



**BENEMÉRITA
UNIVERSIDAD AUTÓNOMA DE PUEBLA**

FACULTAD DE LENGUAS

**ANALYZING SCORES OBTAINED BY LEMO AND
LEI UNDERGRADUATE STUDENTS IN THE
TOEFL® ITP: PERCEPTIONS AND PERFORMANCE.**

THESIS

SUBMITTED TO THE FACULTY OF LANGUAGES FOR THE DEGREE OF:
LICENCIATURA EN LENGUAS MODERNAS

BY:

DAVID HERNÁNDEZ MONTALVO

DIRECTED BY:

M.A. GEORGINA AGUILAR GONZÁLEZ

HERÓICA PUEBLA DE ZARAGOZA, PUE.

SPRING 2014



**Analyzing Scores Obtained by LEMO and LEI Undergraduate Students in the
TOEFL® ITP: Perceptions and Performance.**

A Thesis Submitted to the Faculty of Languages for the Degree of

LICENCIATURA EN LENGUAS MODERNAS

By **David Hernández Montalvo**

Benemérita Universidad Autónoma de Puebla

Spring 2014

**Analyzing Scores Obtained by LEMO and LEI Undergraduate Students in the
TOEFL® ITP: Perceptions and Performance.**

This Thesis has been read by members of the Thesis Committee of

David Hernández Montalvo

And is considered worthy of approval in partial fulfillment of the requirement for the
degree of

LICENCIATURA EN LENGUAS MODERNAS

M.A. Georgina Aguilar González
Thesis Director

Dr. Rebeca Elena Tapia Carlin
Committee Member

Dr. Celso Pérez Carranza
Committee Member

DEDICATIONS

There are many people I would like to thank but a page is not enough. First of all, I would like to express all my gratitude to **GOD**, the one and only, who gave me life and the opportunity to be where I am at this point; He has done as he said.

To my mother **Soledad** for giving me all her love and support in my life; I know life is a difficult path but she was a clear example for me to follow. Since the beginning, she protected me and worked hard for me to study a B.A. and succeed. I'll reward what she has done. Thanks go to my father **Itiel** for his love and support before and during my studies, for his words of advice and all he has done for me. Furthermore, I want to thank my siblings **Big-Vai**, **Mibsam** and **Mesu-Lemet** for sharing precious moments with me and their support whenever I needed it. The three of them are an important part of my life and I would never be able to achieve all my goals if they would not be there to support me.

Special thanks go to my aunt **Isabel** for her support, for those talks we used to have, for all the valuable things I learned from her, and for letting me stay with her and my cousins **Benjis** and **Johanna** during university. I also want to recognize my grandmothers **Hermelinda's** and **Felipa's** words of advice from which I learned so many treasured things. Thanks to my cousins **Betsabe** and **Elizabeth**, and my uncles **Isaí** and **Miqueas** in the United States for their love and support when I was there. I also want to say thank you to the **REST OF MY FAMILY** who are many people but an important part of my life.

I want to emphasize my gratitude to my dear friends **Samitho** and **Erickun** who I knew at the beginning of the B.A. for the time we spent together and the friendship we have now. To **Miriam**, my little friend *Pinguís*, for everything. It was a pleasure to be her partner and friend. To my dear friend **Peter** for his help, support and friendship; I could never stop telling him thanks because he was always there when I needed him.

Moreover, I want to thank **Alice, Sofí, Samantha, Carolina, Getze, Ulises, Mayte** and **Luisa**, for sharing special moments and experiences that I will never forget, for their support and friendship they were willing to give me during university, summer camps and up to now. Thanks to **Edgar, Deyis** and all my friends from the Faculty of Languages.

I want to thank all those teachers who showed me the way to success and taught how to achieve my goals. To my best teacher in primary school **Victoria Zepeda**, I really want to become a good teacher as she is; I will never forget her. Also, I want to thank teachers **Gudelia Rocha** and **Evangelina** who were my first formal contact with English; I really appreciate all their work and how both of them made me get interested in learning this language. Special thanks go to the best teachers we have in the Faculty of Languages of the BUAP, professors **Julio D. Gonzalez, Georgina Aguilar, Ma. Eugenia Garduño, Marsha Way** and **Carmen Martinez** for making me understand the importance of becoming a teacher and learn how to face obstacles. I also want to recognize professor **Ma. Esther Mendez** to whom I own my little French; I did not like French but she made me start loving it.

Finally, the last but not least, I want to thank **Ma. Lupe** and her family and aunt **Veronica** and her daughters for all their love and support, to my best friends from **COBAO** and **COBAEP** with whom I spent an important part of my life and from whom I learned many things. I also want to express my gratitude to **Chayo, Janid** and **Elena** for the amazing childhood we had together and our strong friendship. But most of all, I wish **GOD** may be with all of you forever.

ACKNOWLEDGEMENTS

*¹⁷ Then I applied myself to the understanding of wisdom,
and also of madness and folly, but I learned that this, too,
is a chasing after the wind.*

*¹⁸ For with much wisdom comes much sorrow;
the more knowledge, the more grief.*

Ecclesiastes 1:17-18

Specials thanks go to professor Mtra. Georgina Aguilar González for her time, dedication, and effort to help me develop this research. Without her guidance and persistent help this thesis project would never have been possible. I want to thank her for making me lose fear to present such project in front of a big audience.

Also, I would like to express my gratitude to Dr. Rebeca Elena Tapia Carlin and Dr. Celso Pérez Carranza as part of the Committee Member for their time and dedication to read and contribute with this thesis project with their comments and ideas.

Specials thanks go to my dear best friend Samuel Herrera Cruz who is a really important part of this project as writer, source of ideas and support. Moreover, I would like to thank my dear best friends Alonso Erick Gómez Trujillo and Pedro García Osorio for their help and comments to complete this project.

Finally, I want to thank those LEMO and LEI students who willingly accepted to contribute with the study as participants.

