

In the name of God



Shahrood University of Technology

English Language Department

M.A. Thesis in Language Teaching

Exploring PhD Candidates' Attitude towards TOLIMO

By: Arya Javanmardi

Supervisor:

Dr. Seyyed Ali Ostovar-Namaghi

January, 2018

Dedication

I dedicate this thesis to Professor Dr. Ostovar-Namaghi, my family and my classmates for their help during the fulfillment of my thesis.

Acknowledgment

I would first like to thank my Professor Seyed Ali Ostovar-Namaghi who has instructed me during the writing of this thesis. Also, I would like thank my family who help to complete my education. Finally, I would like to thank participants who allocated time for multiple interview sessions.

تعهد نامه

اینجانب آریا جوانمردی دانشجوی دوره‌ی کارشناسی ارشد رشته‌ی آموزش زبان انگلیسی دانشکده‌ی مهندسی صنایع و مدیریت دانشگاه صنعتی شاهرود نویسنده‌ی پایان‌نامه‌ی **Exploration of PhD Candidates' Attitudes towards TOLIMO**

تحت راهنمایی جناب آقای دکتر سید علی استوار نامقی متعهد می‌شوم.

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ‌جا ارائه نشده است.
- کلیه‌ی حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام «دانشگاه صنعتی شاهرود» و یا «Shahrood University of Technology» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در بدست آوردن نتایج اصلی پایان‌نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه‌ی مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه‌ی مراحل انجام این پایان‌نامه، در مواردی که به حوزه‌ی اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ ۱۳۹۶/۱۱/۱۱

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه‌ی حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون

Abstract

This study aims to explore the attitudes of PhD candidates of TOLIMO in Iran. TOLIMO is a high-stake test in Iran that is used for PhD completion. This qualitative study uses case study for data collection and data analysis. There are different English exams in Iran, but the evidences show that there are not enough explorations about the quality of TOLIMO. The participants of the study are selected by purposive sampling then the investigator takes an unstructured open-ended interview to achieve useful information and code. The data for this study were collected through an unstructured open ended interview with the informant of the study. There are three main propositions in this study: Positive attitude towards TOLIMO, Negative attitude towards TOLIMO and suggestion for improvement. Policy makers by the results of this study can improve the reading and listening section of the TOLIMO. They can use better facilities for holding the test and decrease the length of the reading parts.

Key words: attitude, TOLIMO, PhD candidate, case study, reading, listening

Table of Contents

1. Chapter One: Introduction	
1.1. Overview.....	2
1.2. Statement of the Problem Problem.....	2
1.3. Significance of the Study.....	3
1.4. Purpose of the Study.....	3
1.5. Limitations of the Study.....	3
1.6. Delimitation of the Study.....	4
1.7. Ethical Issues.....	4
2. Chapter Two: Review of the Related Literature	
2.1. Introduction.....	6
2.2. Theoretical Perspective.....	6
2.2.1. Aspects and Definition of Attitude.....	6
2.2.2. Relationship between attitude and Language Learning.....	7
2.2.3. Different Types of Tests.....	8
2.2.4. Relationship between Test and Test Takers.....	9
2.3. Empirical Findings.....	12
2.4. Implication for Practice.....	15
2.5. Summary of Empirical Findings.....	15
2.6. Statement of the Gap.....	16
3. Chapter Three: Methodology	
3.1. Overview.....	18
3.2. Research Design.....	24
3.3. Sampling Procedure.....	24
3.4. Participants.....	26
3.5. Data Collection.....	27
3.6. Data Analysis.....	30

4. Chapter Four: Results	
4.1. Overview.....	34
4.2. Positive Attitudes towards TOLIMO.....	34
4.2.1. Sufficiency of the Time.....	34
4.2.2. Evaluation of the Appropriateness of the Test.....	35
4.3. Negative Attitude towards TOLIMO.....	35
4.4. Suggestions for Improvement.....	45
5. Chapter Five: Discussion and Conclusion	
5.1. Overview.....	48
5.2. Discussion.....	48
5.3. Implication for Practice.....	51
5.4. Suggestion for Further Studies.....	51
6. References.....	52

CHAPTER
ONE
INTRODUCTION

Chapter one

Introduction

1.1. Overview

This study tries to investigate the attitudes of PhD students toward TOLIMO in Iran. TOLIMO is a high-stake nation-wide test administered as a pre-requisite for the completion of the PhD program in Iran. By finding the attitudes of PhD students, policymakers can modify the structure and content of the TOLIMO exam. Therefore, they can omit or add some sections in this type of examination. Despite its importance, PhD candidates' attitude towards standardized proficiency tests such as TOLIMO has rarely been explored. Many researchers study about the TOEFL and the IELTS, but there is a lack of study about the TOLIMO, which is a popular instrument for assessing the English knowledge of PhD students in Iran. Policymakers do not pay attention to the opinion of PhD students whether TOLIMO is suitable for assessment or not.

Attitude is defined as a disposition or inclination to answer positively or negatively towards a certain thing like an idea, object, person, or situation. Students have positive or negative attitudes towards the language they want to learn or the people who speak it. Positive attitude is one of the reasons that students do better on the exams (Poorsoti & Davatgari, 2000). Moreover, a large number of studies investigate the relationships between attitude and proficiency in the language (Bachman, 1990; Malallaha, 2000; Coleman, Strafield, & Hagan, 2003). Similarly, Gardner (1985) believes that attitude and other affective variables are as important as aptitude for language achievement.

1.2. Statement of the Problem

There are many English exams which are used to assess the English ability of students in schools, institutes and universities, one of them is TOLIMO. PhD students must pass this

high- stake nation-wide exam to continue their education and it is pre-requisite for higher education. However, many PhD students cannot pass TOLIMO, they complain about the quality and validity of this test.

1.3. Significance of the Study

The findings of the study can help policymakers to set a new style in foreign language testing. Also, the educational system in Iran can use the findings to improve the quality of similar exams like MSRT. The findings can be useful for candidates who try to pass the TOEFL and IELTS in their educations. The other reason of conducting this study is to change the traditional way of testing in Iranian institutes, schools and universities. The last reason of this study is changing the educational materials that used for English exams preparation.

1.4. Purpose of the Study

This study aims to investigate PhD candidates' attitude towards TOLIMO, a high-stake nation-wide test administered as a pre-requisite for the completion of the PhD program. The general question, "What are your perceptions of TOLIMO? Would you please explain its strengths and weaknesses?" will be used not only to elicit theoretically relevant data but also to direct this study. This research is qualitative one: as a result, there isn't any hypothesis. Subsequent questions were taken from an initial analysis of the data collected from the perspectives of two qualified participants.

1.5. Limitations of the Study

Although there are many people who have participated in TOLIMO but the researcher of this study is not able to get access to such cases. There are different opinions about TOLIMO that can be expressed by many participants of this study but this study just cover the small group.

Also, this study is not information rich enough to achieve to its main purpose because of involving only the one university in Iran to get the meanings out of their experiences.

1.6. Delimitations of the Study

To delimit some limitations of the study, the PhD candidates who attend more than one time in the exam is interviewed. Also, this study tries to cover the opinions of the participants who have both negative and positive opinion about the TOLIMO. I try to interview with the PhD candidates who are accessible for further questions. For improving the quality of the data, the researcher tries to interview with the experienced PhD candidates to narrow down the limitations. The candidates of the study are chosen by the suggestion of the PhD students who attended in TOLIMO.

1.7. Ethical Issues

To address ethical issues, the researcher developed a specific informed consent 'agreement', in order to gain the informed consent from participants which is based on the recommended items by Bailey (1996) and Arksey and Knight (1999). The items concern informing the participants about the purpose of the research, the procedures of the research, the risk and benefits of the research, the voluntary nature of research participation, and the subject's right to stop the research at any time. Generally, these procedures are used to protect confidentiality. Moreover, the participants were endured that they can withdraw any stages of the study.

Chapter Two

Review of the Related Literature

Chapter Two

Review of the related Literature

2.1. Introduction

TOLIMO is one of the exams held in Iran for PhD students. There are different definitions for this type of exam but according to the ministry of science, research, and technology of the Islamic Republic of Iran, TOLIMO is a nationwide large-scale English proficiency test that is held by the Iranian testing organization. There are different exams that are held in Iran and other countries like MSRT, TOEFL, IELTS, etc. However, the evidences show that there is not enough exploration about the quality of TOLIMO and the opinion of the participants. A good test has two aspects: reliability, that means the precision with which a test measures what it is supposed to measure and validity, i.e., if the test really measures what it is supposed to measure. (Bachman, 1990; Walt & Steyn, 2008; Farhady, Jafarpur, & Birjandi, 2006; Nodoushan, 2009). High-stakes tests are those whose results are seen by students, teachers, administrators, parents, or the public as the basis upon which important decisions are made that immediately and directly affect the test-takers and other stakeholders (Salehi, Karimvand, & Naeeni, 2011). The primary use of high-stakes tests has been “to ration future opportunity as the basis for determining admission to the next layer of education or to employment opportunities” (Chapman & Snyder, 2000: 458).

The present study has tried to investigate the topic of exploring PhD candidates' attitude towards TOLIMO by using: 1) establish a theoretical framework in describing the attitudes toward language learning and testing 2) present empirical findings to support the thesis statements and 3) present implications for practice.

2.2. Theoretical Perspectives

2.2.1. Definition and Aspects of Attitude

There are different definitions for attitude. According to Oxford and Shearin (1994) attitude is one factor that influences on motivation in learning of a language. Attitude, according to Allport (1971, p. 13), refers to “a mental and neural state of readiness organized by experience, using a directive or dynamic influence on individual's response to all objects and

situations in which it is related”. In a similar definition, Rasti (2009, p. 111) says that attitude was defined as “a disposition or tendency to respond positively or negatively towards a certain thing such as an idea, object, person or situation”. Likewise, attitude is defined as ‘a psychological status that is expressed by some degree of favor or disfavor’ (Eagly & Chaiken, 1993, p. 1).

There are some definitions of the different aspects of attitude. For instance, Baker (1992) recognizes three aspects of attitude: affective, cognitive, and conative, and classified the five important features of attitudes, as follow: (1) attitudes are cognitive and affective; (2) attitudes are dimensional rather than bilateral; (3) attitudes incline a person to behave in a certain way, but the connection between attitudes and behaviors is not a tough; (4) attitudes are learnt, not innate and (5) attitudes try to persevere but they can be improved by experience. In another study, Murray et al. (2012), pointed out that ‘attitude’ consists of three main components: the first one is belief (the intention that may true or not), opinion (the actual or theoretical actions that should or should not happen), and emotions (feelings about different matters).

2.2.2. Relationship between attitude and Language Learning

There are studies that define the relationship between attitude and motivation in language learning. For example, Oxford and Shearin (1996) recognizes six features that influence motivation in language learning: attitudes, beliefs about self, goal, involvement, environmental support and personal attributes.

Some of the researchers pay attention to the attitude about language and language learning. For example, Holmes (1992) believes those people develop their attitudes towards languages that reflect their opinions about persons who speak the languages, and the conditions and roles with which they are related. Furthermore, attitudes vary from different aspects like anxiety about the language and the situation of language learning, or attitudes of learners about speakers of the L2, the country or city in which it is spoken, the classroom, the teacher, other learners, particular aspects in the learning activities, exams and views about learning (Johnson & Johnson, 1998).

