically, and asks hard questions about method and interpretations" (p. 129). The reviewers were three colleague researchers who were familiar with my research.

The frequency column in Table 5 represents the number of participants stating their perceptions toward subtheme. The analysis processes are very helpful to arrange my data. Creating the table and then comparing table of IELTS and TOEFL tests are helping me to look at the differences between these two tests regarding my participants' testing experiences.

Table 5

*Participants' Perceptions of the Tests*

Themes	Subthemes	Frequency of Participants' Perceptions (N=6)		
		Positive	Neutral	Negative
Testing Experience	Environment			
	Exposure to the test			
	Stakes of the test			
	Technology			
Perceptions of Topic	Familiarity			
Perceptions of Task	Practicality			
	English proficiency measurement			
Perceptions of Time	Length/Duration of the test			
	Allocated time in the tests			
Perceptions of tests score	Representativeness			
Effects of the tests	Perceived English proficiency			
	Societal effects			

**Summary**

This chapter described the rationale of the methodological approach and data collection method of this study. As this study aims at investigating and exploring students' perceptions of TOEFL and IELTS tests, the qualitative approach was applied

for this study. The recruitment of the participants was conducted by adapting purposive and snowball sampling method. By applying these methods, six participants participated in this qualitative research. They were interviewed individually through Skype. Each interview was digitally audio-recorded, and lasted for approximately 45 minutes. The recorded data was transcribed for the analysis. The analysis was done by using themes and coding them. The findings and discussions of the data are presented in Chapter Four.



## CHAPTER FOUR

### RESULTS

This research intends to explicate the differences between students' perceptions of IELTS and TOEFL test taking experiences. Based on the analysis, there are six sets of themes used in this study: testing experience, perceptions of topic, perceptions of tasks, perceptions of time, perceptions of score, and perceptions of effects of the test. The participants' perceptions of IELTS and TOEFL tests data were analyzed into these six themes. Therefore, this chapter has three main sections describing the analyzed data. First section presents the analyzed data of participants' perceptions of the TOEFL test followed by their examples. After that, a table is presented to summarize the TOEFL test section. Second section is the description of the participants' perceptions of IELTS test. Also, a table is presented to conclude this part. The last section of this chapter is a comparison table of participants' perceptions of IELTS and TOEFL tests followed by relative descriptions.

#### **Participants' Perceptions of the TOEFL Test**

This section describes participants' perceptions towards the TOEFL test. As the aforementioned statements, there are six themes emerged from the interview of six participants. Excerpts taken from the interview follow the elaboration of each theme. In addition, all the names used in the excerpts are pseudonyms.

#### **Testing Experience**

In the interview, all participants expressed their TOEFL testing experiences. They mentioned some factors that are prominent influencing their testing experiences: environment, exposure to the test, stakes of the test, and the technology used on the test.

**Environment.** Most of the participants complained about the environment of the TOEFL test. They reported that the environment was not appropriate for test taking because of the noise, the seat, and the space. One of the examples of the participants' complaints is:

Aci: "We like all the test takers. We do we didn't have like separate room for that, so it's like partition and them, so you could hear. You could hear the people next to you speaking." (personal communication, October 28, 2013).

The situation where they could hear other test takers speaking and typing made them have negative experience when they took the test. Besides the other test takers, the location of the test also contributed towards their negative experiences. One participant argued that since the test was conducted at school, she could hear students talking and screaming from the testing room.

**Exposure to the test.** Most of the participants stated that exposure to the test made them have negative feeling towards the TOEFL test.

Shinta: "Like in the practices I knew that it's gonna be very scientific which is not my like. Not My favorite thing so it is my another point" (personal communication, November 30, 2013)

The excerpt above represents the perceptions of other participants who perceived this subtheme negatively as well. They argued that during their preparation, they found that the TOEFL test was about a specific field. Another participant also claimed that during the preparation, she could not find what the TOEFL test really was. She felt that she got lost, because she barely understood the standard of the TOEFL test. Another point is

related to their previous experiences of the test. Since this participant knew how the test would be like, he experienced negative feeling when he was taking the test.

However, there are two participants claiming that the exposure to the test helped them to be familiar with the test.

Aci: “They are familiar enough. It is more like, I am not saying that, I am good at taking test, but it is just that I had more exposure to that kind of test. So that’s why” (personal communication, October 28, 2013)

**Stakes of the test.** When it comes to the stakes of the test, more than half of participants expressed stressfulness and anxiety during test taking. They stated that the position of the test that would determine their future lives (e.g., job and school requirement) made them have negative feeling.

Norah: “ok, the first one is a the stressful time, so, yeah because you try to reach the highest score as much as you can” (personal communication, November 16, 2013)

On the other hand, one participant expressed that she did not experience negative feeling when it comes to the stakes of the test. She said because “it is a test” (Aci, personal communication, October 28, 2013). Unfortunately, there was no further explanation about her statement.

**Technology.** The technology factor is the only factor that all participants agreed contributed to their negative feeling. They reported that the technology created anxiety and difficulties during test taking.

Imam: “I feel worried because it’s yes internet, in [name of a country] the internet connection is really really terrible so I was worried that in the middle of I

am doing the test suddenly there is a lost of connection, and then I can lost everything, yeah that kind of worried, yeah because if it really happens, it will change everything, it mean it will change your mood when you are doing the exam, and it will give you more a:: like what happened if I was in the middle doing it, and was it going to be saved or not or I have to redo it again, so basically in iBT TOEFL, yeah there is also worriedness about internet connection” (personal communication, November 3, 2013)

One participant also argued that the equipment used on the test such as the computer was not compatible. She claimed that the computers “were very very VERY old, and like the keyboard you could barely press the buttons and it was just stupid and horrible, really” (Shinta, personal communication, November 30, 2013).

In conclusion, most of the participants reported negative experiences when they took the TOEFL test. There are four factors: environment, exposure to the test, stakes of the test and the technology. Technology is the only factor that all participants agreed contributed to their negative experiences. Their technology anxiety is not related to their computer familiarity, but rather towards the compatibility equipment used for the test.

### **Topic in the Test**

As one of the important factors that can influence test takers’ success in working on the test, the topics covered in the test should not be leaning towards one mainstream discipline. It can be disadvantageous for test takers who are not coming from that prevailing discipline. Since all participants in this study mentioned this issue frequently, it means this issue is a profound topic that demands more attention from researchers and stakeholders.

This theme is mainly about participants' unfamiliarity of the topic. The participants' perceptions are equally divided into neutral and negative. The participants perceiving it neutrally expressed neither fondness nor disappointment towards the topic. They said that the topic in TOEFL test is academic. One participant argued that the topic in the TOEFL test was not a problem for her.

Aci: "and I did not really have challenge with the Topic of the test. So like the writing and stuff..they. They. They were not really emm a problem. A challenge for me" (personal communication, October 28, 2013)

The other three participants perceiving the topic negatively claimed that the topic was not general interest. The topic directed towards a particular discipline.

Shinta: "I think the worst part is about the topic, everything in that test was about topics that are not general interest, that are not easy for everyone, I know that it's not supposed to be easy but at LEAST something you can, you can build on for your own knowledge, but when you get that test, you don't you don't get anything like you feel that you have never done any studying before, it's completely new things, something so like specialize in one are, and that's hard, REALLY" (personal communication, November 3, 2013)

They argued that the test is mostly about a specific knowledge that is not familiar to them. Since they came from different field, they found the topic in the TOEFL test difficult. Reflecting on one participant's idea in which she expressed her understanding of test difficulties, that it is a nature of test to not be easy, but the topic in the TOEFL test was very specific. Thus, it can be implied that the topic in the TOEFL test is problematic.

To sum up, half of the participants in this study perceived the topic in the TOEFL test negatively. They argued that the topic in the TOEFL test was unfamiliar because they came from different disciplines. They stated that it contributed towards their performance on the test. Participants perceiving it neutrally argued that the TOEFL test is academic without showing either disappointment or fondness towards the topic.

### **Tasks in the Test**

As the TOEFL test intends to measure test takers' English proficiency, the tasks: Speaking, Reading, Writing, and Listening, should be designed properly so that the outcome will provide valid information about test takers' English ability. There are two points that the participants emphasized when they discussed about the tasks in the test: practicality and the measurement of English proficiency. The issue of practicality is related to the practices of these skills in the real life.

**Practicality.** Half of the participants argued that the tasks in the TOEFL test were not practical. Most of the participants emphasized the speaking task in which they had to listen and talk to computer. They stated that the speaking task did not portray a real life activity and hindered them to talk.

Andy: "I thought well first in the speaking part you talk with the computer and it really senseless, and it doesn't really make sense to talk to a computer, I think that if you talk with someone else you can talk more freely and you know share your ideas and experiences while if you speak with the computer or if your speaking is recorded and no one is speaking is actually speaking with you, that really it doesn't really allow you to talk that much" (personal communication, November 3, 2013)

One participant also elaborated that the tasks in the TOEFL test were not applicable because the skills were merely to pass the test.

Imam: “I mean for example I am teaching English education and the class doesn’t necessarily talk about TOEFL things, strategy or so on or skills that people need as the TOEFL, but still the campus the university forces students to pass” (personal communication, November 3, 2013).

However, two participants reported that the tasks were practical because they could apply it in an academic setting. They claimed that the tasks gave them a sense of academic setting, which they would face in the university.

Erica: “the good impression of the TOEFL ibt is ..i think it REALLY tries to measure the performance of the participants in academic setting. So, because from the type of the question that I have I can that the conclusion that they are really trying to know performance of the students in the real university contexts and also outside the classroom but in the real situation. I think they focus on those two contexts. I think that. It really well arranged” (personal communication, November 5, 2013)

**English proficiency measurement.** Almost all of the participants claimed that the TOEFL test did not measure their English proficiency, but rather their understanding of a specific topic. In other words, it is about their knowledge, not their English proficiency. This phenomenon points out a problem in the test. Since many students from different fields take TOEFL test, the topic in the tasks should not give advantage to one mainstream field. The participants argued:

Norah: “Yeah they examine your knowledge about that, knowing about certain thing like astrology and some you know once asked us about medicine construction of something like that” (personal communication, November 16, 2013)

Andy: “yeah, so if you want to hmm give a test to someone who learning English, I don’t think, I don’t think that you would ask him or her about something on geography, something on the surfaces of the earth, something like, something like chemistry biology, I remember biology, something in biology, I don’t think it makes sense, cause the test in here is not to hmm measure the knowledge of the test takers but it intend, this test intend to measure or to know how fluent test takers in English” (personal communication, November 10, 2013).