TABLE OF CONTENTS

CHAPTER I: INTRODUCCION	1
1.1 Introduction to the problem	1
1.2 Purpose of the study	3
1.3 Research questions	4
1.4 Significance of the study	4
1.5 Definition of terms	5
1.6 Chapter conclusions	6
 CHAPTER II: LITERATURE REVIEW	 7
2.0 Introduction	7
2.1 EFL and ESL	7
2.1.1 Testing and Language testing	8
2.1.1.1 Test and types	9
2.1.1.1.1 Placement and Diagnostic tests	10
2.1.1.1.2 Achievement Test	11
2.1.1.1.3 Proficiency Test	12
2.1.1.1.3.1 Language proficiency	13
2.2 TOEFL Test	13
2.2.1 Importance of the TOEFL test	15
2.2.2 Types of TOEFL	16
2.2.2.1 Paper-Based TOEFL (TOEFL® pBT)	16
2.2.2.2 Computer-Based TOEFL (TOEFL® cBT)	17
2.2.2.3 Internet-Based TOEFL (TOEFL® iBT)	17
2.2.2.4 Institutional TOEFL (TOEFL® ITP)	18
2.2.3 TOEFL's sections	19
2.2.3.1 Listening Comprehension Section	19
2.2.3.2 Structure and Written Expression Section	20
2.2.3.3 Vocabulary and Reading Comprehension Section	20
2.2.4 Scoring the TOEFL	21

2.2.4.1 Scoring the TOEFL® pBT	22
2.2.4.2 Scoring the TOEFL® cBT	22
2.2.4.3 Scoring the TOEFL® iBT	23
2.2.4.4 Scoring the TOEFL® ITP	23
2.2.5 TOEFL reliability	24
2.2.6 The TOEFL and the CEFR	26
2.3 TOEFL test taker's perceptions, characteristics and performance	27
2.3.1 Communicative language ability	28
2.3.2 Test methods facets	29
2.3.2.1 Facets of Test Environment	30
2.3.2.2 Facets of Test Rubric	31
2.3.2.3 Facets of Input and Expected Response	32
2.3.3 Background or Personal Characteristics	33
2.3.4 Topical Knowledge and Affective Schemata	34
2.3.5 Attitude and Motivation	34
2.3.6 Perceptions and Expectations about the TOEFL Test	36
2.4 Chapter Conclusions	36
 CHAPTER III: METHODOLOGY	 37
3.0 Introduction	37
3.1 Describing Setting and Subjects	37
3.2 Approach	38
3.3 Describing Instruments	38
3.4 Data Collection Procedures	40
3.5 Data Analysis Procedures	40
3.6 Chapter Conclusions	41
 CHAPTER IV: FINDINGS	 43
4.0 Introduction	43
4.1 Research Questions	43
4.2 Participants' Performance	44

4.2.1 Common Scores, Strengths and Weaknesses	44
4.3 Analysis of Results from the Questionnaire	49
4.3.1 Personal Characteristics	49
4.3.2 Participants' Perceptions towards the TOEFL® ITP Test	50
4.3.3 Participants' Background	55
4.3.4 Participants' Performance	59
4.4 Chapter Conclusions	70
 CHAPTER V: CONCLUSIONS	 71
5.0 Introduction	71
5.1 Findings	71
5.1.1 LEMO and LEI Undergraduate Students' Performance	71
5.1.2 Students' Perceptions Towards the TOEFL® ITP Test	73
5.1.3 Factors that Affected LEMO and LEI Students' Performance	74
5.2 Significance of Findings	74
5.3 Limitations of the Research	75
5.4 Suggestions for Further Research	76
5.5 Reflective account	79
 REFERENCES	 81
Appendix A: Estadísticas Nacionales México 2011	88
Appendix B: TOEFL Test Taker's Questionnaire	89

LIST OF FIGURES

Figure 1: General Scores Obtained in the Diagnostic TOEFL® ITP Test	44
Figure 2: Listening Comprehension Section Scores	45
Figure 3: Structure and Written Expression Section Scores	46
Figure 4: Vocabulary and Reading Comprehension Section Scores	46
Figure 5: Test Takers' Last Score in the Training TOEFL Test	51
Figure 6: Number of Years Participants Have Studied English	56
Figure 7: Appropriate Time for Answering Each Item in the TOEFL® ITP Test	66
Figure 8: The Volume of the Audio in the First Section Was Appropriate	67
Figure 9: Being Motivated to Take the TOEFL® ITP Test in the Future	68
Figure 10: Factors that Motivates Test Takers to Take the TOEFL® ITP in the Future	68
Figure 11: The TOEFL® ITP Test as the Only Way to Demonstrate English Proficiency	69
Figure 12: The Final TOEFL® ITP Test Score Gives a Clear Idea of English Proficiency	70

LIST OF TABLES

Table 1: Participants' Strength and Weakness	48
Table 2: Participants Having Taken the Training TOEFL Test	51
Table 3: Participants' Perceptions Towards the TOEFL® ITP Test	52
Table 4: Test Takers' Age to Start Learning English	55
Table 5: Participants Learning English in Different Educational Levels	56
Table 6: English as Being Used in Participants' Daily Life	58
Table 7: Familiarization with the TOEFL® ITP Test	60
Table 8: Clear Instructions Before the TOEFL® ITP Test Started	61
Table 9: Presence of Personnel Applying the Diagnostic TOEFL® ITP Test Causes Anxiety	62
Table 10: Factors Presented at the Moment Participants Were Answering the Test .	64

ABSTRACT

*Vanity of vanities, saith the Preacher,
vanity of vanities; all is vanity.*

Ecclesiastes 1:2

The TOEFL test has become so important in the certification of the English language that most universities consider it as an appropriate test to evaluate EFL students' proficiency (Boyd et al. 2007, Griffin 2004, Hernandez et al. 2005, Bachman et al. 1990). Most of times, students are required to have a specific score in order to prove their proficiency level (Britt, 2009, p.2). However, there are many internal and external factors as well as perceptions that may have either positive or negative effects on the scores (Poorman, Mastorovich & Molcan 2007, cited in Oermann & Gaberson, 2009, p. 79; Tinsley & Wescot 1990, cited in Xie, 2007). This quantitative-qualitative research aims to find out the most common scores that students from the Faculty of Languages of the Benemérita Universidad Autónoma de Puebla (BUAP) get when they take the TOEFL® ITP test, and reflect about the perceptions students have about this test as well as some factors that might have either positive or negative effects on students' performance through the application of a diagnostic TOEFL® ITP test and the analysis of a questionnaire applied to students. Such study intends to contribute with the improvement of students' performance when doing the TOEFL test, providing experiences and considering students' strengths and weaknesses reflected on the scores in order to help them achieve better results to prove their English proficiency.

CHAPTER I: INTRODUCTION

1.1 Introduction to the problem

In the last century, the evaluation of knowledge of English as a Second Language (ESL) and English as a Foreign Language (EFL) has become an important field of research. Institutions and teachers have been searching for effective ways to evaluate student's knowledge of English language; they have always assessed students to determine their proficiency level of English. Assessing implies the use of tests which can be of achievement which measure what a learner has learned from a particular course or proficiency which measure what a learner knows of a language (Nation & Macalister, 2010). For EFL students, universities and institutions have applied different kinds of tests to determine students' proficiency level of English such as the Cambridge First Certificate in English (FEC) mainly used in Britain and the Test of English as a Foreign Language (TOEFL) applied principally in North America, including Mexico. The TOEFL is considered to be a secure test which is recognized as proof of English proficiency by 2,500 universities in the U.S., Canada, and other countries (Alderson, Krahne, & Stansfield, 1987). This proficiency test has many versions but the principal ones are the Institutional TOEFL (ITP), the Paper-based TOEFL (pBT), the Computer-based TOEFL (cBT), and the International and Official Internet-based TOEFL (iBT) (Britt, 2009).