There are some of the theories about the relationship between the language learning, motivation and attitude of learners. Brown (2000) believes that positive attitudes are beneficial for second language learners and negative attitudes decreased motivation. Similarly, Gardner (1985) believes that attitude and aptitude are important for language achievement. For some of the specialists, learners' attitudes, motivation, beliefs and perceptions of language learning and teaching are valuable. (Cotterall, 1999; Gardner, 1983; Gardner and MacIntyre, 1993; Nunan, 1998; O'Neil, 1991). Some other researchers have similar idea, for example, Finocchiaro (1982) believes that factors like positive learner and teacher attitudes have important role in testing and language learning. Moreover, Holmes (1992) says that if people have a positive feeling about persons who use the language, they would be more highly motivated toward learning the language. Also, Van Lier (1996) in his studies claims that interesting and meaningful expressions of language, improve motivation and positive attitudes to language and language learning. Truitt (1995) theorized that the beliefs of students and their attitudes about learning of new language might vary according to cultural background and previous experiences.

2.2.3. Different Types of Case Study

There are various types of tests. For example, traditional English tests have the fixed formats. Ory (2004) says that traditional form of English language tests are fixed response format with one correct answer. It means that, it is objective. Objective items require students to choose the correct items from several options or fill the blank with a word or short phrase to answer a question or complete a statement.

The aims of using the tests are diverse. Bachman and Purpura (2008) claim that language tests have been used in many cases like the selection of students, giving certifications, and giving promotion. Similarly, Zahedi and Shamsaee (2012) say that language tests open or close opportunities for the test takers' dreams (Zahedi & Shamsaee, 2012). Someone believe that assessment and testing are different. For instance, Leung and Lewkowicz (2006) believes that assessment is a super-ordinate term that used for all types of assessment, while testing means one particular form of assessment.

In many schools, universities and institutes exams are used as a kind of evaluation system that tries to evaluate the knowledge of students. In many countries, language tests are one of the requirements to enter different levels of education or professional life. In other words, language tests used as the gatekeeper (Bridgeman & Cho, 2012). There are many reasons that government and ministry of education in different countries use tests for assessment. For instance, McNamara (2004) stated that because of national standards in educational achievement, some of the countries use tests. As a result, there are different types of exams. For example, competitive examinations refer to two thousand years ago and carry remaining effects on contemporary assessment practices (Carless, 2011). There is high-stakes tests that used in many institutes. Luxia (2005) argued that “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). According to Shohamy (2001a) tests are able to make stakeholders believe that the tested knowledge is the most important knowledge.

2.2.4. Relationship between Test and Test Takers

Some researchers pay attention to the influence of tests on test takers. It has been recognized for many years that assessment has an intense effect on how students experience school and how they identify themselves as learners (Shohamy, 2007). So, the power of tests encourages teachers to teach the tests. Also, test preparation evokes feelings of fear, unfairness, bias, pressure and doubt in students (Shohamy, 1998). In the similar study, Kirkland (1971) concluded that tests could influence factors like student’s self-concept, motivation, goal, study practices, and anxiety. Shohamy (2001a) argues that considering the opinions of test takers about exams direct us to a better understanding about exams specifically high-stake tests. According to Murray, Riazi and Cross (2012) negative attitudes of test takers were personal experiences and feelings of test takers and influence of people. On the other hand, according to Vallerand and Reid (1984) motivation can be enhanced after the exam if the student receives positive feedback. The similar studies try to find out the relationship between tests and performances of test takers. For example, Oxford and Shearin (1994) say that good performances on exams can cause the sense of success. Likewise, As Shohamy (1993) claimed that using tests for authority and controlling the students has led to top-down

educational change strategies. It means that tests decide what students should do. This process can be considered unethical, undemocratic and unbeneficial for test takers.

Researchers have tried to realize why test takers have bad performances on tests. Generally, an investigation of English as a foreign language (EFL) students' problems with English tests is limited (Zeidner & Bensoussan, 1988). Fillipi (2012) is one of the researchers who has done research about factors that cause weak performances of test takers on tests. Some educationalists focused on standardized tests like TOEFL (Test of English as a Foreign Language) and IELTS (International English Language Testing System) (Wait & Gressel, 2009; Aboudan, 2011; Wolf & Steinberg, 2011). There are scientists who concluded that expectations of test takers affect the performances of students in exams (Onwuegbuzie, Baiey, & Daley, 2001). Bachman (1990) believes that the performances of test takers in exams are affected by factors like familiarity of the place and the equipment that used, staffs, time of the test, physical conditions and personal factors such as age, gender and educational background. Some of the problems of test takers were the inability to read the texts, lack of vocabulary knowledge both receptive and productive, weak word choice, the inability to understand questions and lack of knowledge in writing (Al-Buainain, 2006). Studies have shown that English proficiency affects students' performance on tests (Aydn et al., 2006; Teemant, 2010; Korhonen et al., 2012). Alike, Baker and Marshall (2002:1) say that lack of proficiency in language is one of the main reasons that international students cannot match their expectations and experiences. Some of the theorists study about the effect of anxiety in bad performance of students. For instance, Sarason and Sarason (1990) argued that test anxiety is caused by individuals' perception of the test as a form of pressure to do well. Read and Hayes (2003) reported that there is positive feeling about IELTS among teachers and tough motivation in learners. There are various opinions about the weakness of tests in evaluation of test takers. For example, Kohn (2000) believed that the scores of high-stake tests do not reveal the knowledge of students, but rather their skill in taking tests. Doe et al. (2011) investigated about language tests and perception of L1 and L2 test takers toward language tests. Test takers considered the test as unfair, bad, waste of time, and useless because the test did not evaluate their true ability in English. Scientists, researchers and educationalists interested in students' difficulties in learning a foreign language, because it enhances students' performance on tests. Some of the researchers examined teachers' proficiency and teachers' practice and proposed in-service training (Ho, 2008).

Some of the students and educationalists complain about the validity of the tests. The validity refers to the content of the test. According to Bachman (1990), there are two aspects that ensure the content is valid: content relevance and content coverage. It means that the content of the test should be relevant to the areas to be measured and coverage means the supporting of content or ability that purports. Another aspect of the test is fairness. According to Penfield and Camilli (2007) fairness refers to whether test marks measure same things among different groups or not. In other word, fairness refers to construct validity. Some of the educationalists studied about validity of the tests, for example, Shi and He (2012) examined the tests and their validity. Also, some of the educationalists like Teemant (2010) stated that there is not enough research about the students' actual knowledge and their test performance. Murray et al. (2012) pointed out test-takers' attitudes toward test may have significant implications to test validity. There are advantages and disadvantages of standardized testing, for students. Gottlieb (2006) provided a table which clearly shows the advantages and disadvantages of testing.

Advantages	Disadvantages
Procedure reliable data	Often serve as a gatekeeper
Insist on uniform data collection and analysis	Represent shallow curricula, not broad, in-depth knowledge
Provide a national perspective or ranking	Don't allow for student creativity or imagination
Are not confounded by individual teacher effects	Don't take different student learning styles into account
Often are readily scored	Adhere to time limits
Easy train teachers on procedures for administration	Lead to misinterpretation of data and overgeneralization of results

Table 1. **Advantages & Disadvantages of Testing**

2.3. Empirical Findings

Reviewing the theoretical perspective, makes clear the opinions of educationalists and policymakers about the attitudes of test takers and testing. Now, we review the actual teaching and learning. Many studies focus on the role of testing. For example, Uysal (2010) found that language tests are powerful tool in society. Tests have played important roles in test takers lives.

There are findings about the attitudes of test takers about testing. Alderson and Hamp-Lyons (1996) found that many teachers and students have negative attitudes to the test, they feel that tests don't show the reality. Triplett and Barksdale (2005) also did another research about perception of students towards tests. The findings of their study showed that the tests had negative effect on the emotions of students. Students stated their emotion and experiences negatively. In another study, Truitt (1995) studied about the attitude of language learners and found that student beliefs and attitudes about language learning may differ based on cultural background and prior experiences. Similarly, Lamb (2004) in his study about the attitude of language learners found that learners at the beginning of learning have positive attitude towards language but at the end their attitudes toward the experience of formal learning tended to weaken. Murray, Riazi, and Cross (2012) found that tests have negative effect on test-takers' emotion. The result of their study shows that test-takers had bad reactions towards the exams. There is a similar study about the reaction of students towards the tests. For example, Likewise, Triplett, Barksdale, and Leftwich (2003) found that students were worry whenever they talked about tests. Students were worry before taking the test, while they were taking the test, and after the end of the test. Also, the researchers found that fear of students of failing was in their thought when they were taking tests.

The empirical studies focus on the effect of attitude of test performance. Huhta, Kalaja, and Pitkanen-Huhta (2006) found that when the students got a bad score in his or her last exam, they are not happy about it. Therefore, while he or she would take the second test, he or she felt worried and stressed because of his or her failure in the past. They also found that the results of test can affect students' learning either positively or negatively. Similarly, Greene (1929) found that students' feelings of easiness and difficulty of the tests affect their performance.

Some of the researchers try to find out the relationship between attitude towards test and language learning. For instance, Malallah (2000) studied about Arab learner attitudes towards English and found that there are positive attitude towards English language and the proficiency of language learners in exams was related to the positive attitude to English.

Some of the studies try to pay attention to high-stake tests that is held in different countries. For instance, Zhao and Cheng's (2010) have studied about Chinese students' attitudes to the CET, and they identified four components to represent students' attitudes, including test-taking anxiety/lack of concentration, test-taking motivation, belief in CET, and test easiness. There is an example about attitude of language learners towards IELTS in Iran. Rasti (2009) worked on Iranian candidates' attitudes towards IELTS and the findings of his research show that nearly 80% of the participants have positive attitude towards IELTS. Also, there is not any meaningful relationship between Iranian candidates' sex, age, educational background, and IELTS scores with their attitudes towards IELTS.

There are empirical studies that try to find out the effect of standard tests like TOEFL and IELTS on students' performances. Some studies have not found any relationship between IELTS and students' performances (Fiocco, 1987; Graham, 1987; Light, Xu and Mossop, 1987). In the similar study, Kerstjens and Nery's (2000) studied about the relationship between the IELTS test and students' academic performance. They found that at the professional level, IELTS was a significant predictor of academic performance for students. In another study, Feast (2002) have tried to investigate the relationship between scores of IELTS as a language proficiency measurement and performance of students at university. She found an important and positive, but weak relationship between English language proficiency and academic performance.

Other studies focus on attitudes of test takers towards IELTS and TOEFL. For example, Stricker, Wilder, and Rock (2004) found that test takers have positive attitudes about TOEFL CBT in testing centers like Buenos Aires, Cairo, and Frankfurt. In 2007, Lewthwaite found that the nature and type of activities that used by UAE IELTS candidates for exam preparation was a main cause in the balance of positive and negative attitudes. There is a famous empirical study about the attitude of Iranian who participate in IELTS. Rasti (2009) found that candidates who have positive attitude towards IELTS do not differ from

other groups in their performance on the test. Rasti found that having a positive attitude towards IELTS does not secure attainment a good score.

Some of the researchers pay attention to the content of the standard tests and the problems of test takers in taking the high-stake tests. For example, Puspawati (2012) found that according to students, the topic in the TOEFL test is unfair. Similarly, Evans & Green (2007) studied the language problems of Cantonese-speaking students at Hong Kong and found that major percentage of the test takers have problems when studying content subjects in English. In addition they found that students' problems focus on academic writing and speaking. The different aspects of the findings reveal that students' receptive and productive vocabularies are not enough for them. Findings also show that students have fewer difficulties with academic listening than the other three skills. There is similar study reveals that students have problem in writing topics in the LPI (English Language Proficiency Index). This is because of writing topics (He & Shi, 2008).