One participant who perceived it neutrally argued that the TOEFL test measured test takers’ proficiency in some ways, but did not in others. She claimed that the TOEFL test lacked in the speaking test, but good in other tasks such as the reading.

In conclusion, the participants in this study perceived the tasks in the TOEFL negatively due to the impracticality and the content. The use of technology made the test unpractical, because in real life test takers were not facing a computer when they were talking. Furthermore, the skills tested were not applicable for the test takers in their real lives.



## **Time in the Test**

The TOEFL test is conducted for 3 hours and 10 minutes. The test takers have a short break after finishing two tasks. In the interview, the participants also mentioned two themes that they discussed when it comes to the time in the TOEFL test: the duration of the test and the time allocated for each task.

**Length/duration of the test.** The participants, who mentioned the duration of the test, perceived the duration negatively. They argued that the test was too long. The fact that they have to face a computer for almost four hours was exhausting. When I mentioned the break time, one participant said that the break time was useless because they kept thinking about the next two tasks during the break.

Shinta: “yeah, it takes a long, like it takes about 4 hours and you are sitting on the computer, even I think if you are, if you are good to take this test, the timing itself it’s just horrible, like sitting for four hours” (personal communication, November 30, 2013)

Me: “I mean, you got the break right? Like after you finish reading and listening then you have a break 10 minutes I think

Andy: “yeah, which is really useless in my opinion cause it doesn’t really refresh you, because you go outside that room, you take the test, you start thinking about the two other parts of the test or the two sections of the test” (personal communication, November 10, 2013)

The issue of exhaustion/length made it difficult for them to concentrate during the test. They argued that even for the people who are good at taking test, four hours for a test is too long.

**Allocated time in the task.** More than half of the participants claimed that they faced problem with allocated time in the task. They mentioned that the time given was too short, which made them unable to concentrate and finish their task.

Erica: “they SHOULD give more time because I don’t talk with a person. I don’t make any communication with a person and...I think that the time given is kinda strict..kinda rigid and it doesn’t really represent while I write, while I speaking in the real communication context” (personal communication, November 5, 2013)

One participant perceiving the time given in the test positively said that only the speaking task gave a problem due to the short time, “yeah the speaking part section, only the speaking” (Imam, personal communication, November 3, 2013). But, for the rests of the tasks, he perceived the time positively.

One participant perceived the time in the test neutrally. She argued it is more about test takers management rather than the time itself.

Norah: “if you are not expert in time managing you will lost you know, even if the content is easier, you have to manage your time because time is running” (personal communication, November 16, 2013)

To sum up, most of the participants perceived the time in TOEFL test negatively. They argued that the test was too long which made them exhausted. Similarly, they also argued that the time given in each task is too short. The participant perceiving the time in

the test positively also claimed that the time in the speaking task was too short. He stated that “ i have a very very short time to prepare, and it makes me very very anxious” (Imam, personal communication, November 3, 2013). However, he perceived the time in other tasks positively. Finally, the participant positioning herself neutrally argued that the time in the test was not the problem, but the way test takers managed the time.

### **Score of the Test**

In the interview, all the participants also mentioned about the score that they obtained in the test. Half of the participants reported that the score they obtained on the test did not represent their real English ability due to the content and the idea of the test. One participant argued that the score that she obtained was bad. She stated, “Like I always told that my proficiency was not that bad, but the score was really really bad” (Shinta, personal communication, November 30, 2013). Another participant also argued:

Norah: “I have a friend who are physic like medicine major and doctor majoring, they get higher score, because they have already know about it, as I told you, I am thinking that this is tree and she told me “no, it’s an animal”, I have no idea, hmm yeah” (personal communication, November 16, 2013)

She explained that her friends who were not majoring in English were able to get a high score because they knew about the topic. It means the score represented test takers’ knowledge of topic rather than their English ability.

The participants maintaining their neutral position argued that the test has its system that has standard. In addition, another participant just mentioned feedback that she got along with her score. She perceived it neither positively nor negatively.

Me: “do you think that the score represent your English proficiency?”

Imam: “not necessarily, a test is a test” (personal communication, November 3, 2013)

Among six participants, one participant perceived the score positively. She claimed that the score that she obtained reflected her performance.

Erica: “mmm because when I got the result they also provide me with the guidelines or the comments why I got that particular score. I think and in my personal experience it really matches with my performance {laughing}.

Yeah I think TOEFL iBT” (personal communication, November 5, 2013)

In conclusion, most of the participants argued that the scores that they obtained in the TOEFL test were not their real English proficiency. However, two participants claimed that since the test has its standard, they perceived the test score neutrally. One participant claimed that the test score was representative because it matched with her performance in the test. In addition, she also claimed the fact that she did not see the interviewers make the score far from subjectivity issue.

### **Effects of the Tests**

In the interview, participants also talked about effects that they experienced after taking the test. They also put their afterthought towards the effects of test since they are people working in the area of language learning. Particularly they talked about two effects: their perceptions of their English proficiency and the societal effects.

**Perceived English proficiency.** Five participants expressed their perceptions of this issue. Two participants argued that the score gave negative effects towards how they perceived their language proficiency. One participant stated, “I feel like I am, I don’t I

don't like I am not strong enough in English after I got the score," (Shinta, personal communication, November 30, 2013). They argued that the bad score they obtained in the test made them feel not confident with their English.

On the other hand, two participants said that the score that they obtained did not really affect the way they perceived their English proficiency. They took into account the feedback that they got from the TOEFL iBT. They did not perceive her English negatively or positively because of the score. One of the participants said, "for TOEFL iBT, 'you got this because you are like this'. It really like you talking to particular person" (Erica, personal communication, November 5, 2013)

One participant claimed that the good score that he got in the test made him position himself like an "elite" in the society. He claimed, "You feel like you are in a kind a elite if you have passed 550 or 600 on paper, of course" (Imam, personal communication, November 3, 2013). He reported to be more confident in his English because he passed a certain score in the TOEFL test.

**Societal effects of the tests.** Societal effects become one of the prominent issues discussed by the participants. Since they have experienced the consequences of the test, they are able to express their thoughts toward this topic. The fact that they are able to express their perceptions of the societal effects of the test will provide more information about effects of TOEFL test.

Almost all participants perceived the effects of the TOEFL test negatively. They mentioned the issue of fairness, morality, and gatekeeping in the interview.

Andy: "yeah I mean sometimes your admission to a certain school can depend on that score, so if you score like 89 and they ask for 95, then you maybe you

cannot get in to that school, which I think I don't know, it really doesn't make sense" (personal communication, November 10, 2013)

Imam: "I have been also working, TOEFL is also a good business, because many people take it and the preparation, it's a good money, I think right now has become what do you call has less educational value, it's more like, people kinda make it into commercial thing," (personal communication, November 3, 2013)

They stated that the TOEFL test became the gatekeeper for students who wanted to continue to their higher education. If the students cannot pass the certain score, they have to wait for other chances or even bury their dreams to continue their education.

Furthermore, another participant also highlighted the decreasing of moral value because of the test. The test does not become the tools to measure someone's proficiency, but rather as a business. He also added that "some, a lot of people try to bribe you know so they can have fake TOEFL score" (Imam, personal communication, November 3, 2013).

To sum up, most of the participants agreed that TOEFL test has negative effects, which is not only related to them but also related to others.

From the descriptions of the participants' perceptions of the TOEFL test, it can be concluded that most of the participants perceived the TOEFL test negatively. They complained about the testing experiences, the topic in the test, the task given in the test, the time, the unrepresentative score and the effects of the test. Table 6 presents the summary of the participants' perceptions of the TOEFL test.

Table 6

*Frequency of Participants' Perceptions of the TOEFL Test*

Themes	Subthemes	Frequency of Participants' Perceptions (N=6)		
		Positive	Neutral	Negative
Testing Experience	Environment			83.33%
	Exposure to the test	33.33%		66.67%
	Stakes of the test		16.67%	66.67%
	Technology			100.00%
Perceptions of Topic	Unfamiliarity		50.00%	50.00%
Perceptions of Task	Practicality	33.33%		50.00%
	English proficiency measurement		16.67%	83.33%
Perceptions of Time	Length/Duration of the test			33.33%
	Allocated time in the tests	16.67%	16.67%	66.67%
Perceptions of tests score	Representativeness	16.67%	33.33%	50.00%
Effects of the tests	Perceived English proficiency	16.67%	33.33%	33.33%
	Societal effects		16.67%	83.33%
Frequency Total (%)		116.67	183.33	766.67

**Participants' Perceptions of the IELTS Test**

After asking the participants of their perceptions of the TOEFL test, I asked them about their perceptions of the IELTS test. Almost all participants mentioned similar points that they previously mentioned in the TOEFL test. It implies that these characteristics are profound topics in test taking experiences.

**Testing Experience**

In the interview, the participants expressed and described their IELTS testing experience. When they explained this theme, some of the participants tended to compare

it with the TOEFL test. However, it did not lessen the purpose of this section in which to get information of participants' perceptions of their IELTS testing experiences.

**Environment.** Half of participants reported that the environment of the IELTS test was conducive. They described the environment was good for testing because there was no noise and the seat was comfortable. This supportive environment helped them to focus on doing the test.

Shinta: “like in the IELTS we were sitting in a class, like normal exam which we were there was someone (not clear) everyone in her own table, you have your own table, you have your own headset and the room is so so quiet, no one is allowed to speak, so it's more more organized than the TOEFL, you have the feeling that you are really in a test and in a quiet room that help you focus more on the test, so it is more organized” (personal communication, November 30, 2013)

However, one participant argued that the environment when he took the IELTS test was not supportive due to the huge number of candidates in one small room.