Currently, in Mexico, apart from the 41 Test Centers in which the TOEFL test is administered, approximately 80 institutions recognized the TOEFL as an appropriate examination for language testing including prestige universities such as the UNAM (Autonomous National University of Mexico) and the BUAP (Autonomous University of Puebla). In some cases, these institutions make from the TOEFL a requirement to

get a certification of English or getting graduated and getting the degree of a major. As a consequence, lots of books have been published for guiding test takers on the preparation for this test. Furthermore, some research has been developed to study the validity, the reliability, the content and the evaluation of this proficiency test.

Recently, the Institute of International Education generated a graphic of statistics of 145,918 TOEFL® ITP tests administered to Mexican people in general in 2011 from which the state with highest average is Baja California Norte (533 points) and the state with the lowest average is Puebla (457) (See Appendix A). Furthermore, it has been observed that students studying higher education do not get the score required by most universities which, according to information provided in Britt's research (2009, p.2), is usually 550 points. For this reason, universities have been truly worried about the scores obtained by test takers in terms of scoring performance and factors that may influence that performance; they have analyzed those scores from different perspectives and have drawn many conclusions. For instance, researchers from the Department of Modern Languages of the Tecnológico de Monterrey were concerned about scores obtained by their students which varied from 510 to 547, since passing the TOEFL with a score equal to or higher than 550 points is a requirement for getting a degree. They have identified the necessity of an English course to make students increase their scores on the institutional TOEFL exam by analyzing and contrasting the learning content evaluated on this proficiency test and the learning content learned by students from their curricular English courses (Hernández, Salazar, Arellano, Rosales & Garza, 2005).

It is also the case of many students currently studying the Licenciatura en Lenguas Modernas (LEMO) and the Licenciatura en Enseñanza del Inglés (LEI) at

the Faculty of Languages of the Benémerita Universidad Autónoma de Puebla, being the Departamento de Planeación y Evaluación Académica (DEPEA - Department of Planning and Academic Evaluation) which administers the TOEFL to those students who want to prove their English proficiency , or to those students who want to get graduated and get the degree without defending a thesis project. The minimum expected score is 550 points. A training TOEFL is also applied at the DEPEA which presents the same type of tasks that a regular TOEFL® ITP does; this test is free and available for all students to practice before presenting a real TOEFL® ITP.

It is important to mention that in this university there has been a transition between two educational models, from the Proyecto Universitario Fénix that started in 2002 to the Modelo Universitario Minerva (MUM) implemented since 2009. LEMO students are under the Proyecto Universitario Fénix and LEI students are under the MUM educational model. On one hand, the Modelo Fénix establishes for LEMO students to take eight English courses best known as Target Language (TL); on the other hand, the MUM stipulates that LEI students have to take five TL courses and 4 workshops that can be considered as English courses.

1.2 Purpose of the study

The main purpose of this research is to identify the most common scores that students from the Faculty of Languages of the BUAP obtain when they apply for the TOEFL® ITP test as well as the perceptions they have about this test. Moreover, this study is aimed to identify the factors that have either positive and/or negative effects on the students' performance reflected on the scores obtained. This project will focus

on undergraduate LEMO and LEI students from the Faculty of Languages of the BUAP that have already finished their TL courses.

1.3 Research Questions

This project addresses the following questions:

1. What are the most common scores that undergraduate LEMO and LEI students obtain in the TOEFL® ITP examination?
2. What are the LEMO and LEI students' strengths and weaknesses in the TOEFL® ITP examination?
3. What are the student's perceptions about the TOEFL® ITP exam and the scores obtained?
4. What are the factors that influence positively and/or negatively LEMO and LEI students' scores in the TOEFL® ITP test?
5. What conclusions can be drawn about the results obtained by LEMO and LEI students in the TOEFL® ITP exam?

1.4 Significance of the study

Since passing TOEFL® ITP exam is a proof that LEMO and LEI undergraduate students have good level of the English language proficiency, this study intends to contribute with the improvement of students' performance when doing the TOEFL test. This study will also provide information about what LEMO and LEI students have been through and what they need in order to be ready for taking the TOEFL exam. Therefore, TL teachers can realize about these experiences and consider students' strengths and weaknesses reflected on the scores in order to help them achieve

better results. Furthermore, this project suggests the creation of a curricular course that can be implemented to make students get familiar with this test and provide them with the necessary tools to succeed on the TOEFL® ITP examination.

1.5 Definition of Terms

EFL (English as a Foreign Language). “By EFL I mean English as studied by people who live in places where English is not the first language of the people who live in the country, such as in Italy, Saudi Arabia and Korea” (Gebhard 1996, p. 2).

ESL (English as a Second Language). “By ESL I mean English as studied by people who speak other languages as their first language –such as Spanish, Arabic, Chinese, or Swahili– but live in places where English is the first language, such as in Australia, New Zealand, Canada, the United States, and the United Kingdom” (Gebhard 1996, p. 2).

Assessment. It can be defined “as a broad and comprehensive process of collecting quantitative and qualitative data to make informed educational decisions. It is a process that encompasses the full range of procedures used to obtain information about student learning” (Nitko 1996, cited in McDonald 2002, p. 9).

Evaluation. This term refers to “the systematic gathering of information for the purpose of making decisions” (Weiss 1972, cited in Bachman 1997, p. 22).

Language testing. This concept “involves the assessment of some or all aspects of the language ability of individuals in some contexts (not necessarily that of a language class) and for some set of purposes (not necessarily common to all parties)” (Allison 1999, p. 5).

Proficiency. This term refers to “a general type of knowledge of or competences in the use of a language, regardless how, where or under what conditions it has been acquired... performance as measured by a particular testing procedure” (Davies, et al. 2002, p. 153).

Proficiency test. “Proficiency tests are most often global measures of ability in a language or other content area. They are not necessarily developed or administered with reference to some previously experienced course of instruction. These measures are often used for placement or selection, and their relative merit lies in their ability to spread students out according to ability on a proficiency range within the desired area of learning” (Henning 1987, p. 6).

TL (Target Language). This term refers to “the modern foreign language that pupils are learning” (Pachler & Field 2001, p. 86).

TOEFL (Test of English as a Foreign Language). The TOEFL “is a standardized proficiency exam. It is used to test and grade the English ability of non-native speakers. It is published by Educational Testing Service (ETS)” (Britt 2009, p. 2).