The similar studies focus on the manner of testing and the facilities of the place of testing. For example, Sawaki (2001) found that the test-takers complained about the watching the screen of computer because of eye tiredness while test takers were reading the test. Puspawati (2012) found that the students could not focus on test because of the noise of air conditioner and the noise from the keyboard of the computer and other test takers on the speaking section. Also, there is a study that pay attention to the time of testing. Triplett & Barksdale (2005) found that many students were worried about the time of testing. Some students complained that tests were too long; while other students said tests were too short. There is a study about the performance of test takers on TOEFL computer based test. Jamieson, Taylor, Kirsch, and Eignor (1999) found that TOEFL computer-based test (CBT) increased test takers' acceptance of the test.

Other factors that affect the performance of test takers is the structure of test and its questions. For instance, He and Shi (2008, 2012) found that familiarity of students with structure of test will affect the performance of test takers. They found that vocabulary, cultural differences and context of the test are area of difficulty for test takers. Similarly, Bird and Welford (1995) wanted to explore the effect of using both modified and unmodified tests on native and non-native speakers and found differences between the two tests with non-native speakers. There is an empirical study in which Filipi (2012) found that the language

of items in test is an important factor in determining item difficulty. Also, he found that test format is another factor in item difficulty.

Some educationalists refer to lexical knowledge of test takers. Teemant (2010) who studied problems in testing found that vocabulary and grammar were one of the important problems in language testing. Ata (2010) found that poor lexical and cultural knowledge of English by Arab-speaking students can cause numerous negative interferences. His study suggests that IELTS preparation courses should contain texts and tasks that would relate to the social and academic acculturation of test takers. Knowledge of vocabulary is important in determining item difficulty. For example, Korhonen et al. (2012) found out that students with low mathematical performance had low scores in vocabulary, reading comprehension and spelling. Harrison and Morgan (2012) studied about the influence of using simple English on students' performances. They want to identify problems in the language of examinations and found that it is useful to apply simplified English such as using familiar vocabulary, rewording, using simple sentence structure and clear test instructions.

There are other factors that affect students' performance on tests. For instance, Teemant (2010) who wanted to realize ESL university students' problems in testing. He found that students do not know how to express their opinions, memorize in English, paraphrase, or read questions.

2.4. Implications for Practice

This study tried at reviewing (1) weakness and strength of TOLIMO in Iran (2) attitude of PhD students towards high-stake tests like TOLIMO (3) validity and reliability of TOLIMO (4) similarities and differences of TOLIMO with other high-stake tests. The review has clear implications for practitioners since they can use the review as a guide to find the weaknesses of high-stake tests like TOLIMO. The reviewed literature also useful for policy makers who want to change the system of testing in Iran.

2.5. Summary of Empirical Findings

The results of the empirical findings show that:

- Many teachers and students have negative attitude towards testing (Alderson & Hamp-Lyons, 1996).

- The attitude of test takers affect their performances (Greene, 1929).
- There are positive relationship between attitude towards English language and the proficiency of language learners in exams (Malallaha, 2000).
- Test takers have positive attitude towards high-stake tests in different countries (Zhao & Cheng, 2010).
- There is a weak relationship between scores of high-stake tests and tests takers performances (Feast, 2002).
- There is a positive attitude towards IELTS tests between test takers (Rasti, 2009).
- Some of the test takers have problems with the content of high-stake tests (Puspawati, 2012).
- Facilities of the testing are not proper (Sawaki, 2001).

2.6. Statement of the Gap

According to literature, there is not enough explorations and studies about the high-stake tests in Iran, especially TOLIMO. There are many English exams which are used for assessing the English ability of students in schools, institutes and universities, one of them is TOLIMO. Many PhD students must pass this high-stake nation-wide exam to continue their education and it is pre-requisite for higher education. On the other hand, many PhD students cannot pass TOLIMO, they complain about the quality and validity of this test. Also some of the specialist believe that this exam is not standard. Many PhD students believe that TOLIMO is not suitable for assessing the English proficiency.

Generally, literatures show that there is lack of concentration on TOLIMO in Iran. It seems useful to take into account the opinions of test takers to improve the quality of this test. Focusing on weaknesses and strengths of TOLIMO helps policy makers and teachers to enhance the quality of their materials, contents and teaching.

CHAPTER
THREE
METHODOLOGY

Chapter Three

Methodology

3.1. Overview

In this study case study is used as a branch of qualitative studies. There are different definitions for case study. The processes of sampling procedure, data collection, and data analysis were done according to rules of the case study research method.

3.2. Background Information

According to Bromley (1990), in the case study, there is an “organized survey about an event or a set of related events which tries to prove and explain the phenomenon of interest” (p. 302). Becker explains that case study refers to an analysis of the individual case that “one can properly acquire knowledge of the phenomena from intensive exploration of a single case” (p.75). The case study tries to take a comprehensive understanding of the case and develop more theoretical statements about the regularities in the observed phenomena.

The flexibility introduced into the case study is an important and useful attribute as explained by Becker (1970):

It prepares the investigator to deal with unexpected findings and, indeed, requires him to reorient his study in the light of such development. It forces him to consider, however crudely, the multiple interrelations of particular phenomena he observes. And it saves him from making assumptions that may turn out to be incorrect about matters that are relevant, though tangential, to his main concerns. This is because a case study will nearly always provide some facts to guide those assumptions, while studies with more limited data-gathering procedures are forced to assume what the observer making a case study can check on. (p.76).

By reviewing the literature, five features of case study research can be recognized: first one is particularity. Case studies explore the particularity of the case (Stake 1995), and

usually involve a small number of participants who are studied in order to gain understanding of complex social phenomena. The case is selected precisely in order to understand the particular in depth, rather than to find out what is generally true of the many (Cohen et al. 2000; Merriam 1998).

Case studies generate a thick description (Geertz 1973) of the case, which provides rich, in-depth insights and holistic understanding of complex phenomena. This requires that sufficient data be collected to enable researchers to discover systematic connections among experiences, behaviors and relationships, and to explore changes in these over time (Duff 2008; van Lier 2005). Case studies typically make use of multiple sources of data, combining interviews and observations with questionnaires and/or documents, for example, in order to provide different perspectives on the phenomena being studied (Denzin & Lincoln 2003a; Duff 2008; Merriam 1998; Yin 2003). Case studies tend to employ a flexible design which allows the inclusion of an emergent strategy (Anderson & Arsenault 1998) as the study unfolds. Such flexibility, however, requires a rigorous, disciplined and systematic approach to the research design (Richards 2003; Yin 2003). There is an approach that sees the case study as a research strategy, an empirical examination that probes a phenomenon within its real-life context. Case-study research can mean single and multiple case studies, can include quantitative evidence, relies on multiple sources of evidence, and profits from the prior progress of theoretical intentions.

The earliest use of case study referred to Europe, especially to France. The methodology in the United States was most closely connected with The University of Chicago Department of Sociology. From 1900's until 1935, The Chicago School was famous in this field and the source of a great deal of the literature. The sociology is strongly related to the case study research, and during the period leading up to 1935, several problems were propounded by investigators in other fields of study.

Zonabend (1992) stated that case study is conducted by giving special attention to wholeness in observation, reconstruction, and analysis of the cases under study. Case study is fulfilled in a way that integrates the beliefs of the "actors" in the case under study.

A regular criticism of case study methodology is that its reliance on a single case renders it incapable of providing a generalizing conclusion. Yin (1993) presented Giddens' view that considered case methodology "microscopic" because it "lacked a sufficient number" of cases.

Hamel (Hamel et al., 1993) and Yin (1984, 1989a, 1989b, 1993, 1994) strongly argued that the size of the sample that used in cases does not change a multiple case into a macroscopic study. The aim of the study should establish the parameters, and then should be applied to all research. In this way, even a single case could be considered acceptable, provided the established objective.

Tellis (1997) says that study about construct validity causes problem in case study surveys. Because of investigators' bias, construct validity is a source of criticism. Yin (1994) proposed three solutions to solve this problem: using several sources of evidence, forming a chain of evidence, and having a draft case study report reviewed by key informants. Another aspect of validity is internal validity. It is a problem only in causal (explanatory) cases. Internal validity is a problem of "interpretations" in case studies, and can be solved with pattern checking. External validity refers to knowing whether the results are generalizable beyond the instant case. (Tellis 1997).

According to Yin (1994) the case study design must have five components: the research question, proposition, units of analysis, determination of how the data are linked to the propositions and criteria to interpret the findings. Yin concluded that operationally defining the unit of analysis supports with recurrence and efforts at case comparison.

The study's questions are most likely to be "how" and "why" questions, and their definition is the first task of the researcher. The study's propositions sometimes derive from the "how" and "why" questions, and are helpful in focusing the study's goals. An exploratory study, rather than having propositions, would have a stated purpose or criteria on which the success will be judged. The unit of analysis defines what the case is. This could be groups, organizations or countries, but it is the primary unit of analysis. Linking the data to propositions and the criteria for interpreting the findings are the least developed aspects in case studies (Yin, 1994).

Campbell (1975) described "pattern-matching" as a useful technique for linking data to the propositions. Campbell (1975) asserted that pattern-matching is a situation where several pieces of information from the same case may be related to some theoretical proposition.

Stake (1995) stressed that the number and type of case studies depends upon the purpose of the survey: an instrumental case study is used to provide insight into an issue; an intrinsic case study is undertaken to gain a deeper understanding of the case; and the collective case study is the study of a number of cases in order to inquire into a particular phenomenon.

Guba and Lincoln (1981) describe case study "types". These types are factual, interpretative and evaluative. Each case study must outline the purpose, then depending on the type of case study and the actions proposed by the researcher, the researcher could determine the possible products of the study.

3.2.1. Types of Case Study

3.2.2. Exploratory Case Study

An exploratory case study (Yin 2003) is one which aims to identify questions and propositions which can be explored through subsequent study. This study was an exploratory case study which examined participants' own perspectives on experiences related to teacher learning with a view to developing models and theories based on the findings (Bassey 1999; Duff 2008).

3.2.3. Instrumental Case Study

An instrumental case study (Stake 1995) is one in which the case is explored to help generate understanding of wider issues beyond the case itself. This case study was instrumental in that it aimed to provide insights into wider issues regarding teacher learning and teacher education, and facilitate understanding of the relationship between participants' beliefs and actions, rather than studying the cases because of their intrinsically interesting nature.

3.2.4. Multiple Case Study

This was a multiple case study (Yin 2003), sometimes referred to as a collective case study (Stake 1995) or set of cases (Robson 2002), where each of the three participants formed an individual case. Thus, three separate but similar cases were explored within the same context, and subsequent cross-case analysis enabled me to explore similarities and differences among the three individual cases.

An important feature of this study is that it was longitudinal. Multiple data sets were collected at regular intervals over a 20-month period which enabled me to build up a complex vivid picture of each of the participants, to identify turning points and defining moments in the development of their beliefs and teaching, and to explore changes and developments in depth over time (Dörnyei 2007; Duff 2008).

Case study made use of analytic generalization (Yin 2003), whereby findings are generalizable to theoretical propositions rather than to wider populations (I discuss generalizability in greater depth in section 4.5.3 below). Thus, my findings report that something has happened somewhere, posit that it might happen somewhere else, albeit perhaps differently, and invite other researchers to try and find out (Bassegy 1999). The thick description generated by the case study approach also helps readers determine the generalizability of the findings to other contexts (Gall et al. 2007; Richards 2003).

The purposes of case study research may be exploratory, descriptive, interpretive and explanatory (Mariano, 2000). Yin (1993) listed several examples along with the appropriate research design in each case. There were suggestions for a general approach to designing case studies, and also recommendations for exploratory, explanatory, and descriptive case studies.