Imam: “this is about the test site, because they are like, probably more than 20 people and it has small room, and yeah you know, it's very crowded” (personal communication, November 3, 2013)

To sum up, most of the participants argued that the environment of the IELTS test was good for test takers. They reported the seat, the room, and the space on the test enabled them to concentrate on the test. Yet, one participant had different experience because the room was too small to occupy twenty test takers.



**Exposure to the test.** When the participants discussed their exposure to the test, almost all participants perceived it positively. They stated the exposure to the test helped them to know about the test and the content of the test.

Erica: “I know that IELTS has many a..many versions but as I know the standard one for the IELTS is Cambridge so ..so.. I think I was more confident in learning and that impact when I was doing the real test” (personal communication, November 5, 2013)

Norah: “as I told you it’s only one week between the booking and the exam, she give me a book called target Ban, if you know it, it’s a great one, so it’s full of strategies and most of them is work for me,” (personal communication, November 16, 2013)

The participants claimed that the preparation to the IELTS test made them feel confident to face the test. They argued that since the IELTS test had clear standard, it made the test preparation manageable.

On the other hand, one participant perceived the exposure to the IELTS test negatively. Since he had difficulties to buy IELTS books, get access to the IELTS test, which contributed negatively towards his preparation.

Imam: “I mean the type of the questions you know, like what are exactly the content of the test... hmmm it’s not easy to find books about IELTS, most of the time it’s easier to find TOEFL preparation book,” (personal communication, November 3, 2013)

In conclusion, almost all participants agreed that exposure to the IELTS test helped them to get ready for the test. They claimed that because of the preparation, they were able to understand the content and the tasks in the test.

**Stakes of the test.** When it comes to the stakes of the test, most of the participants said that it gave them negative feeling when they took the test. Since the test would play important role for their future lives, such as education and job, they experienced stressfulness during the test.

Shinta: “hm so the first one is stressful too because the IELTS I took the IELTS because I wanted to apply for university so it wasn’t about a job but it was about my higher education so it was somehow stressful” (personal communication, November 30, 2013)

However, two participants in the study perceived the stakes of the test neutrally because the test did not have direct effects toward their lives.

Imam: “oh yeah because I mean first IELTS I took it years I have no motivation it, I just kinda to experience it, so probably people took it for real reason like, to achieve they probably they probably have,..” (personal communication, November 3, 2013)

It can be implied that when a test’s position is at high stakes, it can create negative feeling towards students’ testing experience. During working on the test, they kept thinking about the stakes of the test in which their future lives would depend on. On the other hand, when a test does not affect test takers’ future, it does not create such high-pressure feeling to the test takers.

**Technology.** The IELTS test is a paper-based test. In the listening section, all the test takers listen from the one audio system. Most of the participants expressed fondness to the technology used in the IELTS test. The fact that they could read and write on paper, and then highlighted the things that they thought important made them perceive the IELTS test positively.

Shinta: “number five I like it is more because it’s paper based, I don’t like, like reading from the screen it’s TOO much for me, so this one as even for the writing part, I am not like fast in typing so paper makes me like write more, so I like it because of that” (personal communication, November 30, 2013)

However, two participants perceived the technology used in the IELTS test negatively. They argued that the audio system in the listening section did not support the test takers.

Andy: “for example in the listening part you [inaudible] with the speaker. Sometimes you don’t hear the speaker very well because they use, they use the mm the not a headphone, < ... > at this time we didn’t really use headphone so you may miss something, I mean I did good but for others they maybe you know miss many points” (personal communication, November 10, 2013)

Since the test takers did not have headphone on the test, they were afraid that they might lose some points during the listening. They claimed that one audio system might not reach all test takers. Therefore, the seat arrangement could create either advantage or disadvantage to the test takers.

In conclusion, most of the participants perceived the technology used in the IELTS test positively. They argued that it was more comfortable to do test on paper than computer. Also some participants said that they did not type fast which could create difficulties for them. On the other hand, two participants complained about the technology used in the listening test. They claimed that the unavailability of headphone might hinder them to hear all the listening content. In addition, since IELTS test used one audio system to cover all test takers, the seat could become an important factor to be successful in the listening test.

### **Topic in the Test**

In the interview, all the participants mentioned the issue of topic in the test. There was no participant perceived the topic in the test negatively. They claimed that the topic in the IELTS test considered the test takers' backgrounds.

Norah: "one of the question about coffee which I like it a lot, and story of coffee and talk about the coffee mineral, about one of the machines that make coffee, whenever you read , you know, it's easy, they didn't talk deeply about the machine of coffee, no, in the surface one," (personal communication, November 16, 2013)

Shinta: "I have a reading about museums which is fine, everyone know, Ok I wasn't very aware what was the whole idea about the museums, but still something you can manage, you can predict the meaning of the words" (personal communication, November 30, 2013)

They claimed that the topics were manageable. Even though they were not familiar enough with the topic, they favored it because the discussion was not in detail. They stated that test takers could guess the meaning with their English knowledge. On the other hand, two participants perceived the topic in the test neutrally. They mentioned that the topic in the IELTS test was general.

To sum up, most of the participants perceived the topic in the IELTS test positively due to the level of in-depth knowledge discussed. Two participants in this study perceived the topic in the IELTS test neutrally because they claimed that the topics discussed in IELTS test were general and common.

### **Tasks in the Test**

Almost all of the participants discussed the topic of tasks in the IELTS test. The tasks in tests are important because the score of the test is based on how the participants do in the tasks. This implies that the tasks should be designed carefully to measure test takers English ability. In the IELTS test, the tasks are divided into reading, listening, writing and speaking. For the speaking task, each test taker has it with an interviewer.

**Practicality.** Most of the participants perceived the tasks in the IELTS test were practical. They particularly pointed out the speaking task. They claimed that doing the task with an interviewer made it authentic like a real life performance.

Andy: “I found it really LOVELY actually because I had to speak with someone and then he was asking me you know we use body language, you know gestures not only about you know utterances what we say, it wasn’t about you know verbalizing words, it was really an experience, I would say THIS IS a speaking experience,” (personal communication, November 10, 2013)

One participant also argued that the tasks in the IELTS test are practical in general context. She clarified it by stating:

Erica: “the IELTS test::: it really measures the ability of the second language, the English language learners in four different aspects yeah reading listening and writing and speaking, but not really academically but .. I mean. They really measure in general. In general one. The ability.” (personal communication, November 5, 2013)

**English proficiency measurement.** More than half of the participants claimed that the IELTS test measure their English proficiency.

Shinta: “due to everything we mentioned before, like it’s more about your general knowledge of English. It is something, somehow you could, that could measure your English, somehow not to a great extent, but for example like in the reading, there was like a summary, fill in the blank for the reading and this part you could everything about using your knowledge about part of speech, because they will give you about maybe five or six words and you have only to change the form into to make them fit into the blank and that’s basically based on your knowledge of part of speech and how to make form noun, verb and adjective, just to fill in the blank, so yeah I think it’s somehow measure your ability,” (personal communication, November 30, 2013)

The aspects that make them come to those perceptions are related to the content of the test. They claimed that the test really tried to measure test takers’ English proficiency, not their knowledge of a specific topic.

However, one participant perceived it negatively because he did not believe that this test measure test takers' English as a foreign language. He argued that the test was about measuring test takers' English based on one particular mainstream society.

Imam: "I think as an [nationality] and teacher myself I think, I mean we should more [name of a country] version of standardize English test rather than TOEFL and IELTS because in my experience a lot of my students will not go abroad, I mean, those two exams are not, do not measure the real proficiency in English as a foreign language," (personal communication, November 3, 2013)

In conclusion, the participants in this study perceived the tasks in the IELTS test positively because its practicality and degree of measuring test takers' language proficiency. They stated that the test measured their English proficiency not their knowledge of a topic in English. For the participant who perceived it negatively, he argued that the test measured test takers' ability of a particular English. On the other hand, one participant perceived it neutrally. She maintained her neutral position because she argued that there is no perfect test, each test has its own weaknesses.

### **Time in the Test**

The duration of the IELTS test is 2 hours 44 minutes. The three tasks: reading, listening and writing, are conducted in a row with a short break after each task. The speaking task can be conducted on different days or on the same day depending on the number of the participants.

**Length/Duration.** Most of the participants perceived the duration of the test positively. They argued that the time given was appropriate for a test. Furthermore, since the speaking task was conducted on different days making them feel more relax.

Norah: “IELTS you have like Reading listening and writing AND you have a chance to like speaking on the same day or another day, so it’s like relieving thing, you feel relax, you have like flexibility” (personal communication, November 16, 2013)

Most of the participants favored the idea of the test being conducted for almost three hours, especially because they could do the speaking test on different days. One participant got her speaking test on the same day. However, since she had some break time before the speaking task, she perceived it positively.

**Allocated time in the task.** In the interview, the participants also mentioned their perceptions towards the time allocated in the task of the test. One participant perceived the time in the IELTS test positively. She argued that the time given in the test was fair enough for test takers to finish the task. On the other hand, two participants perceived the time neutrally. They claimed that the time management was important aspect enabling them to cope with the time in the test.

Aci: “I didn’t personally I didn’t have problem with the time. Well my time management was pretty good, and even for the writing section I finished before the time is up, so I didn’t have any problem with that.” (personal communication, October 28, 2013)

In conclusion, none of the participants in the study perceived the time in the IELTS test negatively. The first reason is related to the length of the test, which is not too



long or too short for a test. The second reason is the time allocated for the task in the test. Two participants perceiving the test neutrally argued that time management was the key to deal with time in the test.

### **Score of the Test**

The score in the IELTS test starts from 1 to 9, ranging from non-user to expert-user. Five out of six participants mentioned this topic in the interview. Half of them perceived the representativeness of the IELTS score positively, “somehow yes, yeah to some extent, I feel more confident, ok that could be my level in English,” (Shinta, personal communication, November 30, 2013). They perceived the score positively because the test measured their English proficiency. They argued that the test consisted of questions designed to measure test takers English abilities. Therefore, when they received the score, they accepted it positively.

Yet, one participant argued that her IELTS score did not represent her real English ability because there might be subjectivity aspect during the assessment.