1.6 Chapter Conclusions

Chapter I provided a brief introduction to the topic by establishing the research field, summarizing previous research and presenting the issue. It also provided the purpose of the study as well as the research questions. Then, the significance of study was presented followed by a list of key terms within the theoretical framework of this research.

CHAPTER II: LITERATURE REVIEW

2.0 Introduction

The main purpose of this chapter is to provide the theoretical bases that highlight the issue that is investigated. Introductory notions about evaluation in English as a Foreign Language (EFL), focusing on the TOEFL test, will be presented for a better understanding of the research development. First of all, concepts of English as a Foreign Language and English as a Second Language (ESL) are characterized, and evaluation, assessment as well as testing are contrasted. Secondly, important information regarding test and types of test are presented. Then, the TOEFL test is described according to its importance, types, sections or abilities tested, its score procedures, its reliability and recommended cut values mapped onto the Common European Framework. Finally, research about test taker's characteristics and factors that influence performance will be presented.

2.1 EFL and ESL

English has become one of the most spoken languages in the world; consequently, when it comes to research about a specific aspect of English language learning, it is important to differentiate two wide fields in which English is the central core: EFL and ESL. Sometimes these terms seem to be thought of in the same way; however, they are very different according to researchers in the area of language learning. On the one hand, EFL can be defined as English being studied by people from countries where English is not the first language (Cook, 2003). On the other hand, ESL is “English as studied by people who speak other languages as their first language – such as Spanish, Arabic, Chinese, or Swahili– but live in places where English is the

first language, such as in Australia, New Zealand, Canada, the United States, and the United Kingdom” (Gebhard, 1996, p. 2). From this perspective of language learning, it is clear to see the differences between these two linguistic fields.

2.1.1 Testing and Language Testing

There has always been a necessity for measuring procedures; Henning (1987) expresses the view that science without measurement is impossible. Since more people from different places throughout the world are interested in EFL learning, institutions where English is taught are more aware of the necessity of evaluation of the learner’s knowledge about this language. Therefore, in the processes of EFL evaluation and language teaching, testing plays an important role (Alderson, Krahne, & Stansfield, 1987) which consists on measuring the results of student’s performance (Baxter, 1997). In other words, testing refers to the conversion of the student’s performance into quantitative data in order to be analyzed for various specific purposes.

In the evaluation of the process of learning a foreign language, testing is seen as the easiest and most traditional way to show the learner’s knowledge of a language. According to Cook (2003, p. 7), language testing can be defined as “the assessment and evaluation of language achievement and proficiency, both in first and additional languages, and for both general and specific purposes”. From this definition, it can be said that testing seems to be a reliable source of data for being analyzed in order to make important decisions about what actions might be taken so that the learners may improve their performance. Baxter (1997) mentions a semi-scientific procedure in language testing that consists of four steps: measuring the

student's performance, doing something to affect the performance, measuring the performance again, and compare the results. Concerning Baxter's procedures, it is clear to see the importance of language testing and the use of tests which help any institution not just to measure but to improve the way they work.

2.1.1.1 Test and Types

As it was mentioned before, a test is an important tool for many teachers and institutions to assess students, since it may be considered to be "a substitute for a more complete procedure" (Baker, 1989, p. 4). Carroll (1968, cited in Bachman 1997, p. 20) refers to a test as to a measuring item for obtaining a sample of an individual's behavior which serves for making inferences about his or her features; this definition emphasizes the use of qualitative procedures, and that is what teachers and institutions do for different purposes depending on the issues involved in students' learning process.

It is important to mention that there have been some misconceptions regarding the usage and development of this item as well as a debate about what makes a test the best test (Bachman & Palmer 1996, Alderson, Krahne, & Stansfield 1987). Institutions search for effective ways to evaluate students' knowledge but they might have problems when choosing the kind of test to use due to the fact that designers do tests that may not apply to the institutions in terms of the context (e. g. test takers and level of language). From this idea of developing a specific test, institutions can use a test that could work efficiently when applying it; such idea has caused a debate that has been held about the belief of the existence of the best test. This naïve expectation has become an obstacle for the development of approaches to language

testing design. Bachman & Palmer (1996) suggest an approach that involves developing the ability to design, develop, evaluate and use language tests in ways that are appropriate for a given purpose, context, and group of test takers as well as understanding the fundamental issues and concern in the appropriate use of language tests.

In language testing, there are many classifications about types of tests suggested by many authors depending on their features and purpose; some tests are flexible but others are more restricted. Henning (1987) and Hughes (1989) provide a similar classification of tests depending on the way the test is scored (objective or subjective), the way language is tested (directly or indirectly), a reference (norm-reference or criterion-reference), speediness and difficulty of the test (speed or power), the areas of the language being tested (discrete-point or integrative point), and purpose and content of the test. From this affirmation, it follows that this latter category is composed by four kinds of test (placement test, diagnostic test, achievement test, and proficiency test) that are going to be defined in the following sections.

2.1.1.1.1 Placement and Diagnostic Test

Many teachers require students to take a diagnostic test at the beginning of a specific course which has as main purpose to identify strengths and weaknesses in an individual (Henning, 1987; Alderson, Krahne & Stansfield, 1987; Hughes 1989). That means that the information collected from the results is supposed to lead to interpret test taker's needs concerning learning language abilities. Alderson, Krahne & Stansfield (1987) agree that one of the characteristics of this test is the extensive

measurement of aspects of second language behavior such as listening or reading comprehension. Hence, it is a good instrument for teacher to know what needs to be taught.

On the other hand, plenty of institutions make individuals interested in taking a specific course from their curriculum take a placement test which, according to Hughes (1989, p. 14) has as main purpose to “provide information which will help to place students at the stage (or in the part) of the teaching program most appropriate to their abilities”. Something important to add is the warning that Alderson, Krahne & Stansfield (1987) gives when saying that a placement test is considered as good as long as it includes the same language content that is taught in the instruction program in which the student is expected to be placed as well as usual and familiar tasks.

2.1.1.1.2 Achievement Test

Tests are also designed for evaluating learners about what they have learnt in a specific course taking into account the objectives of the course's program; these tests are well-known as achievement tests (Henning, 1987; Alderson, Krahne & Stansfield, 1987; Hughes, 1989; McNamara, 2000). In Henning's (1987, p. 10) words, “achievement tests are directly related to language courses, their purpose being to establish how successful individual students, groups of students, or the courses themselves have been in achieving objectives”. In addition, McNamara (2000) declares is that sometimes this kind of test is considered to be self-enclosed due to the fact that it is not related directly with the language used outside of the classroom; Rea (1985, cited in Brindley 1989, p. 12) calls it non-communicative performance which means that there is not an approach to real language communication but to the

skills that make communication possible. As a final point, it is important to mention that this type of test may be divided into final achievement test (those applied at the end of a course) and progress achievement test (those applied during); for example, during a course, teachers assess students with a portfolio or classroom work and participation, while at the end, a final exam is applied (McNamara, 2000).