3.2.5. Exploratory Case Study

Exploratory case study or pilot study is used when researcher is seeking to answer a question that sought to explain the presumed causal links in real-life interventions that are too complex for the survey or experimental strategies. (Yin, 2003). In exploratory case studies, research, and data collection may be started prior to definition of the research

questions and hypotheses. This type of study has been considered as an introduction to some social research. However, the framework of the study must be created ahead of time. Pilot projects are very useful in determining the final protocols that will be used. Survey questions may be dropped or added based on the outcome of the pilot study. Selecting cases is a difficult process, but the literature provides guidance in this area (Yin, 1989a). Stake (1995) recommended that the selection offers the opportunity to maximize what can be learned, knowing that time is limited. Hence the cases that are selected should be easy and willing subjects. A good instrumental case does not have to defend its typicality.

3.2.6. Explanatory Case Study

Explanatory cases are suitable for doing causal studies. In very complex and multivariate cases, the analysis can make use of pattern-matching techniques. Yin and Moore (1987) conducted a study to examine the reason why some research findings get into practical use. They used a funded research project as the unit of analysis, where the topic was constant but the project varied. The utilization outcomes were explained by three rival theories: a knowledge-driven theory, a problem-solving theory, and a social-interaction theory.

3.2.7. Descriptive Case Study

Descriptive cases require that the researcher begin with a descriptive theory, or face the possibility that problems will occur during the project. Pyecha (1988) used this methodology to study special education, using a pattern-matching procedure. Several states were studied and the data about each state's activities were compared to another, with idealized theoretic patterns. Thus what is implied in this type of study is the formation of hypotheses of cause-effect relationships. Hence the descriptive theory must cover the depth and possibility of the case under study. The selection of cases and the unit of analysis is developed in the same manner as the other types of case studies.

There is much debate in the research literature about ways in which qualitative research can enhance the validity of its findings and conclusions (Denzin & Lincoln 2003a; Dörnyei 2007; Duff 2008; Gall et al. 2007; Morse 1999; Robson 2002; Seale 1999). Much

of this debate focuses on whether to replace traditionally accepted concepts - validity, generalizability, reliability and objectivity - with Lincoln and Guba's (1985) alternative concept of 'trustworthiness' - credibility, transferability, dependability and confirmability.

3.3. Research Design

This study aims to investigate PhD candidates' attitude towards TOLIMO, a high-stake nation-wide test administered as a pre-requisite for the completion of the PhD program. The general question, "What do you think about TOLIMO? Would you please explain its strengths and weaknesses?" will be used not only to elicit theoretically relevant data but also to direct this study. This research is qualitative one as a result there isn't any hypothesis.

This is a qualitative research. There are different types of qualitative research, one of them is case study. Consequently this study uses exploratory case study for acquiring data. Exploratory case study is used when researcher is seeking to answer a question that sought to explain the presumed causal links in real-life (Yin, 2003).

The participants of the study are selected by purposive sampling then the investigator take an unstructured open-ended interview to achieve useful information and code. After that the process of triangulation is done to verify the validity of data.

3.4. Sampling Procedure

3.4.1. Background Information

In case study survey, the participants of the study is selected voluntarily. That means every persons who met the requirements and agreed to participate in the study was accepted. In case study research, the selection of samples is very important. Many reports in psychology (Bromley, 1986), sociology (Creswell, 1997; Yin, 1984, 1994), and education (Stake, 1978, 1995) have studied the individual as the unit of analysis, and have used the case study method to improve rich and comprehensive understandings about people.

A case selection that is based on representativeness will seldom be able to produce suitable results. When selecting a subject for a case study, researchers will therefore use information-oriented sampling, as opposed to random sampling. Outlier cases (that is, those

which are extreme, unusual or abnormal) reveal more information than the potentially representative case. Alternatively, a case may be selected as a key case, chosen because of the inherent interest of the case or the circumstances surrounding it.

The persons to be included within the group must be distinguished from those who are outside it' (Yin 2003: 24), and this was done by means of a sampling strategy. The sampling strategy greatly influences the quality of research, and a distinction is made in the literature between 'random' and 'purposive' sampling (Cohen et al. 2000; Morse 1991; Patton 1990). Purposive sampling enables researchers to select 'individuals who can provide rich and varied insights into the phenomenon under investigation so as to maximize what we can learn' (Dörnyei 2007: 126). Case studies can use one participant, or a small group of participants. However, it is important that the population of participants remain small. The participants can represent a diverse cross section of society, but this isn't necessary. Often, a brief "case history" is done on the participants of the study in order to provide researchers with a clearer understanding of their participants, as well as some insight as to how their own personal histories might affect the outcome of the study.

3.4.2. Sampling

In this research the purposive sampling is used. The first participant of the study is selected randomly and after that he suggested another interviewee. In the similar case, the participants of the study suggest their friends and suitable cases who attend in the TOLIMO. The researcher looked for those people who had related experiences about TOLIMO and its quality. The purposive sampling was followed by snowball sampling to find additional participants. In this section, participants introduced their friends who already had participated in TOLIMO. This process was followed until gathered data were going to be repeated and saturated. Participants of the study talked about different aspects of TOLIMO. The purposive sampling technique, also called judgment sampling, is the deliberate choice of an informant due to the qualities the informant possesses. It is a nonrandom technique that does not need underlying theories or a set number of informants. Simply put, the researcher decides what needs to be known and sets out to find people who can and are willing to provide the information by virtue of knowledge or experience (Bernard 2002, Lewis & Sheppard 2006).

3.5. Participants

All the population of the study are from Shahrood University. Both male and female students will participate in this study with different background knowledge about English. They are 15 participants and all of them are from different cities of Iran and their fields of study are various. The ages of participants are from 27 to 50 years old.

Student Participants	Age	Gender	Education	Occupation	Years of College
PhD.C1	29	Male	Mathematic	Student	2
PhD.C2	28	Male	Mechanic	Student	1
PhD.C3	28	Male	Mechanic	Student	1
PhD.C4	27	Female	Mathematic	Student	2
PhD.C5	30	Male	Mining	Student	2
PhD.C6	29	Male	Mechanic	Student	3
PhD.C7	30	Female	Mathematic	Student	2
PhD.C8	29	Male	Civil Engineering	Student	3
PhD.C9	32	Male	Physical Education	Student	3
PhD.C10	34	Male	Computer Engineering	Student	4
PhD.C11	30	Male	Chemistry	Student	2
PhD.C12	29	Female	Mathematic	Student	2
PhD.C13	31	Female	Mathematic	Student	2
PhD.C14	30	Male	Physics	Student	2

Table 2. Participants' data

3.6. Data Collection

3.6.1. Background Information

Data collection, and in general the implementation of a good case study, depend significantly upon the competence of the researcher. Unlike, a questionnaire, the evidence to be gathered is defined as it is collected, and the researcher is an active agent in the process. This means that the researcher undertaking data collection must be able to ask good questions, to listen and to interpret the answers. This involves having a sound grasp of the questions and propositions of the case study, and being able to approach the study in an unbiased, and flexible manner.

Interviews, which are traditionally one of the most important means of eliciting introspective data in case study research (Yin 2003). The interviews were also progressively focused (Silverman 2000; Verschuren 2003), both within and across cases, such that each stage of interviews informed the next.

In case study, qualitative rather than quantitative data was collected in order for analysis, because as Stake (1995, p. 37) points out, qualitative researchers “have forced for explanation and control” as well as “understanding the sophisticated interrelationships among all that exists”. Stake (1995) stated that the protocols that are used to ensure accuracy and alternative explanations are called triangulation. Triangulation is defined as “the process of collecting data from several different sources or in different ways in order to provide a complete understanding of a phenomenon” (Richards & Schmidt, 2003, p. 565). Using triangulation in a case study ensures the resources are fully used so that all the important data is triangulated (Stake, 1995). The need for triangulation arises from the ethical need to confirm the validity of the processes. In case studies, this could be done by using multiple sources of data (Yin, 1984).

Depending on whether researchers have chosen to use a single or multi-modal approach for the case study, they may choose to collect data from one or any combination of these sources: documents, archival records, interviews, direct observation, participant observation and artifacts.

Documents could be letters, notes, programs, administrative documents, newspaper articles, or any document that is useful to the investigation. In the interest of triangulation of evidence, the documents serve to corroborate the evidence from other sources. Documents are also useful for making inferences about events. Archival documents can be service records, organizational records, lists' of names, survey data, and other such records.

Interviews are one of the most important sources of case study information. There are several forms of interviews that are possible: open-ended, focused, and structured or survey. In an open-ended interview, key respondents are asked to comment about certain events. They may propose solutions or provide insight into events. They may also corroborate evidence obtained from other sources. The researcher must avoid becoming dependent on a single informant, and search for the same data from other sources to confirm its validity. The focused interview is used in a situation where the interviewee is interviewed for a short period of time, usually answering set questions. This technique is often used to confirm data collected from another source. The structured interview is similar to a survey, and is used to gather data in cases such as neighborhood studies. The questions are detailed and developed in advance, much as they are in a survey.

Direct observation occurs when a field visit is conducted during the case study. It could be as simple as casual data collection activities, or formal protocols to measure and record behaviors. This technique is useful for providing additional information about the topic being studied. The reliability is enhanced when more than one observer is involved in the task.

Participant-observation makes the researcher into an active participant in the events being studied. This often occurs in studies of neighborhoods or groups. Physical artifacts can be tools, instruments, or some other physical evidence that may be collected during the study as part of a field visit. The perspective of the researcher can be broadened as a result of the discovery.

In the field of composition research, these six sources might be: a writer's drafts, school records of student writers, transcripts of interviews with a writer, transcripts of conversations between writers (and protocols), videotapes and notes from direct field

observations, hard copies of a writer's work on computer. Protocols, that is, transcriptions of participants talking aloud about what they are doing as they do it, have been particularly common in composition case studies.

It can be said then, that cross checking data from multiple sources can help provide a multidimensional profile of composing activities in a particular setting. Merriam, S. (1985) suggests "checking, verifying, testing, probing, and confirming collected data as you go, arguing that this process will follow in a funnel-like design resulting in less data gathering in later phases of the study along with a similar increase in analysis checking, verifying, and confirming.

3.6.2. Data Collection in this Study

The data for this study will be collected through an unstructured open ended interview with the informant of the study. According to Glaser (2001), interviewing was a passive process of listening. As each interview finishes it will be analyzed, concepts and categories will be drawn.

Interviews will be recorded with the permission from interviewee, however, they will be sure that their real names and identities will not be announced, instead researcher will use fabricated names. Interview is a tool which permits an in-depth exploration of a particular topic and goes beneath the surface of ordinary conversation (Charmaz, 2006). Interviewer will start with one broad introductory questions that can be followed by relevant follow-up questions. The first question is that: What are your perceptions of TOLIMO? The participant should be active whereas the interviewer should listen actively and encourage the participant to talk and to clarify details (Charmaz, 2006), for example; “that is interesting, tell me more about it”.

3.7. Data Analysis

3.7.1. Background Information

Generally, researchers interpret their data in one of two ways: holistically or through coding. Holistic analysis does not attempt to break the evidence into parts, but rather to draw conclusions based on the text as a whole. However, composition researchers commonly interpret their data by coding that is by systematically searching data to identify and/or categorize specific observable actions or characteristics. These observable actions then become the key variables in the study.

This exploratory case study aimed to generate theory rather than to test existing theory, and as such the analytical approach I took was essentially iterative and abductive (Blaikie 1993; Coffey & Atkinson 1996; Constan 1992; Morse 1994; Scott & Usher 1999). Whereas deductive approaches begin with, and inductive approaches begin without, a set of a priori categories, abduction is a strategy which implies a dialogic relationship between analysis and theory and a cyclical rather than a linear logic.