Erica: “yeah, there is even like a joke like in IALF there were interviewers or several evaluators for the writing and the speaking and WE kinda like “oh I hope I will not get that particular person because his so strict, so I hope I can get the normal one” that is kinda joke {laughing}” (personal communication, November 5, 2013)

One participant perceived the score neutrally. She showed neither satisfaction nor disappointment towards the score she obtained. She explained the content of the score that she obtained.

Aci: “And ...mmmm yeah, and then the when you take a look at the result, the the what do you call it, the document.like, you would see clearly for ..for IELTS like they separate the score for each section and then you see that ok for example like oh you got nine for this, you got nine for this and stuff and stuff.” (personal communication, October 28, 2013)

In conclusion, half of the participants argued that the score they obtained on the IELTS test represented their real English ability. They claimed that the test measure their proficiency, therefore the score they obtained was their real abilities. One participant argued that the score did not reflect her real proficiency because there might be a subjectivity element in scoring. One participant perceived it neutrally because she did not show either negative or positive perception of IELTS score regarding the representativeness.

### **Effects of the Test**

**Perceived English proficiency.** Half of the participants in this study reported that the score that they obtained in the IELTS test influenced the way they perceived their English ability. They said that they felt more confident when they got the IELTS test result.

Shinta: “I got a better score which is like, there was about one year between the two tests. And I got six in the TOE in the IELTS which is very good score, like it compared 100 in the TOEFL, so I feel like like more confident I guess about it,” (personal communication, November 30, 2013)

They obtained good score from the test that they perceived positively. Therefore, it made them feel more confident towards their English. The rest of the participants did not discuss this topic in the discussion.

**Societal effects of the tests.** Most of the participants perceived the societal effects of the IELTS test neutrally. One participant claimed that IELTS test could be used for screening students. She argued, “so as I told you, certain major in university should have IELTS” (Norah, personal communication, November 16, 2013). Since students are positioned to have their English proficiency score to register to universities, the IELTS test is appropriate, but only for specific majors such as social science.

One participant perceived the effects of the IELTS test negatively. She stated that the problem of standardized test was gatekeeping. She argued that standardized tests made people experience stressfulness and anxiety. It can also be inferred that the other societal effect is related to the impact of the test to the test takers.

Shinta: “they are similar because they are all of them standardized test, and they are like stressful because people take them because to do something else with them not just taking them for fun.” (personal communication, November 30, 2013)

Based on the descriptions, it can be concluded that most of the participants perceived the IELTS test positively. They claimed that their testing experiences were good due to the environment and the technology used. Also, they argued that the topic, the tasks and the time in the test were manageable. Regarding the score, most of the participants argued that the IELTS score is representative. For the effects of the tests,

participants mainly perceived it positively and neutrally. Table 7 presents the summary of participants' perceptions of IELTS test.

Table 7

*Frequency of Participants' Perceptions of the IELTS Test*

Themes	Subthemes	Frequency of participants' perceptions (N=6)		
		Positive	Neutral	Negative
Testing Experience	Environment	50.00%		16.67%
	Exposure to the test	83.33%		16.67%
	Stakes of the test		33.33%	50.00%
	Technology	50.00%		33.33%
Perceptions of Topic	Unfamiliarity	66.67%	33.33%	
Perceptions of Task	Practicality	83.33%		
	English proficiency measurement	66.67%	16.67%	16.67%
Perceptions of Time	Length/Duration of the test	83.33%		
	Allocated time in the tests	16.67%	33.33%	
Perceptions of tests score	Representativeness	50.00%	16.67%	16.67%
Effects of the tests	Perceived English proficiency	50.00%		
	societal effects		50.00%	16.67%
Frequency Total (%)		600.00	183.33	166.67

**The Comparison of Participants' Perceptions of the IELTS and TOEFL Tests**

Having discussed participants' perceptions of each test, this section presents the comparison of participants' perceptions of both tests. As this research intends to explicate the differences between these two tests, the table from each test is combined together to show the differences.

Table 8

*Comparison of Participants' Perceptions of the IELTS and TOEFL Tests*

Themes	Subthemes	Frequency of participants' perceptions (N=6)					
		Positive		Neutral		Negative	
		IELTS	TOEFL	IELTS	TOEFL	IELTS	TOEFL
Testing Experience	Environment	50.00%				16.67%	83.33%
	Exposure to the test	83.33%	33.33%			16.67%	66.67%
	Stakes of the test			33.33%	16.67%	50.00%	66.67%
	Technology	50.00%				33.33%	100.00%
Perceptions of Topic	Unfamiliarity	66.67%		33.33%	50.00%		50.00%
Perceptions of Task	Practicality	83.33%	33.33%				50.00%
	English proficiency measurement	66.67%		16.67%	16.67%	16.67%	83.33%
Perceptions of Time	Length/Duration of the test	83.33%					33.33%
	Allocated time in the tests	16.67%	16.67%	33.33%	16.67%		66.67%
Perceptions of tests score	Representativeness	50.00%	16.67%	16.67%	33.33%	16.67%	50.00%
Effects of the tests	Perceived English proficiency	50.00%	16.67%		33.33%		33.33%
	Societal effects			50.00%	16.67%	16.67%	83.33%
Frequency Total (%)		600.00	116.67	183.33	183.33	166.67	766.67

Table 8 shows the comparison of participants' perceptions between IELTS and TOEFL tests. Among six themes of the data, there are some significant differences between IELTS and TOEFL test. On testing experience theme, 50% of the participants perceived the environment of the IELTS test positively. Meanwhile 83.33% of the participants perceived the environment of the TOEFL test negatively. Regarding the technology, 100% of the participants claimed that the technology used in the TOEFL test negatively influenced their testing experience. In contrast, 50% of the participants favored the technology used in the IELTS test. In the issue of stakes of the tests, 50% of the participants perceived the IELTS test negatively. Similarly, 66.67% of the participants perceived the TOEFL test negatively.

When it comes to the perceptions of the tasks, 83.33% of the participants claimed that the tasks in the IELTS test were practical. In contrast, 50% of the participants said that the tasks in the TOEFL test were impractical. In relation to the English proficiency measurement, 66.67% of the participants perceived this issue positively for the IELTS test, and 83.33% of the participants perceived negatively for the TOEFL test. This result shows that most of the participants agree that tasks in the IELTS test are more practical, and measure test takers' English proficiency effectively compared to TOEFL test.

The next theme is the participants' perceptions of the time in the test. 83.33% of the participants perceived the duration of the IELTS test positively. Meanwhile, 33.33% of them perceived it negatively for the TOEFL test. Concerning the time allocation for the tasks, 33.33% of the participants positioned themselves in a neutral area for the IELTS test, while 66.67% of the participants positioned themselves in a negative area for the TOEFL test. In terms of the effects of the test, 83.33% of the participants claimed that

TOEFL test has negative effects towards society, while 50% of them perceived the effects of IELTS test neutrally. In conclusion, most of the participants perceived the IELTS test positively. Meanwhile, they perceived the TOEFL test more negatively in all themes. There are two themes where the distribution of their perceptions is almost equally divided: perceptions of the topic in the test and perceptions of the test score.

### **Conclusion**

This chapter provides results of the analyzed data gathered from the participants. There are six sets of themes emerged from the analysis: testing experience, perceptions of topic, perceptions of tasks, perceptions of time, perceptions of score and perceptions of effects of the test. The results show that participants in this study have different perceptions of IELTS and TOEFL tests. The prominent differences are: the testing experience, perceptions of the task, perceptions of the time, and perceptions of the effects of the tests. The rest of the themes, perception of the topic and perceptions of the score, did show different perceptions, yet equally divided into positive, neutral and negative perceptions. The next chapter presents the discussions of these findings.

## CHAPTER FIVE

### DISCUSSIONS AND CONCLUSIONS

This study aims to explicate differences between students' perceptions toward two high-stakes tests, IELTS and TOEFL. Having conducted data collection, data analysis and presenting the result of analyzed data, I discuss the result of the analyzed data in this chapter. Following the discussion part, I present the conclusion, the implications of this study, the limitations, and recommendations for future studies.

This study is an endeavor to give test takers chances to vocalize their perceptions towards tests because “it is very rare that a test taker could protest, complain or claim that the test did not fit his or her knowledge or that he or she was not expected to master that knowledge” (Shohamy, 2001, p. 154). Some previous studies focusing on test takers' voices proved that giving test takers chances to share their experiences would provide information which is valuable in designing and using tests. This phenomenon asserted Shohamy's (2001a) idea where she argued, “listening to the voices of test takers can provide testers with a new and unique perspective and a deep insight into tests and their meanings” (p. 7). In a sense, listening to test takers is a profound issue to bring into discussion because they are the people who have experienced the tests and the consequences. This attempt will enable stakeholders, especially testers, to understand that listening to takers' voices is important.

#### **Different Perceptions of IELTS and TOEFL Tests**

There are three aspects of IELTS and TOEFL tests that the participants in this study perceived differently: on the tests (e.g., the environment and the technology), in the tests (e.g., the tasks and the time), and the effects of the tests. These three aspects also



provide information on how participants in this study perceived high-stakes tests, conceptualized ideal tests, and criticized the effects and the uses of the tests.

### **On the Tests Aspects**

The differences between participants' perceptions of IELTS and TOEFL test taking experiences lie in the environment and the technology used on the tests. According to the participants, the environment of the TOEFL test was not appropriate for test taking because of the noise, the space, and the other test takers. Puspawati's (2012) research also showed similar findings in which test takers criticized the environment of the TOEFL test due to the same factors. The technology used in the TOEFL test, such as the internet and the equipment, also made the participants in this study experience anxiety. In addition, they reported eye fatigue because of the long exposure to computer monitors. Technology is mentioned by Puspawati (2012) and Sawaki (2001) as a factor that can contribute towards test takers' performance on tests.

In contrast, participants in this study perceived technology and environment in the IELTS test positively. They claimed that the good environment (e.g., quiet room) and technology (such as paper-based test and audio system) in the IELTS test made them able to work better. Luce and Kapler (2007) and Puspawati (2012) argued that when test takers felt uncomfortable with the environment or the technology used in the test, they were not able to perform to the best of their ability.