2.1.1.1.3 Proficiency Test

The last type of tests to be mentioned, and the one that is the basis of this study is commonly known as proficiency test. This kind of test implies global measurement of the testee's language proficiency. In contrast to achievement tests, proficiency tests are usually applied without taking into consideration any institutional program. McNamara (2000) makes such a distinction referring to achievement tests as the past and proficiency as the future since the first one measures the results of teaching and the second one focuses on the future situation of language use without previous teaching reference; however, such distinction is sometimes considered to be a bit unclear (Brindley, 1989; Hughes, 1989).

In the last centuries, most proficiency tests have been commercialized worldwide with different purposes; most of times proficiency tests are applied to determine if the examinee is well prepared for tasks that require English such as getting a job, studying abroad or even getting a degree or certification (Alderson, et al., 1987). In some universities proficiency tests are considered as reference for entrance such as the British Council ELTS test in British universities, and the TOEFL in American universities (Baker, 1989).

2.1.1.1.3.1 Language Proficiency

As it was already mentioned, a proficiency test intends to measure language proficiency which basically means to perform language abilities. There has been a debate of the definition of this term, but most authors agree that proficiency has to do with a learner's ability to use language (Farhady, 1982, cited in Ingram & Wylie, 1993, p. 221). In this study, language proficiency will refer to a global consideration of competences regarding language usage in how, where and under what conditions it has been acquired, measuring performance with a testing procedure (Davies *et al.*, 2002). Moreover, the American Council on the Teaching of Foreign Language (ACTFL) presents two implications entangled to language proficiency which are progression and continuum; progression has to do with communicative growth which expresses levels in a hierarchical continuity of performance ranks, and continuum involves subsuming preceding levels (Galloway, 1987). In general terms, the importance of proficiency lies on the learners' capability to play an effective role in real life.

2.2 TOEFL Test

According to information provided by Wall and Hórak (June 2006) in a report from the Educational Testing Service (www.ets.org), in 1963 the National Council on the Testing of English as a Foreign Language which was formed through the cooperative effort of more than 30 public and private organizations concerned with testing the English proficiency of nonnative speakers of the language applying for admissions to institutions in the United States, developed the worldly known as the Test of English as a Foreign Language better known as TOEFL. In 1965, the ETS and the College

Board assumed joint responsibility for the program; and nowadays ETS wholly owns and administers the TOEFL test. It is clear to see that the TOEFL test emerged because of the big demand from foreign students to study in the U.S.A.

In the last century, many proficiency tests have been commercialized throughout the world being the United States and the United Kingdom the two principal countries that developed proficiency tests (Alderson, Krahne & Stansfield, 1987); the TOEFL is one of these tests which is the core of this research. The Educational Testing Service (ETS) declares that the TOEFL is a test that measures nonnative speakers' ability to communicate in the college or university classroom by using English whose results taken into consideration in admission processes at more than 8,000 institutions worldwide. Angoff and Sharon (1971) attempt to enrich the TOEFL definition by saying that:

The TOELF is intended to measure the English proficiency of foreign students applying for college admission in the United States. As such, it consists of items and item types addressed to the linguistic problems of nonnative speakers of English, and it is designed to assess the degree of facility with those nuances of English that seem to cause foreign students difficulties in pursuing college studies (p. 129).

Thanks to these definitions it is possible to have a clearer idea about what TOEFL test is as well as its main purpose. However, there is an implicit hypothesis in the design of this test that states that while TOEFL may differentiate adequately among foreign students and may identify their English language proficiency, it will not accomplish the same purposes for native English speakers (Angoff & Sharon, 1971), and Clark (1997) confirms it in his study when concluding that the TOEFL is not

psychometrically appropriate for discriminating among native speakers of English regarding English language competence.

2.2.1 Importance of the TOEFL test

As English started becoming an international language, it was important to have a standard test that could measure the English proficiency of nonnative English speakers, which led to the development and implementation of the TOEFL. This test is considered to be usual because it is the most researched and used test (Douglas, n.d.). According to statistics provided by the ETS, almost a million people take the TOEFL test yearly to show their English language proficiency, and over 27 million people worldwide have taken the TOEFL test since it was administered for the first time. Hence, it is clear to see that the TOEFL test has become so important due to the fact that:

- it is the most widely available test of English language which means that there are over 4500 testing center of TOEFL in over 165 countries (Douglas, n.d).
- it is an important requirement in nonnative English student who want to study in an English-speaking country (Douglas, n.d., ETS 1986, 1989, 1997, 2012, Kim & Lee, 2006; Bachman, Davidson & Foulkes, 1990).
- many people take it because it is a requirement for getting a job (Griffin 2004, Bachman, Davidson & Foulkes, 1990) or to demonstrate proficiency in English to a company for which the examinee works.
- it is accepted by more than 8,500 universities and colleges in over 130 countries
- it is reliable for federal agencies and institutions.

- it is used by immigration departments to issue residential and work visas (Duran *et al.*, 1985).
- it is used by medical and licensing agencies for professional certification purposes.
- it is used by some universities or colleges as a requirement to get graduated (Boyd *et al.* 2007, Griffin 2004, Hernandez *et al.* 2005, Bachman *et al.* 1990).
- it is used for becoming licensed to practice a profession.

2.2.2 Types of TOEFL

Since 1963, the TOEFL has had different versions and changes. The ETS has published lots of research, and in the volume 6 of the 'TOEFL® iBT Research Insight Series' called 'TOEFL Program History', it is stated that changes in the theories of language proficiency were the main factors that motivated the redesign of the test's construct and content. As a result, the ETS developed four main types: the TOEFL® pBT, the TOEFL® cBT, the TOEFL® iBT, and the TOEFL® ITP. Other tests that are also administered by this institution are the TOEFL® Junior Test, the Test of Written English (TWE), the Test of Spoken English (TSE), the Speaking Proficiency English Assessment Kit (SPEAK® Kit), and the Secondary Level English Proficiency Test (SLEP® Test). The difference among all these tests is that the first tests mentioned are especially for higher education students.

2.2.2.1 Paper-Based TOEFL (TOEFL® pBT)

The ETS affirms that The TOEFL® pBT, better known as the 'Standard' TOEFL, is administered in a paper-based format, which estimates your capability to use and

understand English in a classroom within a college or university level. The aspects measured are Listening Comprehension, Structure and Written Expression, Reading Comprehension, and the writing skill. Besides, whoever wants to take the TOEFL[®] pBT test is required to take the Test of Written English (TWE) which is a thirty-minute test in which the testee has to write a short composition or essay about a specific topic. The purpose of the TWE is to measure the examinee's ability to write in English, i.e., it evaluates how the test takers generate, organize and support their ideas as well as the usage of standard written English formats.