I constantly checked and revised these until new categories emerged, taking care to avoid either imposing prior categories of analysis or prematurely forming such categories as both may have contaminated the data or influenced my perspective (Morse 1994; Patton 1990; Silverman 2000). The explicitness of research procedures enhances the credibility of research (Lincoln & Guba 1985).

3.7.1.1. Pre-coding:

The first stage of data analysis, pre-coding, consisted of three elements; (1) transcription of data, (2) writing of analytic memos, and (3) initial development of categories.

I developed my own system for transcribing (as recommended by Dörnyei 2007; Duff 2008; Lapadat 2000; Roberts 1997), and corrected minor linguistic errors so as to create a more natural and readable discourse.

I then read, re-read and reflected on each transcription and wrote analytic memos for each. These served as summary sheets for each set of data collected (Duff 2008). The third step in pre-coding was to begin developing categories of analysis.

3.7.1.2. Coding:

The second stage of data analysis involved 'reducing the data' in order to highlight key concepts and categories and to start linking these to broader concepts (Hammersley & Atkinson 1995; Morse 1994; Silverman 2000; Tesch 1990).

3.7.1.3. Theorizing:

The final stage of data analysis involves interpreting, theorizing and drawing conclusions (Coffey & Atkinson 1996; Dörnyei 2007; Hammersley 1992; Miles & Huberman 1994; Morse 1994; Richards 2005; Silverman 2001).

Miles and Huberman (1984) suggested analytic techniques such as rearranging the arrays, placing the evidence in a matrix of categories, creating flowcharts or data displays, tabulating the frequency of different events, using means, variances and cross tabulations to examine the relationships between variables, and other such techniques to facilitate analysis. Yin (1994) presented two strategies for general use: One is to rely on theoretical propositions of the study, and then to analyze the evidence based on those propositions. Pattern-matching is another major mode of analysis. This type of logic compares an empirical pattern with a predicted one. Internal validity is enhanced when the patterns coincide. If the case study is an explanatory one, the patterns may be related to the dependent or independent variables. If it is a descriptive study, the predicted pattern must be defined prior to data collection. Yin (1994) recommended using rival explanations as pattern-matching when there are independent variables involved. This requires the development of rival theoretical propositions, but the overall concern remains the degree to which a pattern matches the predicted one.

Stake (1995) recommended categorical aggregation as another means of analysis and also suggested developing protocols for this phase of the case study to enhance the quality of the research. He also presented ideas on pattern-matching along the lines that Yin (1994)

presented. Runkel (1990) used aggregated measures to obtain relative frequencies in a multiple-case study. Stake (1995) favored coding the data and identifying the issues more clearly at the analysis stage. Eisner and Peshkin (1990) placed a high priority on direct interpretation of events, and lower on interpretation of measurement data, which is another viable alternative to be considered.

Merriam, S. (1988) suggests seven analytic frameworks for the organization and presentation of data: the role of participants, the network analysis of formal and informal exchanges among groups, historical, thematically, resources, ritual and symbolism, critical incidents that challenge or reinforce fundamental beliefs, practices, and values.

There are two purposes of these frameworks: to look for patterns among the data and to look for patterns that give meaning to the case study.

While most researchers begin their case studies expecting to look for particular observable characteristics, it is not unusual for key variables to emerge during data collection. Typical variables coded in case studies of writers include pauses writers make in the production of a text, the use of specific linguistic units (such as nouns or verbs), and writing processes (planning, drafting, revising, and editing).

3.7.2. Data Analysis

In this study, once written records are available, analysis involves the coding of data and the identification of salient points or structures. Data analysis may begin informally during interviews or observations and continue during transcription, when recurring themes, patterns, and categories become evident. There are three main propositions in this study: Positive attitude towards TOLIMO, Negative attitude towards TOLIMO and finally suggestion for improvement. For the positive attitude we had extracted two subcategories and in negative attitude part we extracted seven subcategory.

CHAPTER
FOUR
RESULTS

Chapter Four

Results

4.1. Overview

The purpose of this study is to investigate PhD candidates' perceptions of TOLIMO. Case study was used to collect and analyze data. The interviewees in this study state their experiences and understanding from TOLIMO. The main goal of this chapter is to present the factors that shape attitudes towards TOLIMO.

Analyzing the data revealed three main categories: positive attitude towards TOLIMO, negative attitude towards TOLIMO and suggestion for improvement. The candidates who believe that TOLIMO is a standard test talked about the sufficiency of the timing of exam, reasonable relationship between the timing of exam and number of tests and finally they say it can evaluate participants exactly. Also, other participants mentioned: insufficiency of timing of the exam, substandard administration of listening section, low quality administration, inappropriateness of listening section, substandard text length in reading part and inappropriateness of number of the questions as the main drawbacks of the test.

4.2. Positive Attitude towards TOLIMO

Analysis revealed negative attitude towards TOLIMO as the rule and positive attitudes towards the test as an exception. Only a small group of candidates believe that TOLIMO effectively evaluates the participants' proficiency. They take into account the timing of exam and good evaluation of the test.

4.2.1. Sufficiency of the Time

Some PhD candidates state that the exam timing is sufficient. It means that there is a good correlation between number of questions and timing. If you are ready for the exam, you can answer the questions in allocated time. PhD.C3 explains this issue:

There are many questions in this exam, but in my opinion it's not weakness. The candidate who has studied for exam can easily pass the

test, as a result it's not related to the time of the exam. The candidates compete with the other participants of the test so the time of exam is the same for everyone.

4.2.2. Evaluation Appropriateness of the Test

Other groups of interviewees believe that TOLIMO is a good choice for evaluating the PhD candidates. In comparison with MSRT, another similar exam, TOLIMO is chosen by many PhD students. They say that this exam can categorize good and bad participants. Also weak and strong students in this test have not the same performance. PhD.C14 explains:

I think the questions of TOLIMO are more standard than MSRT. If you have studied for exam you can get a good grade. In my opinion. Candidates are compared with other participants and after that your final score is calculated.

PhD.C5 explains that TOLIMO is a good tool for evaluation of English knowledge. This group of students believe that good score depends on sufficient knowledge of English:

TOLIMO can exactly distinguish poor candidate from knowledgeable ones. Every candidate who has the sufficient knowledge in English can take a suitable score. I have taken the score that I expected. Sometimes you receive the score that it is higher than your expectation.

4.3. Negative Attitude towards TOLIMO

4.3.1. Insufficiency of Timing of Exam

There are many PhD students who complain about the exam timing. They say that no relationships between timing of exam and number of questions. As it was mentioned by the candidates, they cannot answer to all questions just because of the shortage of time. PhD.C4 says:

I didn't see any relationship between number of questions and timing. Many candidates did not enough time to answer questions. I think the

difficulty of the questions influence on the performance of the students so that they don't have time. It, means there is relationship between timing and difficulty of test.

Some of the students believe that the problem of timing, relate to the English knowledge of students. It means that candidates are not knowledgeable and their level is not acceptable. PhD.C2 explains that:

The timing of exam is not sufficient, I think many students like me are not ready for exam, so that they cannot manage their time in the session of exam. Also students are very weak in English. Therefore, they focus on each question and lose the time. Candidates should be able to manage time, they are PhD students and before that pass similar exams.

Other groups of students believe that the amount of time is directly related to the number of questions. In their opinions, managers should decrease the number of questions to help candidates. PhD.C7 states that:

There is not good relationship between time and amount of questions. Time is very short so you cannot think about the questions. The authorities should increase the time or decrease the number of questions.

The amount of time depends on the knowledge of students and their readiness but nearly all the interviewees believe that time is insufficient. Candidates suggest the time increased. PhD.C5 says:

It seems that time of TOLIMO is very short. The participants of TOLIMO are from different fields therefore, they are not knowledgeable in English. I think managers of the test can add few minutes to the total time. Many candidates translate the questions into the Persian and this process consume the time.

4.3.2. Substandard Administration of Listening Section

TOLIMO is held in different universities, institutes and schools. But sometimes the facilities of these places are not acceptable. Many participants criticize the atmosphere of the sessions and the classes used for testing. In their opinions the authorities don't pay attention to the facilities and place of testing. The other group of students focus on hardware section of listening part. For example, PhD.C8 says that:

The physical condition is awful. For example, there isn't good response from speakers in listening section. Because of this problem, many students are not able to answer the questions. As a result candidates are cheating or sleeping during the listening section. I believe that listening section is difficult but it can be improved by using the best facilities.

Some candidates criticize the place of testing. They believe that the classes and place of testing are small, also sounding system is not suitable for this kind of exam. PhD.C6 explains:

The atmosphere of the place of testing is ambiguous for me. The classes that used for exam is small, also I think these classes are not suitable for this important exam. The authorities don't know anything about the exam. You can easily understand the problem in sounding system, many students cannot understand the speaker. I should say that some of the listening problems related to the poor knowledge of candidates.

There are many complains about the quality of sound system of testing. Some candidates state that the performances of students in listening section related to the quality of the sound of the speakers. PhD.C7 states that:

I had attend in TOLIMO twice. In the first exam I was not ready for the listening part and the quality of the speakers was good, but in the second test the sound of the speakers was unclear. I was ready for listening but I couldn't hear what the speaker says.

Although, TOLIMO is held in famous universities, it doesn't warranty the quality of the listening section. Candidates criticize the facilities of universities especially the sounding system. PhD.C11 explains his opinion:

I don't like the manner of holding exam. The managers of the exams used the terrible speakers. I can't understand the voice of listening section. The exam was held in Elm o Sanaat University so that I expected the high quality sound in listening section.

According to some candidates the behavior of the proctors besides sounding system is a big problem. The candidates say that proctors don't pay attention to the candidates and talk to each other. Also in the listening section the talking between proctors annoy the candidates. PhD.C9 says that:

I think the listening section is difficult for all candidates and the reason for this problem refers to the sounding system of the test. The managers of the exam use the sound software that make the conversations unreal. Another problem is talking of proctors during the test. They don't take into account the participants and talk to each other in listening section. Sometimes they tell joke and funny things to other persons in session.

There are students who state that the quality of listening section is not proper for this type exam. They compare the facilities of TOLIMO with IELTS and TOEFL. PhD.C10 explains that:

If you attend in IELTS and TOEFL exams, the listening section is held in laboratory, but you don't this that in TOLIMO. The cost of TOLIMO is few so that managers don't provide the sufficient things for you. There are speakers at the end of the saloon and you cannot hear the voice of speaker of listening section.

4.3.3. Low Quality of Presentation of Exam

TOLIMO just held in big cities with huge population like Tehran, Mashhad, Isfahan, Tabriz and Shiraz. The candidates from other parts of the country gather in these cities for testing but some of them believe that it's not fair. There are many PhD students in different cities and many universities and institutes have PhD students in small provinces so the holding of exam in few cities is a big problem. For example PhD.C12 says:

TOLIMO is not held in small cities, I think it's a problem. PhD students should travel to the big cities like Tehran and Mashhad for exam, in my opinion the distance and tiredness influence on their performances. Candidates have not a suitable place for resting and in the session they are tired and sleepy.

The other group of students have good suggestions for the problem of distance and tiredness of candidates. They suggest to hold the TOLIMO in center of provinces for helping the candidates. PhD.C12 states that:

TOLIMO is held in few cities and this is a weakness. I think the authorities can find a solution, in my opinion TOLIMO can be held in the center of each province so that candidates should not travel to big cities. One of the main weaknesses of travelling to other cities is cost of trip.

4.3.4. Inappropriateness of Listening Section

According to interviewees, listening part is the most difficult section of TOLIMO. Many reasons are introduced by candidates as a cause of difficulty. PhD candidates complain about the texts that used in listening, time, instructions and vocabulary.