Besides the differences, there is one similarity between their perceptions of IELTS and TOEFL testing experiences: the stakes of the tests. Since the tests had direct effects in their lives such as pursuing higher education or applying for a job, it made them experience negative feelings when they worked on the tests. This result supports findings

from previous studies such as Triplett, Barksdale, and Leftwich (2003), Triplett and Barksdale (2005), Puspawati (2012), and Murray, Riazi, and Cross (2012) whose studies focused on students' perceptions on high-stakes tests. They reported that students felt nervous, stressed, and anxious during the tests.

On the other hand, two participants in this study did not experience negative feelings when they took the IELTS test. Their reason for taking the test was not for any requirements, but rather to experience it. Therefore, the participants in this study experienced different levels of stressfulness when the tests have different consequences toward their lives. This finding deepens our understanding of tests' position's influences on test takers' testing experience.

Altogether, IELTS and TOEFL tests gave negative experiences to the students when the tests were at high stakes. This study reveals that comfortable environment and supportive technology will help test takers work better on the tests. This study also supports Puspawati's (2012) findings that the way the TOEFL test is conducted is not supportive for the test takers.

### **In the Tests Aspects**

IELTS and TOEFL tests look similar based on skills tested: Reading, Writing, Listening, and Speaking. These two tests claim to possess the authority to measure individuals' English proficiency. However, the participants in this study perceived these two tests differently. They lean more positively toward the IELTS test instead of the TOEFL. There are two prominent reasons for their position: tasks and the time given in the tests.

The first point is the tasks. The participants argued that the tasks in the TOEFL test were impractical, and did not measure test takers' English proficiency. This finding is similar to Puspawati (2012) in which she argued that participants did not find the TOEFL test practical, nor did it measure their English ability. In this study, most of the participants said that the format of the test, such as speaking with a computer, did not represent the real use of English. The participants also argued that the skills that were being tested on in the TOEFL did not have relation to the use of English in the class. Furthermore, the content of the test assured them that the tasks in the TOEFL test did not measure English proficiency, but rather their knowledge of specific subjects. They claimed that the topics in the TOEFL test were unfamiliar and discussed in great detail, making it difficult for them to understand the content. This finding asserts Puspawati's(2012) and He and Shi's (2008, 2012) findings in which they found that topic unfamiliarity created difficulties to test takers.

On the contrary, the participants perceived the tasks in the IELTS test positively. IELTS test practicality for the real life performance and its content assured them that the test intended to measure their English proficiency. This finding aligns with Rasti's (2009) research where she found that test takers also had positive attitudes toward the test. Participants in her study argued that the IELTS test measured their "linguistics and communicative competence precisely" (p. 145). In addition, participants in this study perceived the IELTS test's topics positively even though they were not familiar. They argued that since the topic in the IELTS test was not discussed in deep detail, they found it more manageable. One participant asserted this claim by stating that the IELTS test considers the test takers' backgrounds. This argument is in line with Shohamy's (2001b,

2004) notion of democratic assessment, pointing out the issue of incorporating knowledge of test takers in developing tests. Furthermore, the issue of unfamiliarity connects with the fairness issue in the test. Tests should be designed without excluding any group in society (Xi, 2010). Since test takers come from many different fields of knowledge, it will be fair if topics in the test do not lean towards one mainstream field, or are discussed in deep detail.

The last topic is the time, covering the duration and time allocated for each task. Participants in this study perceived the time in the TOEFL tests negatively because the duration is too long. In addition, they complained of the limited time allocated to finish the tasks. This result is in line with Puspawati's (2012) finding in which her participants criticized the time given in the TOEFL test.

In contrast, the participants in this study found the time in the IELTS test fair enough. They favored the duration of the test and the time allocated which helped them to perform better. This finding supports Rasti's (2009) result in which she argued that time given in the IELTS test was perceived positively by her participants.

In conclusion, the participants in the study perceived IELTS and TOEFL tests differently. They lean towards the IELTS test because it is similar to their concept of ideal tests. They argued that tests should consider test takers' backgrounds. Generalizing test takers' qualifications such as cultures, countries, and fields of study were regarded by participants as unfair. Therefore, considering and including test takers' knowledge in creating tests is essential (Shohamy, 2004). Furthermore, their concept of ideal tests is related to the time given in the tests. The timing should not create disadvantage to test

takers to work on the test. These findings show that applying democratic assessment in the high-stakes tests is achievable.

These findings also highlight that even though the tests look similar, the content, the manifestation, and the way the test is conducted are important matters which differentiate these two tests. The participants argued that an ideal test should consider those three elements. The content should consider diverse knowledge of test takers, the manifestation should be delivered clearly, and the way it is conducted should be appropriate to support test takers in working on the test.

On a different level, based on the data, it is clear that participants in this study perceived the tasks and the time in the IELTS test positively. These factors might positively affect their perception toward score representativeness. However, some of the participants who had positive attitudes toward the IELTS perceived the score either neutrally or negatively. They argued that the IELTS score might not represent their real English ability. One of the factors was the interviewers' bias which has become one of the critical topics for the IELTS test. Brown (2003) found that interviewer variation in eliciting test takers' communicative competencies could affect test takers' performance. This bias could influence raters' judgments of test takers' abilities.

For the TOEFL test, they argued that since they got clear feedback of their performance, it gave them a sense that the score represented their real performance on the test. In contrast, they argued that the IELTS test's feedback was too general. It is not addressed to a particular test taker.

To sum up, based on participants' perceptions these two tests were still lacking in representing test takers' real abilities. However, they claimed the score in the IELTS test

was more representative than the TOEFL test. This implies that test takers still question the representativeness of the tests. In addition, they highlighted their need for feedback. By being able to provide comprehensible feedback of test takers' performance in the test, testers show their responsibility of their judgment. Shohamy (2004) argued that testers have to be responsible for their actions.

### **Effects of the Tests Aspects**

One of the effects of these two tests is related to the way participants in this study perceived their English proficiency. Even though participants in this study were aware that the score of the tests did not represent their real English ability, it more or less influenced the way they perceived their English proficiency. The effect is parallel with their score. Test takers obtaining higher scores stated that they felt more confident of their English and vice versa. This phenomenon asserts the powerful effect of tests towards test takers. Since tests possess features such as the power of numbers (Shohamy, 2001a), they can affect the way people perceive their abilities. The dimension of numbers in the test score does not only relate to the matter of representativeness but also self-recognition from the view of test takers. Therefore, there is a need to create a test that is able to present test takers' real ability.

Another effect that participants in this study pointed out is the societal effects: gatekeeping, fairness, morality, and ethicality (Shohamy, 2001a). Since tests are conducted by powerful organizations and required by influential institutions such as universities, the power of tests has expanded beyond its original intention, measuring test takers' proficiency. Participants in this study criticized the use of the tests as the requirement to enter school or apply for a job. They stressed particularly on the use of

the TOEFL test. Since they claimed that the TOEFL test itself is unfair, it will be unfair if the test is also used as a requirement. This shows that the participants in this study felt that they were treated unfairly by the test. The unfairness issue of the test occurs not only when they took the test, but also after they took the test.

Furthermore, they mentioned another example of the effect of the tests, the morality and ethicality (Shohamy, 2001a). They stated that tests were able to make people change their behavior in order to meet test demands or to be successful in the test. For example, people may pay other people to take tests on their behalf. In addition, tests also affect test takers' ways to interpret passages in the test. Test takers are in the powerless position where they "could not refuse to take the test, and were forced to look for the exact answers the test makers intended" (p. 49).

However, despite the fact that the IELTS test is used as a gatekeeper to enter universities, participants in this study showed more leniencies towards it. They argued that since IELTS test considers test takers background regarding the content of the test, it made them perceive the use of the test neutrally. There were fewer unfairness issues in the IELTS test. Based on participants' data, if the test is designed and conducted by considering and including knowledge from test takers, participants might not negatively perceive tests as a gatekeeper.

### **Conclusion**

The aim of this study is to explicate the differences between test takers' IELTS and TOEFL testing experiences. As discussed in this chapter, participants perceived three aspects of the two tests differently: on the tests, in the tests, and effects of the tests. First, participants' perceptions of the on test aspect highlighted that technology and

environment in the TOEFL test were not supportive for the test takers. In contrast, IELTS test provided test takers with a more comfortable and convenient environment and technology. Second, participants' perceptions of the in test aspect emphasized that the TOEFL test did not measure test takers' language ability due to its impracticality. Participants also expressed their opinion that the TOEFL test did not intend to measure language ability, but rather knowledge of a subject. However, their perceptions towards the IELTS test were the opposite. Finally, participants' perceptions of the effects of tests aspects pointed out more negatives than positives, referencing gatekeeping, fairness, ethicality, and morality.

In addition, this study also provides information of high-stakes tests testing experience, a concept of ideal tests, and critiques of the effects of the tests. The first one is the parallel correlation between the tests and their influences on test takers' experiences. For instance, there is an equivalent connection between the position of the test and their level of stressfulness. This means, the higher stakes of the tests entails the higher level of stressfulness the test takers will feel. The second point is the concept of ideal tests. Test makers should consider test takers' backgrounds in creating a test since test takers come from diverse groups. Therefore, it is important to create a test that is not leaning towards one mainstream society. Also, supportive environment and technology should be provided. The last topic is the effects of the test. The effects of tests can be minimized if the tests are conducted in democratic ways aiming at "monitoring and limiting the uses of tests, especially those that have the potential to exclude and discriminate against groups and individuals" (Shohamy, 2001b, p. 373). In order to achieve this, five democratic practices should be applied: incorporating diverse groups'



knowledge on tests, cooperating test takers in conducting tests, applying Critical Language Testing (CLT), protecting test takers' rights, and addressing test makers and testers' responsibilities (Shohamy, 2001b).