2.2.2.2 Computer-Based TOEFL (TOEFL[®] cBT)

The ETS introduced the computer-based version of the TOEFL in 1998. It was the first time that ETS used electronic technology to test more complex skills such as writing, listening and reading. As a result of many institutions requesting for a section that could measure the writing skill of English as a foreign language, a new section that consisted in writing an essay was added to the TOEFL[®] cBT to evaluate such skill. In the listening and reading sections, new and more complex questions were added to these sections. In the listening skill there were some visual effects to improve this part of the test. The scores were valid for two years from the testing date. However, the TOEFL[®] cBT is not officially longer available since September 2006.

2.2.2.3. Internet-Based TOEFL (TOEFL[®] iBT)

In 2005 the ETS introduced the TOEFL[®] iBT to the world with a communicative competence perspective, that is, the ability to use the language in relevant contexts.

The TOEFL® iBT test is administered via Internet and its purpose is to measure the ability of nonnative speakers to use and understand English at the university level. This test has four sections that evaluate each skill; reading, listening, speaking and writing. The most common modern-day keyboard layout QWERTY is used in the TOEFL® iBT.

2.2.2.4 Institutional TOEFL (TOEFL® ITP)

According to information provided by the ETS, the Institutional TOEFL or TOEFL® ITP (Institutional Testing Program) is a paper-based test that measures the English proficiency of nonnative English speakers. This kind of test is used for placement, progress, evaluation, exit testing and other situations that do not require a secure testing environment, and it is administered by an institution and it does not replace the need for the official TOEFL test. The scores are established according to the Common European Framework of References for Languages (CEFR) and test takers are provided with a certificate of achievement. It is important to emphasize that the TOEFL® ITP is just nationally recognized, i.e. it does not have an international validation. Finally, it is important to mention that the Institutional TOEFL has two levels of evaluation. The level 1 is a two-hour test that measures the English language proficiency from an intermediate to an advanced level, which is composed by 140 questions; in contrast, the level 2 is a 70-minute test that measures the English language proficiency from a high beginning to an intermediate level, and which consists of 95 questions.

2.2.3 TOEFL's Sections

Apart from developing different versions of the TOEFL, the ETS made various changes with respect to the design of each version. At the beginning, the common sections in the TOEFL were five, but in 1976, the ETS redesigned and introduced TOEFL tests with three sections: Listening Comprehension, Structure and Written Expression and, Vocabulary and Reading Comprehension. Six years before, the ETS started developing new sections to evaluate speaking and writing skills within the TOEFL having as a result the development of the Test of Spoken English (TSE) to assess the oral skill of test takers, and the Test of Written English (TWE) in 1986 to assess the writing skill. Nevertheless, according to information provided by the ETS (1986, 1989) in its workbook 'Understanding TOEFL', the actual structure of most TOEFL tests still consists of three sections.

2.2.3.1 Listening Comprehension Section

The EST designed this section to assess the ability of nonnative speakers to understand English as it is spoken in North America because all the content in this section such as sentences and conversations are samples of what TOEFL test takers might hear if they were speaking with native North American speakers. The combinations of sounds and words that are usually difficult for nonnative speakers and the grammar that is used in spoken English are included in the questions of this section.

The Listening Comprehension Section has three different parts: 'A', 'B' and 'C'. The first part consists on having the test taker to listen to 20 sentences one by one, each sentence having four options; the examinee has to choose one sentence out of

the four options that is closest in meaning to the sentence that s/he heard. In the second part, the testee has to listen to 15 short conversations between two people; at the end of each conversation, a third person asks a question about the conversation which has four possible answers but only one is correct. The last part consists on having test taker to listen to a long conversation, a short talk or a lecture about a variety of subjects; after the examinee has listened to it, s/he is asked from three to five questions related to what s/he has just listened to that have four possible answers but only one is correct.

2.2.3.2 Structure and Written Expression Section

In comparison to the Listening Comprehension Section, the structure and written expression section tests language in a more formal way; it measures the test taker's knowledge about structural and grammatical points used in standard written English. This section is divided into two parts. In the first one, there are fifteen incomplete sentences, each sentence having four options which can be words or phrases; the test taker has to select one option out of the four that best completes the sentence. In the second one, there are twenty-five sentences in which some words or phrases are underlined; in this case, the testee has to identify the one underlined word or phrase that would not be correct in standard written English.

2.2.3.3 Vocabulary and Reading Comprehension Section

The vocabulary and reading comprehension measures the ability to read and the knowledge of different kinds of words such as nouns, verbs, adjectives and adverbs that the examinee has. This section is also divided into two parts. In the first part,

thirty sentences or questions are presented having a word or group of words underlined; every sentence has four options that might be words or group of words that can substitute the underlined word in the sentence. The second part consists of thirty questions that can be answered by reading and understanding the passages provided on the test; the questions are about main ideas, secondary ideas, analogies and inferences.

2.2.4 Scoring the TOEFL

The ETS utilizes automated scoring means and human raters to guarantee TOEFL test takers an innate accuracy of their level of English proficiency. Even though automated scorings means have a big advantage over measuring linguistic features and ensuring consistent and quality scores, these methods do not assess the effectiveness of the language response and the appropriateness of its content. Consequently, the ETS uses human raters to attend a wider variety of features of the language that automated scores cannot measure, such as the quality of ideas and content as well as form. Besides that, the ETS states that there is no a standard passing or failing TOEFL score since every educational institution and agencies set their own TOEFL score requirements. Nevertheless, TOEFL scores are valid for two years after the test date and there is no limit to the number of times people can take the test. In addition to the final score, the testee receives feedback about their performance level and a brief description of common performance that test takers in the reported rank typically do.

2.2.4.1 Scoring the TOEFL® pBT

The ETS states that the scores of the TOEFL® pBT's sections are reported on a scale from 31 to 67/68 points, and then they are converted by statistical means to a number from 310 to 677. As it was mentioned before, the TWE is part of the TOEFL® pBT but its scores are reported and printed separately on a scale from 1 to 6. The score may contain two digits and can also be approved for example, 3.5, and 5.5; nonetheless, the score is not added to the TOEFL score.

TOEFL® pBT scores

Section	Score Range	
	Min.	Max.
Listening Comprehension	31	68
Structure/Written Expression	31	68
Reading Comprehension	31	67/68
Total Score	310	677

2.2.4.2 Scoring the TOEFL® cBT

The TOEFL® cBT's sections are reported on a scale from 0 to 30 points, and then they are converted by statistical means to a number from 0 as the lowest score to 300 the highest one.