Students say that they are not ready for listening so that they suggest to omit listening in TOLIMO. PhD.C7 says that:

Listening was very difficult for me, I had not proficiency in this part. Also I was not familiar with oral instructions of this part. Unfamiliarity with instructions made the listening difficult. I suggest to omit the listening, because by elimination of listening students can focus on reading and grammar. These parts are more suitable than listening for candidates.

Weak participants have problem in listening. They don't practice for this part and just focus on reading and grammar. Candidates state that basic knowledge of PhD students are very poor. PhD.C3 says:

The most difficult part of exam is listening. Many PhD students have not basic knowledge about English, therefore, they try to get the acceptable mark. On the other hand, listening needs practice. You cannot answer the questions of this part with one month practice.

Other group of students believe that holding exams like TOLIMO is for article writing. PhD.C8 states that:

Listening is difficult for all candidates. Authorities hold exams like this to improve the writing skills of students but there is useless parts like listening. Candidates cannot have good performance in this section because they have not listening proficiency.

Some of interviewees believe that the performance of students related to the materials they use for improving listening skill. For example PhD.C9 says:

Many students have not good exercise for listening, therefore, listening is difficult for them and finally they eliminate this part. I think using hard question is easier for listening than grammar and reading. Some

of the students have not access to the proper sources. Also I think there are good books for reading and grammar.

In all courses, students should attend in the different classes. For TOLIMO preparation, candidates say that English classes can be helpful. Some of the students say that attending in the listening class is beneficial. PhD.C13 explains that:

If you want to have good performance in listening section of TOLIMO, attend the listening classes of institutes. You must practice about one year to have proficiency in listening. I couldn't answer the listening question because I was not ready. I suggest to introduce CDs for listening practice.

Candidates of TOLIMO are from different fields of study so they are not proficient in English. Candidates state that listening should be designed according to the level of participants. They complain about rate of speaking and time of listening part. PhD.C14 states:

When you listen to the audios of listening section the sound interrupted immediately. Students are from different fields of study so I think it's better to allow participants think about listening questions, I mean add more time to this part.

In many cases, participants state that listening is not related to their studies. PhD students try to write ISI articles and thesis, therefore, listening is not necessary for them. They say that listening and speaking are useful for students who want to continue their studies in abroad. PhD.C1 says that:

Listening was very difficult for me. In TOLIMO there isn't negative point for false answers so I omit listening section. Listening has not positive effect for participants, PhD students use English for translation, therefore, listening is not useful for them. They want to study the scientific websites and software in English actually grammar and reading are more important than listening.

4.3.5. Substandard Text Length in Reading Part

One of the important parts of TOLIMO is reading. In this exam, there are four or five readings. In many TOLIMO exams, readings are very long with many questions. Participants say that they lose time by reading the texts. Also they believe that time for this part is very short. PhD.C2 says:

In my opinion reading is the main weakness of this test. Every year the texts of exam are selected from special books so that candidates answer questions by memorizing. For example, one of my friends have attend in TOLIMO many times and he has passed the exam by memorizing the texts. Many readings of the test are reiterative and sometimes texts refer for special theme.

Some of the interviewees say that texts are selected from special sources. Selection of texts from specific sources is the main mistake. Participants have access to the sources and by memorizing they can answer the questions. PhD.C4 explained:

Texts and readings are originated from special book. When you attend in this exam, you face with question bank. I didn't answer the reading question because texts were long.

According to PhD candidates, the content of readings doesn't related to the fields of study of participants. Texts are selected from scientific and historical sources, on other hand, the unknown vocabulary is used in these texts. PhD.C6 explained:

When I participated in the exam, I faced with the texts that related to the astronomy and cardiology. My field of study in university is physical education so I was not familiar with professional words of these fields. I think one of the problem of TOLIMO is that it doesn't cover all fields.

Reading section can be useful because you should learn grammar and vocabulary.

Other participants have similar idea and say that text and its vocabulary are not match with their fields. PhD.C10 says:

The readings are long and boring. TOLIMO is very similar to Konkoor, in which you can find five long texts. The time that is allocated to the reading section is not enough. I can easily read my special books and texts that related to my studies, but I cannot understand the meaning of the words that used in TOLIMO. One of the techniques that used by designers of TOLIMO is that they select a text and substitute the easy words with difficult ones.

4.3.6. Insufficiency for the Personal Studies

In many interviews PhD candidates of TOLIMO say that holding English exams like TOLIMO and MSRT is not helpful. Half of the interviewees believe that their English knowledge doesn't change after test. According to expressions of participants, content, structure and manner of TOLIMO is not related to their studies. PhD.C4 said:

Today, it's not necessary to have TOLIMO degree, because it's not useful for students. TOLIMO is not valid internationally so I believe it should be omitted. The content of the exam is ordinary and it doesn't related to the fields of students. As a result, if you are not successful in this exam, you don't lose anything.

One of the important parts of TOLIMO is grammar. In this section, the grammar questions are presented to candidates. Many of these questions related to the tenses. Some of the students believe that past tense is more useful than other tenses. They state that holding exam is for article writing. PhD.C14 explained:

First of all, we should determine the aim of English exams. Authorities say that English exams help participants to write article. On the other, there are useless tenses like future tense in exam. When a PhD candidate writes an article, he or she uses just past and present tense. I think the usage of other tenses in grammar section is inappropriate.

There are other candidates who believe their English knowledge is same without any change. PhD.C2 explained:

My English knowledge hadn't change after this test. An exam cannot change the knowledge of students. In TOLIMO there isn't any difference between the persons who pass the test and failed ones. I don't know how this exam evaluate the candidates, also it's not clear what' the aim of this exam. Helping for article writing or improving English knowledge?

4.3.7. Inappropriateness of Number of the Questions

Many participants of the TOLIMO believe that there are many questions in the exam. They say that it is not essential to pass many questions that waste their time and energy. TOLIMO tries to cover all areas of the English knowledge so different types of questions are used to evaluate candidates accurately. The candidates say that they cannot manage the time because of number of questions. PhD.C5 explained that:

There are many questions in the test, on the other hand the time of exam is not sufficient so that I cannot think about the answers. I am not proficient in English, therefore, I need time for thinking. By large amount of questions I lose my concentration and energy for thinking.

Some of the candidates believe that number of questions is the weak point of TOLIMO. In their opinion students are tired during the test, they have travelled from different cities so they lose their energy. As a result, they are not ready for many questions. PhD.C7 explained that:

The number of questions is very large, it made me tired. At the end of the exam I was tired therefore, I couldn't focus of the questions. There is not any relationship between the number of questions and time. The managers should decrease the number of questions. In my opinion the weak point of TOLIMO is questions.

4.4. Suggestions for Improvement

The participants of this study according to their experience suggest the matters that can help test designers. Many of the PhD candidates attend in TOLIMO exam more than one time, as a result, they understand the negative and positive points. They suggest to include: 1. speaking and writing section 2. introducing source for studying

4.4.1. Including Speaking and Writing Section

Candidates in the study believe that TOLIMO should include writing score because of article writing of the PhD students. They explain that article writing can be substitute with test. PhD.C12 explains:

I think the test should be arranged according to the candidates' necessity. For example some of the students want to apply for foreign universities, the exams like TOLIMO should designed for these students. The candidates who want to apply for other universities should be able to write and speak, on the other hand, the candidates who want to continue his or her study in Iran should focus on article writing. I suggest that PhD students who publish ISI article are knowledgeable in English, therefore, they shouldn't pass exams like TOLIMO. I think the theme of TOLIMO should be related to the field of studies of Participants.

In TOLIMO, there isn't speaking section for evaluating the speaking ability of participants. It means that the speaking is not important. Some of interviewees suggest to include the speaking. They believe that PhD candidates should be able to

communicate with foreigners, therefore, they need speaking ability. PhD.C1 says that:

If you want more academic test, you should include the score of speaking section. The tests like TOEFL and IELTS have speaking section so the participants can evaluate their ability in all the aspects. Learning new language is not grammar and reading, speaking play important role in language learning.

4.4.2. Introducing Sources for the Test

TOLIMO like other high-stake tests has not any source for studying. In this exam, test designers have not introduced specific book, CD, article and text for test takers. Participants of the test suggest authorities to introduce the resources for candidates. PhD.C4 explains that:

There isn't any source for TOLIMO, I suggest test designers to introduce source for grammar and vocabulary section of the test. In these books writers can explain the grammar rules and questions that similar to the questions of grammar section of TOLIMO. Also it would be good if authorities arrange CD for listening section. The questions of the exam should not be taken from the books and CDs, because many students try to memorize the content of the sources.

CHAPTER

FIVE

DISCUSSION AND

CONCLUSIONS

Chapter five

Discussion and Conclusion

5.1. Overview

The purpose of this study is to explore the attitude of PhD candidates towards TOLIMO by qualitative case study as well as the suggestion of participants for improvement of the test is taken into account. To this end, some PhD students who take TOLIMO are chosen to introduce their attitude and experience about TOLIMO.

This chapter presents a summary of the findings, a discussion of the findings in relation to the former studies in this field, the pedagogical implications of the study, and finally the suggestions for additional research.

5.2. Discussion

Participants of this study state their experiences and opinions about the quality of the TOLIMO. There are three main propositions about TOLIMO. The first group of interviewees have positive attitude towards TOLIMO. The second group have negative attitude and the last group have some suggestions for improvement.

The candidates who have positive attitude say that TOLIMO is a good instrument for evaluation of the PhD candidates. The goal of using the tests are diverse. For example, Bachman and Purpura (2008) claims that language tests used in many cases like selection of students, giving certifications, and giving promotion. Small group of participants of the present study says that TOLIMO is a good choice for selection the good candidates. These group of interviewees claim timing, number of questions and difficulty of the test are normal.

The second group of participants of the study state their negative point of view about TOLIMO. According to Murray, Riazi and Cross (2012) negative attitudes of test takers were personal experiences and feelings of test takers and influence of people. On the other hand, according to Vallerand and Reid (1984) motivation can be enhanced after the exam if the

student receives positive feedback. The negative attitudes of the interviewees have divided into six subcategories: 1)insufficiency of timing of exam, 2)substandard administration of listening section, 3)low quality of presentation of exam, 4)inability of test to evaluate candidates, 5)substandard text length in reading part, 6)insufficiency for the personal studies and 7)inappropriateness of number of the questions.

Analysis of the data shows that timing of exam is insufficient. Participants say that timing is limited so that weak students cannot manage the time. In their opinion, there are many questions in TOLIMO, therefore they need extra time for completing the test. In the study, Bachman (1990) believes that performances of the students depend on timing of the exam. In the similar study, Triplett & Barksdale (2005) found that many students were worried about the time of testing. Interviewees believe that test takers should decrease the number of questions or increase the time of exam.

The second subcategory in negative attitude refers to administration of the listening section. Based on results of this study, the facilities of the place of testing are not suitable. Participants say that they cannot easily hear the sound from speakers. Also the place of the exam like classes and saloons are small. Sawaki (2001) found that the test-takers complained about the facility of the place of testing. Similarly, Puspawati (2012) found that the students could not focus on test because of the inappropriateness of the place of testing. Another reason that is elicited from interviews is that, proctors are distracting the participants of the test. Some of the proctors talk to each other so that candidates cannot focus on the exam. There are interviewees who complain that listening section should be held in laboratory like TOEFL and IELTS.

The results of the study show that many students are not satisfied with the cities in which TOLIMO are held. TOLIMO is held in big cities like Tehran, Isfahan, Shiraz, Mashhad and Tabriz so candidates from small cities should travel to big cities. They are tired and sleepy because of trip the so that they cannot perform very well in exam. Shohamy (1998) says that some of the students feel fear, unfairness, bias, pressure and doubt during the test preparation. In another study, Bachman (1990) claims that the performances of test takers in

exams affected by factors like familiarity of the place, the equipment that used and staffs. Some of the students suggest that TOLIMO can be held in the center of each province.