### **Implications of the Study**

There are some implications that can be derived from this study. First, this current study exposes the factors (e.g., tasks and time in the tests) becoming the concerns of test takers in the high-stakes tests. This means, these factors are prominent elements in tests that need to be looked at more thoroughly by those involved in testing act. Second, this study also addresses issues that will remain as questions, such as whether the problem in language tests are about the specific test or the language tests in general, if studies only focus on investigating test takers who have taken one test. The participants in this study pointed out bigger problems in the TOEFL test rather than in the IELTS test. This issue raises awareness that some aspects in both tests need to be revisited. In addition, participants conceptualized and criticized the use of the tests based on their experiences. Participants in this study revealed that if the test were designed fairly, they would perceive the use of test as gatekeeper neutrally. This means that the content, the manifestation, and the way tests are conducted are important matters to be reviewed carefully. Also, regarding the three elements mentioned above, this study highlights the different values that IELTS and TOEFL tests use. The IELTS test should be used in the Humanities field and the TOEFL in the Science field. Third, applying the practices of democratic assessment emphasizing on inclusion and non-discrimination is achievable. Finally, this study implies the importance of giving valid and reliable feedback to test takers

### **Limitation of the Study**

There are two major limitations of this study. The first limitation is the small number of the participants. Therefore, the finding of this study cannot be over generalized. The second limitation lies on the fact that I, as researcher, do not have information about my participants' test taking time. In other words, there is no data about which test was taken first and or when it was taken. This detail might or might not influence the participants' perceptions towards IELTS and TOEFL tests.

### **Recommendations for Future Studies**

This study reveals the importance of feedback in testing. This can be the new area of study focusing on high-stakes tests. To date, I have not found any studies investigating students' perceptions of the high-stakes tests feedback. Furthermore, this study also exposes a correlation between participants' test score and their perceived English proficiency. Conducting this research can be the way to have better understanding about the powerful effect of tests which not only affect test takers' future lives, but also the way they perceive their English ability. Finally, since the participants in this study are English majors, it will be valuable to know how test takers from other majors perceive IELTS and TOEFL tests. The result of this future study can be compared with the current study, and eventually deepen our understanding about test takers' perceptions of high-stakes tests.

### **Final Reflections**

I realized that conducting this study might not change the way IELTS and TOEFL tests are conducted. However, by conducting this research I recognized that test takers have been treated unfairly. I did not use to pay attention to this issue because I just accepted it as a part of systems that I had to follow. This research enlightened me that test

takers are in the powerless position because they are treated as recipient. Furthermore, this study helped me to understand that listening test takers voices is a way to fight the unfairness and other consequences. They are the individuals who experience direct effects of the tests. I hope that this research can raise stakeholders' awareness towards the influences of high-stakes tests. In addition, I learned many things during the process of writing this thesis. Every single step I have done with the guidance from my professor helped me to understand the meaning of research. Conducting a research is also a way for me to be more aware about phenomenon and issues around me.

## References

- About the IELTS. (2013). International English Language Testing System (IELTS). Retrieved from [http://www.ielts.org/test\\_takers\\_information/what\\_is\\_ielts/ielts.aspx](http://www.ielts.org/test_takers_information/what_is_ielts/ielts.aspx)
- About the TOEFL. (2013). English Testing Services (ETS). Retrieved from <http://www.ets.org/toefl/ibt/about/>
- Alderson, J. C. (2009). Test of English as a foreign language: Internet-based test (TOEFL iBT). *Language Testing*, 26(4), 621-631. doi:10.1177/0265532209346371
- Alptekin, C. (2006). Cultural familiarity in inferential and literal comprehension in L2 reading. *System*, 34, 494-508.
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty and student learning. *Education Policy Analysis Archives*, 10(18), 1-74.
- Attali, Y. (2007). Construct Validity of e-rater® in Scoring TOEFL® Essays (Research Report Vol. 21). Retrieved from ETS Research Reports website: <http://www.ets.org/Media/Research/pdf/RR-07-21.pdf>
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Au, W. (2009). Social studies, social justices: W(h)ither the social studies in high-stakes testing?. *Teacher Education Quarterly*, 36(1), 43-58. doi:10.3102/0013189X07306523
- Au, W. (2011). Teaching under the new Taylorism: High-stakes testing and the standardization of the 21<sup>st</sup> century curriculum. *Journal of Curriculum Studies*, 43(1), 25-45. doi:10.1080/00220272.2010.521261

- Bachman, L. F., & Purpura. J. E. (2008). Language assessments: Gate Keepers or door openers? In B. Spolsky, & F.M. Hult (Eds). *The Handbook of educational linguistics* (pp. 456-468). Malden, MA: Blackwell.
- Beitin, B. K. (2012). Interview and sampling: How many and whom. In J.F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds). *The SAGE Handbook of interview research: The complexity of the craft* (2<sup>nd</sup> ed.) (pp. 243-254). New York, NY: Sage.
- Bridgeman, B., Powers, D., Elizabeth, S., & Mollaun, P. (2011). TOEFL iBT speaking test scores as indicator of oral communicative language proficiency. *Language Testing*, 29(1), 91–108. doi:10.1177/0265532211411078
- Bridgeman, B., & Cho, Y. (2012). Relationship of TOEFL iBT<sup>®</sup> to academic performance: Some evidence from American universities. *Language Testing*, 29(3), 421-442. doi:10.1177/0265532211430368
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20(1), 1-25. doi:10.1191/0265532203lt242oa
- Brown, A. (2006). An examination of the rating process in the revised IELTS Speaking Test (Research Report Vol. 6). Retrieved from IELTS Research Report website: <http://ielts.org/pdf/Volume%206,%20Report%202.pdf>
- Brown, A., & Hill, K. (1998). Interviewer style and candidate performance in the IELTS oral interview (Research Report Vol. 1). Retrieved from IELTS Research Reports website: <http://ielts.org/pdf/Vol1Report1.pdf>

- Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, 28(2), 221–249. doi:10.1177/0265532210384253
- Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20, 19–27. doi: 10.1111/j.1745-3992.2001.tb00072.x
- Cohen, N., & Arieli, A. (2011). Field research in conflict environments: Methodological challenges and snowball sampling. *Journal of Peace Research*, 48(4), 423–435. doi:10.1177/0022343311405698
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124–130.
- Doe, C., et al. (2011). What has experience got to do with it? An exploration of L1 and L2 test takers' perceptions of test performance and alignment to classroom literacy activities. *Canadian Journal of Education*, 34(3), 68-85.
- Dooley, P. (2008). Language testing and technology: Problems of transition to a new era. *ReCALL*, 20(01), 21-34. doi:10.1017/S0958344008000311
- Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: What does the test-taker have to offer? *Language Testing*, 19(4), 347-368. doi:10.1191/0265532202lt235oa
- Elder, C., & Wigglesworth, G. (2003). An investigation of the effectiveness and validity of planning time in part 2 of the IELTS speaking test. *IELTS Research Reports*, 6, 1-28.

- Faircloth, C. A. (2012). After the interview: What is left at the end. In J.F. Gubrium, J.A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds). *The SAGE Handbook of interview research: The complexity of the craft* (2<sup>nd</sup> ed.) (pp. 269-277). New York, NY: Sage.
- Feast, V. (2002). The impact of IELTS scores on performance at university. *International Education Journal*, 3(4), 70-85.
- Fulcher, G., & Reiter, R. M. (2003). Task difficulty in speaking tasks. *Language Testing*, 20(3), 321-344.
- Gan, Z., Humphreys, G., & Hamp-Lyons, L. (2004). Understanding successful and unsuccessful EFL students in Chinese universities. *The Modern Language Journal*, 88(2), 229-244. doi:10.1111/j.0026-7902.2004.00227.x
- Gomm, R. (2004). *Social research methodology: A critical introduction*. New York, NY: Palgrave Macmillan.
- Gordon, D. (2004). "I'm tired. you clean and cook." Shifting gender identities and second language socialization. *TESOL Quarterly*, 38, 437-457. doi: 10.2307/3588348
- Green, A. (2005). EAP study recommendations and score gains on the IELTS Academic Writing test. *Assessing Writing*, 10(1), 44-60.
- Green, A. (2006). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing*, 11(2), 113-134. doi:10.1016/j.asw.2006.07.002
- He, L., & Shi, L. (2008). ESL students' perceptions and experiences of standardized English writing tests. *Assessing Writing*, 13(2), 130-149. doi:10.1016/j.asw.2008.08.001

- He, L., & Shi, L. (2012). Topical knowledge and ESL writing. *Language Testing*, 29(3), 443-464. doi:10.1177/0265532212436659
- Hill, K., Storch, N., & Lynch, B. (1999). A comparison of IELTS and TOEFL as predictors of academic success (Research Report Vol. 2) Retrieved from IELTS Research Report website: <https://www.ielts.org/pdf/Vol2Report3.pdf>
- Hill, Y. Z., & Liu, O. L. (2012). Is There Any Interaction Between Background Knowledge and Language Proficiency That Affects TOEFL iBT® Reading Performance. Retrieved from TOEFL iBT Research Report website: <http://144.81.87.152/Media/Research/pdf/RR-12-22.pdf>
- Hirsh, D. (2007). English language, academic support and academic outcomes: A discussion paper. *University of Sydney papers in TESOL*, 2(2), 193-211.
- Horn, C. (2003). High-stake testing and students: stopping or perpetuating a cycle of failure?. *Theory into Practice*. 42(1), 30- 41.
- Hruska, B. L. (2004). Constructing gender in an English dominant kindergarten: Implications for second language learners. *TESOL Quarterly*, 38, 459-485. doi:10.2307/3588349
- Hsu, L. (2012). Applications of generalizability theory to estimate the reliability of EFL learners' performance-based assessment: A preliminary study. *Educational Research*, 3(2), 145-154.
- Huhta, A., Kalaja, P., & Pitkänen-Huhta, A. (2006). Discursive construction of a high-stakes test: the many faces of a test-takers. *Language Testing*, 23: 326. doi:10.1191/0265532206lt331oa



- Johnston, L. G., & Sabin, K. (2010). Sampling hard-to-reach populations with respondent driven sampling. *Methodological Innovations Online*, 5(2), 38–48.  
doi:10.4256/moi.2010.0017
- King, K., & Ganuza, N. (2005). Language, identity, education, and transmigration: Chilean adolescents in Sweden. *Journal of Language, Identity & Education*, 4(3), 179-199. doi:10.1207/s15327701jlie0403\_1
- King, N., & Horrocks, C. (2010). *Interviews in qualitative research*. Thousand Oaks, CA: Sage.
- Klinger, D.A., & Luce-Kapler, R. (2007). Walking in their shoes: Students' perceptions of large-scale high-stakes testing. *The Canadian Journal of Program Evaluation*, 22(3), 29-52.
- Kohn, A. (2000). Burnt at the high stakes. *Journal of Teacher Education*, 51(4), 315–327.  
doi:10.1177/0022487100051004007
- Kokhan, K. (2012). Investigating possibility of using TOEFL scores for university ESL decision-making: Placement trends and effect of time lag. *Language Testing*, 29(2), 291–308. doi:10.1177/0265532211429403
- Kubota, R., & Lin, A. (2006). Race and TESOL: Introduction to concept and theory. *TESOL Quarterly*, 40 (3), 471-493. doi:10.2307/40264540
- Kumagai, K. (2013). *How accent and identity influence each other: An investigation of L2 English speakers' perceptions of their own accents and their perceived social identities*. (Unpublished master's thesis). Indiana University of Pennsylvania, Indiana, PA.