TOEFL® cBT scores

Section	Score Range	
	Min.	Max.
Listening Comprehension	0	30
Structure/Written Expression	0	30
Reading Comprehension	0	30
Total Score	0	300

2.2.4.3 Scoring the TOEFL® iBT

The TOEFL® iBT is the only test that has four sections which are reported on a range from 0 to 30 points with 120 points as total score. In contrast to the other TOEFL tests, the points in each section are ranged by levels. The reading and listening scores may be high (from 22 to 30 items answered correctly), intermediate (from 15 to 21 items), or low (from 0 to 14 items). On the other hand, the speaking and writing sections are rated in good, fair, limited, and weak with different division of points.

TOEFL® iBT scores

Section	Score Range	
	Min.	Max.
Listening Comprehension	0	30
Structure/Written Expression	0	30
Reading Comprehension	0	30
Total Score	0	300

2.2.4.4 Scoring the TOEFL® ITP

Since the TOEFL® ITP is divided into two level tests, the three sections in each test are reported on different scales. The scores in the level 1 are reported from 310 to 677 points (first chart), and the scores in the level 2 may differ from 200 to 500 points (second chart).

TOEFL® ITP Level 1 scores

Level 1 (intermediate to advanced) - 2 hours				
Section	Questions	Time	Score Range	
			Min.	Max.
Listening Comprehension	50	35 min	31	68
Structure and Written Expression	40	25 min	31	68
Reading	50	55 min	31	67
Total	140	115 min	310	677

TOEFL® ITP Level 2 scores

Level 2 (high beginning to intermediate) - 1 hour and 10 minutes				
Section	Questions	Time	Score Range	
			Min.	Max.
Listening Comprehension	30	22 min	20	50
Structure and Written Expression	25	17 min	20	50
Reading	40	31 min	20	50
Total	95	70 min	200	500

2.2.5 TOEFL Reliability

The study 'The TOEFL Exam: How reliable is it?' by Denis Griffin (2004), from the ITESM campus Zacatecas, was carried out to provide evidence to the reader to doubt about the precision of the TOEFL to measure the English proficiency of English nonnative speakers. This study was comprised of a comparison of students' scores in the ITP TOEFL with the results of the same students in an exam elaborated by the author. The hypothesis of this study was "the format of multiple-choice options in the TOEFL comprises an artificial technique for examining students' ability in English and this indirect form of testing students does not represent a reliable means of evaluation their overall command of English" (Brown 1994, cited in Griffin 2004, p.4). In this study, the author supports her study by citing different authors that have wrote about language teaching and testing (e.g. Brown, 1994; Broukal, 1995; Harmer, 1991; Hughes, 1989). In this study, the author emphasizes Brown's idea about how nowadays the world is full of standardized, norm-referenced tests that are: timed, multiple-choice, tricky, long and artificial. The authors state that these characteristics are a clear description of what the TOEFL is really. To prove her hypothesis, the author proposed a more authentic English proficiency exam. This proposal consisted

of a designed exam to resemble the TOEFL except that it had open-ended charts for the Listening Comprehension Section cloze type questions for the structure and written expression section and open-ended questions for the reading section. However, the results from this exam were not very useful in order to prove her hypothesis because the reliability of the designed exam could be questioned due to the fact that it had less items than the TOEFL® ITP does; besides, the TOEFL® ITP tested indirectly or directly more language skills such as idioms, vocabulary, synonyms and conditional in the three sections. As the method of the author was ambiguous to test the TOEFL® ITP reliability, she decided to implement two other ways to test the TOEFL® ITP reliability, the first one consisted of two TOEFL® ITP exams administered to the same students within a two week period and these scores were compared. In the second one the author converted an exam with a format identical to the TOEFL® ITP into an exam with open-ended answers. Later the same exam would be given section by section to the same students but in multiple-choice format and the results would be compared with the open-ended answers exam. The results of this part of the study are still in progress and there is not a final conclusion. In sum, it is clear to see that, even though, the author of this study used different stages to attempt to test the reliability of the TOEFL® ITP, she did not get conclusive results to prove the hypothesis that supported this study: consequently, the TOEFL® ITP reliability cannot be discarded.

Although, the TOEFL is a precise and reliable test that measures the English proficiency as a foreign language, the ETS in the TOEFL – Test and Score Manual recognizes that “no test score is entirely without measurement error” because examinees’ scores are not perfectly consistent, due to many factors. The ETS (1997)

defines reliability as “the extent to which test scores are free from errors in the measurement process” (pp. 29, 30). The two most common statistical indices for test reliability are the reliability coefficient and the standard error of measurement. On the one hand, the reliability coefficient refers to sources of error that can be found from variations in the sample of tasks required by the testing instrument, or in the way that examinees respond during the course of a single test administration; the reliability coefficient is better known as a measure of internal consistency, this means that it focuses on a single point in time. On the other hand, the standard error of measurement (SEM) refers to the error inherent in a test score due to the imprecision of the measurement process. The ETS gives an example about this, if a group of people with the same English level were to take the TOEFL test. Despite, their equal English proficiency, every person would not get the same scores. Some of them could get a higher scores than the other some of them a lover scores and most of them would get the same score average that represent their English proficiency. Hence, the variety of the scores could be accredited to differences in motivation, attentiveness and to the particular items on the TOEFL test. That is why the SEM works as an index that shows how examinees that have the same English proficiency can be expected to vary.

2.2.6 The TOEFL and the CEFR

Since the Common European Framework (CEFR) has become very important in the field of Teaching and Learning Languages, test designers have decided to develop tests according to the CEFR which “provides a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across

Europe; it describes in a comprehensive way what language learners have to learn in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively” (CEFR, Council of Europe, 2001, p. 1). Therefore, according to Tannenbaum & Baron (2011), the TOEFL ITP recommended cut scores for A2 (basic user - wastage), B1 (independent user - threshold), B2 (independent user - vantage), and C1 (proficient user – effective operational proficiency) are 337, 460, 543 and 627 scaled points.

2.3 TOEFL Test Taker’s Perceptions, Characteristics and Performance

Scoring performance is a hard task to do in language testing since it is “the essential step to arrive at a measure, in addition to any qualitative, descriptive information obtained from the test takers’ responses” (Bachman & Palmer, 1996, p. 193). In other words, scoring is the process of transforming performance into quantitative or qualitative data. But language testing performance is always affected by the test taker’s perceptions and characteristics; that is why scoring also implies to take into consideration factors that have either a positive and/or negative effect on the examinees’ performance (Bachman, 1990; Kunnan, 1995; Bachman & Palmer, 1996) in order to achieve more informed construct validation results (Kunnan, 1995). As Bachman states:

The problems currently facing language testers have both practical and theoretical implications, and fall into two general areas. First is the problem of specifying language abilities and other factors that affect performance on language tests precisely enough to provide a basis for the test development and for the interpretation and use of test scores. The second problem is

determining how from language test behave as quantifications of performance

(p. 8).