The result of chapter four show that listening section is very difficult for some group of participants. They believe that unclear instructions, lack of proficiency in English and ambiguous sounds make this section difficult. In the similar study, Shohamy (1993) claimed that the use of tests has led to top-down educational change strategies and this can be considered unethical, undemocratic and unbeneficial for test takers. It means that the candidates try to learn and memorize the materials of listening section without any internalization. The previous studies have shown that English proficiency affects students' performance on tests (Aydn et al., 2006; Teemant, 2010; Korhonen et al., 2012).

Another important weak point of TOLIMO according to participants of the study is length of reading texts. Candidates believe that reading part included long and boring texts. Also the number of questions in this part are excessive. Some group of interviewees say about the content of the texts. In their opinion special books are used for this part. Also the scientific contents that doesn't related to many field are used in TOLIMO. Sometimes the content of the text is not related to any field of study.

Results make clear for us that TOLIMO is not suitable for personal studies of PhD students. They told us that attending in TOLIMO doesn't help them in article writing, reading text and speaking. Participants of the study say that holding exams like TOLIMO and MSRT cannot help students to improve their knowledge in English. In similar study, Rasti (2009) found that there is not any meaningful relationship between Iranian candidates' educational background, and IELTS scores with their attitudes towards IELTS.

Another weak point of TOLIMO according to participants of the study is number of questions. They believe that there are many questions in the test, as a result, there is not meaningful relationship between number of questions and timing of exam.

There are good suggestions for improvement of the TOLIMO. Shohamy (2001a) claimed that considering the opinions of test takers about exams direct us to a better understanding about exams specifically high-stake tests. Some of the candidates suggest to

arrange tests according to the necessities of the participants. They believe that TOLIMO should be related to the fields of studies of students, also there should be focus on writing parts for students who want to apply for foreign universities. Participants of the study say that reading parts should be related to the fields of studies of students.

Another suggestion is that PhD candidates who write ISI article should be excused from TOLIMO. These group of interviewees say that the student who write ISI article is very proficient in English. The other group of interviewees suggest to include speaking section for TOLIMO. They believe that PhD students want to travel foreign countries for seminars so that they need speaking ability.

5.3. Implication for Practice

There are many high-stake English tests like TOEFL, IELTS, TOLIMO and MSRT in Iran that PhD candidates attend. The policy makers should try to improve the quality of Iranian exams like TOLIMO and MSRT, because these exams are used for evaluation. Timing is an important element in every exam, as a result, authorities should try to increase the time as well as the place of testing is very important for participants.

Policy makers by the results of this study can improve the reading and listening section of the TOLIMO. They can use better facilities for holding the test and decrease the length of the reading parts. Speaking section can be added to the TOLIMO and MSRT to help students. The test designers can include more fields of studies in reading parts. Also, TOLIMO can be held in every province and city.

5.4. Suggestion for Further Studies

This study tries to search about the attitudes of PhD candidates who attend in TOLIMO. Other researchers can focus on the similar test, MSRT. The present study ask PhD candidates about their opinions but these questions can be asked from other participants who are not PhD students. TOEFL and IELTS are held in Iran for evaluation of PhD candidates, the similar study can be done for participants of IELTS and TOEFL.

References:

- Allport, G. W. (1971). *Attitudes*. In K. Thomas (Ed.), *Attitudes and behavior*. Harmondsworth, UK: Penguin.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation course: a study of wash-back. *Language Testing*, 13(3), 280-297.
- Anderson, G. & Arsenault, N. (1998) *Fundamentals of Educational Research*. 2nd Edition London: Falmer.
- Ata, A. W. (2010). Grammatical interference from Arabic to English can pose big challenges for students. *E l gazette (Britain)*, 4, 2 -23.
- 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).
- Bachman, L. (2000). Forward. In G. Ekbatani & H. Pierson (Eds.). *Learner directed assessment in ESL* (pp. ix-xiii). New Jersey: Lawerance Erlbaum Associates.
- Baker, C. (1992). *Attitudes and language*. Clevedon, UK: Multilingual Matters.
- Bassey, M. (1999) *Case Study Research In Educational Settings*. Buckingham: Open University Press.
- Bernard, H.R. 2002. *Research Methods in Anthropology: Qualitative and quantitative methods*. 3rd edition. AltaMira Press, Walnut Creek, California.
- Bird, E., & Welford, G. (1995). The effect of language on the performance of second language students in Science examinations. *International Journal of Science Education*, 17(3), 389-397.
- Blaikie, N. (1993) *Approaches to social enquiry polity press*, Cambridge UK.
- Broadfoot, P. (1996). *Education, assessment and society: a sociological analysis*. Buckingham: Open University Press.
- Birjandi, P., & Alemi, M. (2010). The impact of test anxiety on test performance among Iranian EFL learners. *Broad Research in Artificial Intelligence and Neuroscience (BRAIN)*, 1(4), 44-58.
- Bridgeman, B., & Cho, Y. (2012). Relationship of TOEFL iBT® to academic performance: Some evidence from American universities. *Language Testing*, 29(3), 421-442.

- Bromley, D. B. (1986). *The case-study method in psychology and related-disciplines*. Chichester: John Wiley & Sons.
- Bromley, D. B. (1990). Academic contributions to psychological counselling: I. A philosophy of science for the study of individual cases. *Counselling Psychology Quarterly*, 3(3), 299-307.
- Brown, D. H. (2000). *Principles of language learning and teaching* (fourth ed.). Pearson Education.
- Campbell, D. (1975). Degrees of freedom and the case study. *Comparative Political Studies*, 8, 178-185.
- Carless D (2011) From Testing to Productive Student Learning: *Implementing Formative Assessment in Confucian-heritage Settings*. New York: Routledge.
- Chapman, D. W., & Snyder, C. W. (2000). Can high-stakes national testing improve instruction: reexamining conventional wisdom. *International Journal of Educational Development*, 20, 457-474.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage Publications.
- Coffey, A. & Atkinson, P. (1996). *Making Sense of Qualitative Data Analysis: Complementary Strategies*. Thousand Oaks CA: Sage.
- Cohen, L, Manion, L and Morrison, K. (2000). *Research Methods in Education*. (Fifth edition) London and New York: Routledge. Chapter 3, 4, 5, 9, 13, 14, 15
- Constas, M. A. (1992). Qualitative analysis as a public event: The documentation of category development procedures. *American Educational Research Journal*, 29(2), 253-266.
- Cotteral, S. (1999). Key variables in language learning: what do learners believe about them? *System*, 27, 493-513.
- Creswell, J. W. (1997). *Qualitative inquiry and research designs: Choosing among five traditions*. Thousand Oaks, CA: Sage.
- Denzin, N. K. and Y. S. Lincoln (2003). *Collecting and interpreting qualitative materials*. Thousand Oaks, Calif., Sage.

- 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.
- Duff, Patricia A. (2008). *Case Study Research in Applied Linguistics*. New York: Lawrence Erlbaum. Pp. 248.
- Eisner, E., & Peshkin, A. (Eds.). (1990). *Qualitative inquiry in education*. New York: Teachers College Press.
- Feast, V, 2002, 'The impact of IELTS scores on performance at university. *International Education Journal*, 3(4), pp 70-85.
- Filipi, A. (2012). Do questions written in the target language make foreign language listening comprehension tests more difficult? *Language Testing*, 29(4), 511-532.
- Finocchiaro, M. (1982). *Motivation: it's crucial role in language learning*. ERIC Document Reproduction Service.
- Fiocco, M, 1987, 'English Proficiency Levels of Students From a Non English Speaking Background: A Study of IELTS as an Indicator of Tertiary Success', research report, Centre for International English, Curtin University of Technology.
- Gardner, R. C. (1983). Learning another language: A true Social Psychological Experiment. *Journal of Language and Social Psychology*, 2, 219-239.
- Gardner, R. C., & MacIntyre, P.D. (1993). A student's contributions to second language learning. Part II: Affective Variables. *Language Teaching*, 26, 1-11.
- Geertz, Clifford. (1973). *The interpretation of cultures: selected essays*. New York: Basic Books
- Glaser, B. (2001). *The Grounded Theory Perspective: Conceptualization Contrasted with Description*. Mill Valley, CA: Sociology Press.
- Gottlieb, M. (2006). *Assessing English language learners: Bridges from language proficiency to academic achievement*. Thousand Oaks, California: Corwin Press.

- Greene, E.B. (1929). Achievement and confidence on true-false tests of college students. *The Journal of Abnormal and Social Psychology*, 23(4), 467-478.
- Green, E., & Evans, S. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, 6, 3–17.
- Graham, J.G. (1987). 'English Language Proficiency and the Prediction of Academic Success. *TESOL Quarterly*, 21(3), pp 505-521.
- Guba, E.G. and Y.S. Lincoln (1994), "Competing paradigms in qualitative research" in Denzin and Lincoln (eds.), *Handbook of Qualitative Research*, Sage, Thousand Oaks.
- Hamel, J., Dufour, S., & Fortin, D. (1993). *Case study methods*. Newbury Park, CA: Sage Publications.
- Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in Practice, 2nd edition*. London: Routledge.
- Harrison, S. & Morgan, R. (2012). Using simplified English to identify potential problems for non-native speakers in the language of engineering examination papers. *The Language Learning Journal*, 40(1), 113-123.
- He, L., & Shi, L. (2008). ESL students' perceptions and experiences of standardized English writing tests. *Assessing Writing*, 13(2), 130-149.
- Holmes, J. (1992). *An introduction to sociolinguistics*. London: Longman.
- Ho, Y. (2008). *ESL/EFL teachers from overseas-Yes or No?* Paper presented at the Annual Meeting of the MWERA Annual Meeting, Western Great Southern Hotel, Columbus, Ohio Online 22/6/2012. Institute of Education Sciences (ies). (2012).
- 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.
- Kirkland, M. C. (1971). The effect of tests on students and schools. *Review of Educational Research*, 41(4), 303–350.
- Lamb, M. (2004). It depends on the students themselves: Independent language learning at an Indonesian state school. *Language, Culture, and Curriculum*, 17, 229–245.

- Langerfeld, K., Thurlow, M., & Scott, D. (1997). *High-stakes testing for students: Unanswered questions and implications for students with disabilities* (Synthesis Report 26). Minneapolis: University of Minnesota, National Center for Educational Outcomes.
- Lapadat, J. C. (2000). Teacher research on classroom discourse in northern Canadian communities. *Networks: An Online Journal for Teacher Research*, 3(2).
- Leung C, Lewkowicz J (2006) Expanding horizons and unresolved conundrums: language testing and assessment. *TESOL Quarterly* 40(1): 211–34.
- Lewis, J.L. & S.R.J. Sheppard. 2006. *Culture and communication: can landscape visualization improve forest management consultation with indigenous communities?* *Landscape and Urban Planning* 77:291–313.
- Lewthwaite, M. (2007). Teacher and Student Attitudes to IELTS writing tasks: positive or negative wash-back? *Action Research Report, UGRU Journal*, 5, 1-23.
- Light, RL, Xu, M and Mossop, J, 1987, 'English proficiency and academic performance of international students. *TESOL Quarterly*, 21(2), pp 251-261.
- Li, M.; Baker, T. & Marshall, K. (2002). 'Mismatched expectations: a case study of Asian students in New Zealand', Paper presented to the New Zealand Applied Business Education Conference, *Christchurch College of Education*, Christchurch, 25-27 September.
- Luxia, Q. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142–173.
- Jamieson, J., Taylor, C., Kirsch, I., & Eignor, D. (1999). *Design and evaluation of a TOEFL CBT tutorial* (TOEFL Research Rep. No. 62; ETS Research Rep. No. 99-01). Princeton, NJ: ETS.
- Johnson, K., & Johnson, H. (eds.) (1998). *Encyclopedic Dictionary of Applied Linguistics*. Blakwell Publishers Ltd.
- Kerstjens, M and Nery, C, 2000, 'Predictive validity in the IELTS test: A study of the relationship between IELTS scores and students' subsequent academic performance. *IELTS Research Reports*, 3, IELTS Australia, Canberra.
- Korhonen, J. Linnanmaki, K. & Aunio, P. (2012). Language and Mathematical performance: a comparison of lower secondary school students with different level of Mathematical skills. *Scandinavian Journal of Educational research*, 56(3), 333-344.