- Lambert, M. (2012). *A beginner's guide to doing your education research project*.  
Chennai, India: Sage.
- Lee, S. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, 57(1), 87-118.
- Leibowitz, B. (2005). Learning in an additional language in a multilingual society: A South African case study on university-level writing. *TESOL Quarterly*, 39, 661–681. doi:10.2307/3588526
- Leung, C., & Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly*, 40(1), 211–234.  
doi:10.2307/40264517
- Li, H., Zhong, Q., & Suen H. K. (2012). Students' perceptions of the impact of College English Test. *Language Testing in Asia*, 2(3), 77–94.
- Lichtman, M. (2009). *Qualitative research in education: A user's guide* (2<sup>nd</sup> ed.).  
Thousand Oaks, CA: Sage.
- Luxia, Q. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142–173.  
doi:10.1191/0265532205lt300oa
- Macpherson, S. (2005). Negotiating language contact and identity change in developing Tibetan-English bilingualism. *TESOL Quarterly*, 39, 585–607.  
doi:10.2307/3588523

- McNeil, L. (2011). Investigating the contributions of background knowledge and reading comprehension strategies to L2 reading comprehension: An exploratory study. *Reading Writing, 24*, 883-902.
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Fransico, CA: Jossey-Bass.
- Merrylees, B. (1999). An investigation of speaking test reliability with particular reference to examiner attitude to the speaking test format and candidate/examiner discourse produced (Research Report Vol. 2). Retrieved from IELTS Research Reports website: [http://www.ielts.org/PDF/Vol2\\_Report1.pdf](http://www.ielts.org/PDF/Vol2_Report1.pdf)
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research, 45*(1-3), 35-44.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256. doi:10.1177/026553229601300302
- Mickan, P., Slater, S. & Gibson, C. (2000). A study of response validity of the IELTS writing subtest (Research Report Vol. 3). Retrieved from IELTS Research Reports website: [http://www.ielts.org/pdf/Vol3\\_Report2.pdf](http://www.ielts.org/pdf/Vol3_Report2.pdf)
- Morita, N. (2004). Negotiating participation and identity in second language academic communities. *TESOL Quarterly, 38*, 573–603. doi: 10.2307/3588281
- Morse, J. M. (2012). The implication of interview type and structure in mixed method designs. In J.F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds). *The SAGE Handbook of interview research: The complexity of the craft* (2<sup>nd</sup> ed.) (pp. 193–204). New York, NY: Sage.

- Murray, J. C., Riazi, A., & Cross, J. L. (2012). Test candidates' attitudes and their relationship to demographic and experiential variables: The case of overseas trained teachers in NSW, Australia. *Language Testing*, 29(4), 577-595.  
doi:10.1177/0265532212440690
- Nakatsuhara, F. (2011). Effect of test-taker characteristics and the number of participants in group oral test. *Language Testing*, 28(4), 483–508.  
doi:10.1177/0265532211398110
- Noubandegani, P.A. (2012). Students' perceptions of computerized TOEFL test. *Language Testing in Asia*, 2(2), 73–101.
- O'Loughlin, K. (2000). The impact of gender in the IELTS oral interview. *IELTS Research Reports*, 3, 1-28.
- O'Sullivan, B. (2012). A brief history of language testing. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff (Eds.) *The Cambridge guide to second language assessment* (pp. 9–19), Cambridge, UK: Cambridge University press.
- O'Sullivan, B., & Lu, Y. (2006). The impact on candidate language of examiner deviation from a set interlocutor frame in the IELTS Speaking Test (Research Report Vol. 6). Retrieved from IELTS Research Reports website:  
[http://www.ielts.org/PDF/Vol6\\_Report4.pdf](http://www.ielts.org/PDF/Vol6_Report4.pdf)
- Perry, J. F. (2005). *Research in applied linguistics*. New York, NY: Routledge.
- Poland, B. D. (2002). Transcription quality. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context & method* (pp. 629-649). Thousand Oaks, CA: Sage.

- Puspawati, I. (2012). *EFL/ESL (English as Foreign/Second Language) students' perceptions toward the TOEFL (Test of English as a Foreign Language) test*. (Unpublished master's thesis). Indiana University of Pennsylvania, Indiana, PA.
- Rajadurai, J. (2010). "Malays are expected to speak Malay": Community ideologies, language use and the negotiation of identities. *Journal of Language, Identity & Education*, 9(2), 91-106. doi:10.1080/15348451003704776
- Rasti, I. (2009). Iranian candidates' attitude towards IELTS. *The Asian EFL Journal Quarterly*, 11(3), 110–155.
- Sadeghi, K. (2010). Cloze validation against IELTS Reading Paper: Doubts on correlational validation. *Journal of English Teaching and Learning*, 53, 217.
- Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: A multiple data source approach. *Language Testing*, 17(1), 85-114.
- Sawaki, Y., & Nissan, S. (2009). Criterion-related validity of the TOEFL iBT listening section. TOEFL iBT (Research Report Vol.08). Retrieved from ETS Research Reports website: <http://www.ets.org/Media/Research/pdf/RR-09-02.pdf>
- Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. *Language Testing*, 26(1). 005–030.  
doi:10.1177/0265532208097335
- Sawaki, Y. 2001: Comparability of Conventional and Computerized Tests of Reading in a Second Language. *Language Learning & Technology*, 5(2), 38–59.
- Scott, G., & Garner, R. (2013). *Doing qualitative research: Design, methods, and techniques*. New Jersey, NJ: Pearson.

- Shaw, S. (2006). IELTS Writing: Revising assessment criteria and scales (Conclusion) (Research Notes Vol. 24). Retrieved from Cambridge English Research Notes website: <http://www.cambridgeenglish.org/images/22627-research-notes-24.pdf>
- Shohamy, E. (2001a). *The power of tests: Critical perspective on the uses of language tests*. Essex, London: Pearson Education.
- Shohamy, E. (2001b). Democratic assessment as an alternative. *Language Testing*, 18(4), 373–391.
- Shohamy, E. (2004). Assessment in multicultural societies: Applying democratic principles and practices to language testing. In B. Norton, & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 72–92). Cambridge, Cambridgeshire: Cambridge University.
- Sulaiman, T. (2012). A Review on IELTS Writing Test, Its Test Results and Inter Rater Reliability. *Theory and Practice in Language Studies*, 2(1), 138-143.
- Teemant, A. (2010). ESL student perspective on university classroom testing practices. *Journal of the Scholarship of Teaching and Learning*, 10(3), 89–105.
- Templer, B. (2004). High-stakes testing at high fees: Notes and queries on the international English proficiency assessment market. *Journal for Critical Education Policy Studies*, 2(1). Retrieved from <http://www.jceps.com/print.php?articleID=21>
- Terry, M. (2003). IELTS preparation materials. *ELT Journal*, 57(1), 66-76.
- Triplett, C. F., Barksdale, M. A., & Leftwich, P. (2003). Children’s perceptions of high stakes testing. *Journal of Research in Education*, 13(1), 15-21.

- Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high-stakes testing. *Journal of Literacy Research, 37*(2), 273-260.
- Tsai, Y., & Tsou, C.-H. (2009). A standardised English language proficiency test as the graduation benchmark: Student perspectives in higher education. *Assessment in Education: Principles, Policy, & Practice, 16*(3), 319-330.
- Uysal, H. H. (2010). A critical review of the IELTS writing test. *ELT Journal, 64*(3), 314-320.
- Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing 30*(1), 49–70. doi:10.1177/0265532212442634
- Wall, D. (2000). The impact of high-stake testing on teaching and learning: Can this be predicted or controlled? *System, 28*, 499–509.
- Weigle, C.S. (2010). Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability. *Language Testing, 27*(3), 335–353.  
doi:10.1177/0265532210364406
- Winke, P., Gass, S., & Myford, C. (2012). Raters' L2 background as potential source of bias in rating oral performance. *Language Testing, 30*(2), 231–252.  
doi:10.1177/0265532212456968
- Wise, L. L., et al. (2004). Independent evaluation of the California high school exit examination (CAHSEE) (Research Report vol.1). Retrieved from Human Resources Research Organization website:  
<http://www.cde.ca.gov/ta/tg/hs/documents/biennial04.pdf>

- Xi, X. (2010). How do we go about investigating test fairness. *Language Testing*, 27(2), 147-170. doi:10.1177/0265532209349465
- Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Model. *Language Testing*, 30(1), 49-70. doi:10.1177/0265532212442634
- Zahedi, K., & Shamsaee, S. (2012). Viability of construct validity of the speaking modules of international language examinations (IELTS vs. TOEFL iBT): evidence from Iranian test-takers. *Educational Assessment, Evaluation and Accountability*, 24(3), 263-277.
- Zhang, H. (2010). *Assessing test Reliability: Comparing Two Versions of Reading Comprehension Testing the TOEFL test*. (Doctoral dissertation). Kristianstad University, Kristianstad.
- Zhengdong, G. (2009). IELTS preparation course and student IELTS: A case study in Hong Kong. *RELC Journal*. 40(1), 23-40. doi:10.1177/0033688208101449



## APPENDIX A

### IRB Approval



Indiana University of Pennsylvania  
[www.iup.edu](http://www.iup.edu)

Institutional Review Board for the  
Protection of Human Subjects  
School of Graduate Studies and Research  
Stright Hall, Room 113  
210 South Tenth Street  
Indiana, Pennsylvania 15705-1048

P 724-357-7730  
F 724-357-2715  
[irb-research@iup.edu](mailto:irb-research@iup.edu)  
[www.iup.edu/irb](http://www.iup.edu/irb)

October 22, 2013

Hartati Suryaningsih  
686 South 5th Street  
Indiana, PA 15701

Dear Ms. Suryaningsih:

Your proposed research project, "Students' Perceptions of IELTS (International English Language Testing System) and TOEFL (Test of English as Foreign Language)," (Log No. 13-265) has been reviewed by the IRB and is approved. In accordance with 45CFR46.101 and IUP Policy, your project is exempt from continuing review.