- Kohn, A. (2000). Burnt at the high-stakes. *Journal of Teacher Education*, 51(4), 315–327.
- Madaus, G. F. (1985). Public policy and the testing profession: You've never had it so good? *Educational Measurement: Issues and Practice*, 4(4), 5-11.
- Malallaha, S. (2000). English in an Arabic environment: current attitudes to English among Kuwait University students. *International Journal of Bilingual Education and Bilingualism*, 3, 19-43.
- Mariano, C. (2000). *Case study: the method. Chapter 10. In P. Munhall & C. Oiler Boyd, Eds. Nursing Research. A Qualitative Perspective (2nd ed.)* (pp. 311- 337). Sudbury, MA: Jones and Bartlett Publishers.
- McNamara, T. (2004). Language testing. In A. Davies, & C. Elder (Eds.), *the handbook of applied linguistics* 763-783.
- Merriam, S. B. (1985). The Case Study in Educational Research: A Review of Selected Literature. *Journal of Educational Thought*, 19.3, 204-17.
- Merriam, S. (1998). *Qualitative research and case study applications in education*. San Francisco (USA): Jossey-Bass. p. 179.
- Miles, M., & Huberman, M. (1984). *Qualitative data analysis: A source book for new methods*. Thousand Oaks, CA: Sage Publications.
- Morse, J.M. (1999). *Nursing Research: The application of qualitative approaches*. London: Chapman & Hall.
- Murray, J. C., Riazi, A. M., & 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.
- Nunan, D. (1988). *The learner Centred curriculum: A study in Second language Teaching*. Cambridge: Cambridge University Press.
- Onwuegbuzie, A. Baiey, P. & Daley, C. (2001). Cognitive, affective personality and demographic predictors of FL achievement. *The Journal of Educational Research*, 94(1), 3-15.
- O'Neill, R. (1991). The plausible myth of learner centredness: Or the importance of doing ordinary things well. *ELT Journal*, 45, 293-304.

- Ory, J. (2004). *Improving your test questions*. EES Memorandum #27. Evaluation and Examination Service. University of Iowa.
- Oxford, R., & Shearin, J. 1994. Language learning motivation: Expanding the theoretical framework. *Modern Language Journal*, 78(1), 12-28.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods, 2nd ed.* Thousand Oaks, CA: Sage Publications.
- Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In S. Sinharay & C.R. Rao (eds.) *Handbook of Statistics Psychometrics* (pp. 125-167). New York: Elsevier.
- Puspawati, I. (2012). *EFL/ESL (English as Foreign/Second Language) students' perceptions toward the TOEFL (Test of English as a Foreign Language) test.*
- Pyecha, J. (1988). *A case study of the application of non-categorical special education in two states.* Chapel Hill, NC: Research Triangle Institute.
- Ragin, Charles C. & Becker, Howard S. (1992). *What is a Case? Exploring the Foundations of Social Inquiry.* Cambridge, USA: Cambridge University Press.
- Rasti, I. (2009). Iranian candidates' attitude towards IELTS. *The Asian EFL Journal Quarterly*, 11(3), 110-155.
- Read J & Hayes B. The Impact of IELTS on preparation for academic study in New Zealand. *IELTS Australia Publication*, 4(5). From IELTS Research Abstracts, 2003, Abstract No.5
- Reeve, J. (1996). *Motivating others: Nurturing inner motivational resources.* Massachusetts: Allyn & Bacon.
- Richards, K. (2003) *Qualitative Inquiry in TESOL.* Palgrave Macmillan, Basingstoke.
- Richards, J.C. & Schmidt, R. (2002). *Longman Dictionary of Language Teaching & Applied Linguistics (3rd edition).* Essex: Pearson Education Limited.
- Robson C (2002) *Real World Research.* Blackwell, (2nd edition).
- Runkel, P. (1990). *Casting nets and testing specimens: Two grand methods of psychology.* New York: Praeger.

- Salehi, H., Karimvand, P. N., & Naeeni, S. K. (2011). *Various aspects of high-stakes school tests: A close look at the phenomenon of washback*. In Proceedings of 24th International Congress for School Effectiveness and Improvement (ICSEI, 2011), Limassol, Cyprus.
- Sarason, I. G., & Sarason, B. R. (1990). Test anxiety. In H. Leitenberg (Ed.), *Handbook of social and evaluation anxiety* (pp. 475-493). New York: Plenum Press.
- Scott, D., & Usher. R. (1999). *Researching education data, methods and theory in educational enquiry*, London and New York: Cassell.
- Seale, C. F. (1999). *The quality of qualitative research*. London: Sage.
- Shi, L. & He, L. (2012). Topical knowledge and ESL writing. *Language Testing*, 29(3), 443-469.
- Sawaki, Y. 2001: Comparability of Conventional and Computerized Tests of Reading in a Second Language. *Language Learning & Technology*, 5(2), 38–59.
- Shohamy E (2007). The power of language, the power of English language and the role of ELT. In: Cummins J, Davison C (eds) *International Handbook of English Language Teaching, Vol. 15*. New York: Springer, 521–31.
- Shohamy E. Donitsa-Schmidt S. Ferman I. Test impact revisited: Washback effect over time. *Language Testing*. 1996; 13(3):298-317.
- Shohamy, E. (2001a). *The power of tests: Critical perspective on the uses of language tests*. Essex, London: Pearson Education.
- Silverman, D. (2000). *Doing Qualitative Research: A Practical Handbook*. London: Sage.
- Smith, M. L. (1991a). Put to the test: the effects of external testing on teachers. *Educational Research*, 20(5), 8-11.
- Stake, R. E. (1978). The case study method in social inquiry. *Educational Researcher*, 7(2), 5-8.
- Stake, R. (1995). *The art of case research*. Thousand Oaks, CA: Sage Publications.
- Stricker, L. J., Wilder, G. S., & Rock, D. A. (2004). Attitudes about the computer-based Test of English as a Foreign Language. *Computers in Human Behavior*, 20, 37-54.
- Teemant, A. (2010). ESL student perspectives on university classroom testing practices. *Journal of the Scholarship of teaching and learning*, 10(3), 89-105.

- Tellis, W. M. (1997). Introduction to Case Study. *The Qualitative Report*, 3(2), 1-14.
- Tesch, R. (1990). *Qualitative research: Analysis Types and Software Tools*. New York: Falmer.
- Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high-stakes testing. *Journal of Literacy Research*, 37(2), 273-260.
- 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.
- Truitt, H. (1995). Beliefs about language learning: A study of Korean University students learning English. *Texas Papers in Foreign Language Education*, 2, 12-28.
- Uysal, H. H. (2010). A critical review of the IELTS writing test. *ELT Journal*, 64(3), 314-320.
- Vallerand, R. J., & Reid, G. (1984). On the causal effects of perceived competence and task feedback on intrinsic interest. *Journal of Personality and Social Psychology*, 6, 94-102.
- Van Lier, L. (1996). *Interaction in the language curriculum: Awareness: Autonomy and authenticity*. London: Longman.
- Van Lier, L. (2005). Case Study. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 195-208). Mahwah, NJ: Erlbaum.
- Verschuren, P. J. M. (2003). Case study as a research strategy: some ambiguities and opportunities. *International Journal of Social Research Methodology*, 6, Issue 2, pp. 121-139.
- Wait, I.W. & Gressel, J.W. (2009). Relationship between TOEFL score and academic success for international engineering students. *Journal of Engineering Education*, 98(4), 389-398.
- Yin, R. (1984). *Case study research: Design and methods (1st ed.)*. Beverly Hills, CA: Sage Publishing.
- Yin, R. (1989). *Case study research: Design and methods (Rev. ed.)*. Beverly Hills, CA: Sage Publishing.
- Yin, R. (1989). *Inter-organizational partnerships in local job creation and job training efforts*. Washington, DC: COSMOS Corp.
- Yin, R. (1993). *Applications of case study research*. Beverly Hills, CA: Sage Publishing.

- Yin, R. (1994). *Case study research: Design and methods (2nd ed.)*. Beverly Hills, CA: Sage Publishing.
- Yin, R., & Moore, G. (1987). The use of advanced technologies in special education. *Journal of Learning Disabilities, 20*(1), 60.
- Xu, M. (1991). The impact of English-language proficiency on international graduate students' perceived academic difficulty. *Research in Higher Education, 32*(5), 557-570.
- Ya-Chin, T. & Yi-Chih, L. (2012). Test anxiety and foreign language reading anxiety in a Reading-Proficiency Test. *Journal of Social Sciences, 8*(1): 95-103.
- 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.
- Zeidner, M. & Bensoussan, M. (1988). College students' attitudes towards written versus oral tests of English as a foreign language. *Language Testing, 5*, 100-113.
- Zhao, J., & Cheng, L. (2010). Exploring the relationship between Chinese university students' attitude towards the College English Test and their test performance. In L. Cheng & A. Curtis (Eds), *English language assessment and the Chinese learner* (pp. 190- 201). New York: Routledge, Taylor & Francis Group.
- Zonabend, F. (1992, spring). The monograph in European ethnology. *Current Sociology, 40*(1), 49-60.



دانشگاه صنعتی شاهرود

دانشگاه صنعتی شاهرود

گروه زبان انگلیسی

پایان نامه کارشناسی ارشد آموزش زبان انگلیسی

بررسی نگرش دانشجویان دکتری نسبت به آزمون تولیمو

نگارنده:

آریا جوانمردی

استاد راهنما:

سید علی استوار نامقی

بهمن ماه ۱۳۹۶

چکیده:

این تحقیق تلاش دارد تا نگرش دانشجویان دکتری نسبت به آزمون تولیمو در ایران را بررسی کند. تولیمو آزمونی است که در ایران برای ادامه‌ی تحصیل در مقطع دکتری به کار گرفته می‌شود. این تحقیق، تحقیقی کیفی است که از روش تحقیق موردپژوهی استفاده می‌کند. آزمون‌های زبان گوناگونی در ایران وجود دارد اما شواهد نشان می‌دهد که تحقیق کافی درباره‌ی تولیمو انجام نشده است. شرکت‌کنندگان در این تحقیق به روش نمونه‌گیری هدفدار انتخاب شده‌اند و پس از آن محقق از مصاحبه‌ای با پایان باز برای جمع‌آوری داده استفاده کرده است. سه طبقه‌ی اصلی در این تحقیق وجود دارد: نگرش مثبت نسبت به تولیمو، نگرش منفی و پیشنهاد برای بهبود آزمون. سیاست‌گذاران با استفاده از نتایج این تحقیق می‌توانند نسبت به بهبود کیفیت بخش خواندن و شنیدن تولیمو اقدام کنند، همچنین استفاده از امکانات بهتر برای برگزاری آزمون و کاهش اندازه‌ی متن‌های خواندن از جمله پیشنهادهای این تحقیق است.

کلمات کلیدی: نگرش، تولیمو، دانشجویان دکتری، موردپژوهی، خواندن، شنیدن