It is also important for you to note that IUP adheres strictly to Federal Policy that requires you to notify the IRB promptly regarding:

1. any additions or changes in procedures you might wish for your study (additions or changes must be approved by the IRB before they are implemented),
2. any events that affect the safety or well-being of subjects, and
3. any modifications of your study or other responses that are necessitated by any events reported in (2).

The IRB may review or audit your project at random *or* for cause. In accordance with IUP Policy and Federal Regulation (45CFR46.113), the Board may suspend or terminate your project if your project has not been conducted as approved or if other difficulties are detected.

Although your human subjects review process is complete, the School of Graduate Studies and Research requires submission and approval of a Research Topic Approval Form (RTAF) before you can begin your research. If you have not yet submitted your RTAF, the form can be found at <http://www.iup.edu/page.aspx?id=91683>.

I wish you success as you pursue this important endeavor.

Sincerely,

A handwritten signature in blue ink that reads "J. Mills".

John A. Mills, Ph.D., ABPP  
Chairperson, Institutional Review Board for the Protection of Human Subjects  
Professor of Psychology

JAM:jeb

Cc: Dr. David Hanauer, Thesis Advisor  
Ms. Brenda Boal, Secretary

## APPENDIX B

### RTAF Approval



Indiana University of Pennsylvania  
[www.iup.edu](http://www.iup.edu)

Office of Assistant Dean for Research  
School of Graduate Studies and Research  
Stright Hall, Room 113  
210 South Tenth Street  
Indiana, Pennsylvania 15705-1048

P 724-357-7730  
F 724-357-2715  
[www.iup.edu/research](http://www.iup.edu/research)

October 22, 2013

Hartati Suryaningsih  
686 South 5<sup>th</sup> Street  
Indiana, PA 15701

Dear Ms. Suryaningsih:

Now that your research project has been approved by the Institutional Review Board for the Protection of Human Subjects, I have reviewed your Research Topic Approval Form and approved it.

**The Thesis/Dissertation Manual, additional resources, and information to help you start writing can be found at <http://www.iup.edu/graduatestudies/thesis/default.aspx>.**

Your RTAF indicates your anticipated graduation date as May 2014. This means that if your program requires a defense, you must defend **no later than April 1, 2014** and all necessary documents are due by this date. A description of the required documents can be accessed at <http://www.iup.edu/page.aspx?id=116439>. Your thesis or dissertation must be submitted to the School of Graduate Studies & Research by April 15, 2014 if you desire to graduate by your anticipated date. You must apply for graduation by May 1, 2014. For deadlines for subsequent graduation dates, please access <http://www.iup.edu/page.aspx?id=16683>.

You are now eligible to receive a FREE copy of Adobe Professional! This software will help you to create an electronic thesis or dissertation. Attached is a copy of the Adobe Agreement form that you need to print, fill out, and sign. Once you have completed the form, you can take it to the IT Support Center in G35 Delaney Hall to obtain a copy of the software. If you are not able to come to campus, you will need to scan the completed form and send it as an attachment from your IUP e-mail account to [it-support-center@iup.edu](mailto:it-support-center@iup.edu). You can also fax the completed form to 724-357-4983. Please indicate in your email or fax that you are a graduate student requesting the Adobe Professional software and include your Banner ID and mailing address so the software can be mailed to you.

Finally, if you change your topic, the scope or methodology of your project, or your committee, a new Research Topic Approval Form must be completed.

I wish you well and hope you find this experience to be rewarding.

Sincerely,

A handwritten signature in black ink, appearing to read 'Hillary E. Creely', is written over a circular stamp.

Hillary E. Creely, J.D., Ph.D.  
Assistant Dean for Research

xc: Dr. Yaw Asamoah, Dean  
Dr. Gloria Park, Graduate Coordinator  
Dr. David Hanauer, Thesis Chair  
Ms. Julie Bassaro, Secretary

HEC/bb

## APPENDIX C

### Informed Consent Form

TITLE: Students' Perceptions of IELTS (International English Language Testing System) and TOEFL (Test of English as a Foreign Language) tests

You are invited to participate in this research study. The following information is provided in order to help you to make an informed decision whether or not to participate. If you have any questions, please do not hesitate to ask. You are eligible to participate because you are a second language learner of English, an undergraduate or a graduate student who has or pursues a degree in TESOL/TEFL/TESL/ Applied Linguistics, and have taken both IELTS and TOEFL tests. This study attempts to gain insights into students' perceptions of IELTS and TOEFL tests.

You will be asked to have an interview with the researcher by using Skype. The interview will take approximately 45 minutes where you will be asked some questions addressing the purpose of the study. In the interview, you have rights and power not to answer certain questions which may cause you to feel uncomfortable. This interview will be digitally audio-recorded so that I can keep track of our conversation. Therefore I ask your permission to record our interview. I will transcribe the interview for analysis. In the transcript, I will use pseudonym in lieu of your real name in order to protect your identity. You may request to see these sections if you wish.

Your participation in this study is **voluntary**. You are free to decide to participate in this study or to withdraw at any time without adversely affecting your relationship with the researcher. If you choose to participate, you may withdraw at any time by notifying me through E-mail, phone, or mail. Upon your request to withdraw, all information pertaining to you will be destroyed. If you choose to participate, all information will be held in strict confidence. The information obtained in the study may be published in scientific journals or presented at scientific meetings but your identity will be kept strictly confidential. If you are willing to participate, please sign the statement below. Thank you.

**Researcher:**  
**Hartati Suryaningsih**  
**M.A. Candidate**  
**TESOL Program**  
**English Department**  
**Indiana University of Pennsylvania**  
**Indiana, PA 15705, U.S.A**  
**Phone: 724-467-0740**  
**E-mail: [zrws@iup.edu](mailto:zrws@iup.edu)**

**Thesis Advisor:**  
**Dr. David I. Hanauer**  
**Professor**  
**Composition and TESOL**  
**English Department**  
**Indiana University of Pennsylvania**  
**Indiana, PA 15705, U.S.A**  
**Phone: 724-357-2274**  
**E-mail: [Hanauer@iup.edu](mailto:Hanauer@iup.edu)**

**This project has been approved by the Indiana University of Pennsylvania Institutional Review Board for the Protection of Human Subjects (Phone: 724-357-7730)**

**Informed Consent Form (continued)**

**VOLUNTARY CONSENT FORM:**

**I have read and understand the information on the form and I consent to volunteer to be a subject in this study. I understand that my responses are completely confidential and that I have the right to withdraw at any time. I have received an unsigned copy of this informed Consent Form to keep in my possession.**

Name (please print): \_\_\_\_\_

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

**I certify that I have explained to the above individual the nature and purpose, the potential benefits, and possible risks associated with participating in this research study, have answered any questions that have been raised, and have witnessed the above signature.**

\_\_\_\_\_

**Date**

\_\_\_\_\_

**Investigator's Signature**

## APPENDIX D

### E-mail Protocol

Dear -----,

My name is Hartati Suryaningsih, and I am a Master candidate majoring in TESOL (Teaching English to Speakers of Other Languages) at Indiana University of Pennsylvania. I am currently working on my thesis focusing on students' perceptions of IELTS (International English Language Testing System) and TOEFL (Test of English as Foreign Language) tests.

You are invited to participate in this research study. You are eligible to participate because you are a second language learner of English, an undergraduate or a graduate student who has or pursues a degree in TESOL (Teaching English to the Speaker of Other Languages)/TESL (Teaching English as a Second Language)/TEFL (Teaching English as a Foreign Language)/Applied Linguistics, and have taken both IELTS and TOEFL tests.

You will be asked to have an interview with me by using Skype. The interview will take approximately 45 minutes where you will be asked some questions addressing the purpose of the study.

If you are interested in participating in this study which can be the way for you to share your experiences and perceptions of both tests as well as contributing to research on language testing, please contact me by replying this email.

Following your email response, I will explain this research in greater detail; therefore you can decide whether or not to participate in this study.

Thank you for your time and consideration.

Sincerely,

Hartati Suryaningsih  
M.A. Candidate  
TESOL Program  
English Department  
Indiana University of Pennsylvania  
Indiana, PA 15705, U.S.A  
Phone: 724-467-0740  
E-mail: [zrws@iup.edu](mailto:zrws@iup.edu)

## APPENDIX E

### Interview Protocol

#### TOEFL test

I would like you to think carefully about the TOEFL test. Please try to remember details of doing the test.

1. Could you please make a list of 5 things that characterize your experience of TOEFL test?
2. Could you tell me about each one and describe it as detail as you can?
3. What would you say about the TOEFL test? Please provide positive and negative impressions.

#### IELTS test

I would like you to think carefully about the IELTS test. Please try to remember details of doing the test.

4. Could you please make a list of 5 things that characterize your experience of IELTS test?
5. Could you tell me about each one and describe it as much detail as you can?
6. What would you say about IELTS test? Please provide positive and negative impressions.

#### Both IELTS and TOEFL tests

7. Based on your perception, how are these two tests different? Or how are these two tests similar?
8. Would you like to tell me which test is fairer? Could you explain why?
9. Would you like to tell me which test do you prefer? Could you explain why?