Hence, one of the aspects to be considered in this study is the identification of test takers' characteristics since they may influence on the performance when applying the TOEFL. Researchers in the field have mentioned different test takers' characteristics regarding test administration and scoring since there is a concern "whether an individual's test performance can be interpreted as an indication of his competence, or ability to use language appropriately and affectively in non-test context" (Bachman, 1990, p. 11). The most common characteristics are communicative language ability, test methods facets, personal characteristics, random factors (Bachman, 1990), affective schemata, topical knowledge (Bachman & Palmer, 1996), and attitude, anxiety and motivation (Kunnan, 1995). Random factors are made up of correlations among components of factors that are going to be presented.

2.3.1 Communicative Language Ability

TOEFL test takers' communicative language ability is the first aspect that predominantly influences performance or scores when taking the exam and it refers to the capacity of using language competence appropriately in the test. According to Candlin (1986, cited in Bachman 1990, p. 84), communicative language ability is:

the ability to create meanings by exploring the potential inherent in any language for continual modification in response to change, negotiating the value of convention rather than conforming to established principles.

In sum,... a coming together or organized knowledge structures with a

set of procedures for adapting this knowledge to solve new problems of communication that do not have ready-made and tailored solutions.

In addition, Bachman (1990) proposes three important components of communicative language ability which are language competence, strategic competence, and psychological mechanism. Language competence refers to “a set of specific knowledge components that are utilized in communication via language” (Bachman, 1990, p. 84) which is divided into organizational and pragmatic competence; organizational competence has to do with the formal aspects of language like grammatical competence and textual competence whereas pragmatic competence regards functional aspects of language such as illocutionary and sociolinguistic competence (Kunnan, 1995). Strategic competence is defined as “the mental capacity that relates language competence, or knowledge of language to the language user’s knowledge structures and the features of the context in which communication takes place” (Bachman, 1990, p. 107), involving assessment, planning and execution of language functions in a specific contextualized situation in order to assure effective communication. Finally, psychophysiological mechanism refers to “neurological and psychological process involved in the actual execution of language as a physical phenomenon... been the channel and mode in which competence is implemented” (Bachman, 1990, pp. 84, 108).

2.3.2 Test Methods Facets

Test performance may also be affected by the characteristics of the methods used which are interconnected with the features or ‘facets’ of the context of the testing

situation (Bachman, 1990). These aspects have to do with the way the test is planned to be administered.

2.3.2.1 Facets of Testing Environment

The first facets are related to the characteristics of the testing environment which may make testees perform differently; such facets are familiarity of the place and equipment, personnel, time of testing, and physical conditions. “Test takers tested with familiar equipment such as paper and pencil or tape recordings might perform better than those tested with unfamiliar equipment, such as computer” (Bachman, 1990, p. 118). For instance, with the introduction of the TOEFL® cBT in 1998, many researchers were concerned that the computer proficiency would have a negative effect on language proficiency. For that reason, Kirsh, Jamieson, Taylor, and Eignor’s (1998) carried out a study in which general TOEFL test-taking population was interviewed about their computer familiarity when taking the test which could be low, moderate or high. 1,100 ‘low-computer-familiar’ and ‘high-computer-familiar’ examinees were identified and administered a computer tutorial. Finally, they took the TOEFL test again, and ‘high-computer-familiar’ testees obtained higher scores than those with ‘low-computer-familiar’. Kirsh *et al.* (1998) declared that there were little literature discussing the effects of computer familiarity on performance on computer-based language tasks” (p. 1) after analyzing four studies which, from their perspective, may not be considered generalizable to TOEFL’s population of international testees or to language proficiency test. This study led the authors to conclude that there was no relationship between the testee’s computer familiarity and the testee’s performance on the TOEFL® cBT; nevertheless, it was found that TOEFL

'high-computer-familiar' examinees usually have more opportunities for language, computer instruction and use.

In concern to personnel, Bachman (1990) states that examinee may perform differently when being administered by familiar or unfamiliar personnel as well as with a different number of individuals. In the case of the TOEFL, the test is administered by different personnel depending on the test centers and institutions affiliated to the ETS which is the case of many universities where the TOEFL is also administered. In test centers, test takers get in touch with supervisor or proctors, who are instructed to exercise extreme vigilance during a test administration to prevent examinees from giving or receiving assistance in any way (ETS, 1997). In contrast, the TOEFL in colleges, universities and institutions is administered most of the times by personnel known by the testees who at the same time are students from those institutions.

Test takers' performance may also differ from each other due to the time of testing, i.e. a test may be administered early or late in the day, after having eating or without having eating, etc. The ETS administered TOEFL tests internationally on regularly scheduled test dates through the Friday and Saturday testing program which consists of 12 dates yearly. Most universities and institutions administered the TOEFL test on dates convenient for them. Physical conditions also interfere with the expected performance which can be noise, temperature, humidity, seating arrangement, and lighting (Bachman, 1990).

2.3.2.2 Facets of Test Rubric

There are also facets of test rubric which have as purpose to let the test takers know what they are expected to do when taking the exam; these facets may be test

organization regarding silence, sequence and relative importance of parts, time allocation, instructions regarding language and channel, specification of procedures and tasks, and explicitness of criteria for correctness (Bachman 1990). After lots of research reported by the ETS since the first administration of the TOEFL, this institution has implemented a rigorous organization in order to avoid negative effects on performance. For instance, according to the ETS, each part of the TOEFL has been weighted equally with reference to a final score; however, the parts of the TOEFL test are timed separately and the testees do not have the opportunity to answer in the order in which they want. Also, the time given for the whole test as well as for each part affects performance even when the ETS has established parameter of times based on previous research. Finally, instructions are very important due to the fact that performance may reflect the testees' understanding of procedures to be followed as well as the kind of tasks to perform in the test. Unclear or inaccurate instructions and inadequate time allocation may be sources of test anxiety which is reflected in test performance (Madsen 1982, cited in Bachman, 1990).

2.3.2.3 Facets of Input and Expected Response

Facets of input and expected response may also have an effect on performance and they are related since they share two aspects: format and nature of language. Format refers to channel, mode, form, vehicle and language of presentation, identification of the problem, and degree of speediness; nature of language refers to lengthiness, and organizational and pragmatic characteristics mentioned before (Bachman, 1990).

2.3.3 Background or Personal Characteristics

Most TOEFL testees share related characteristics since they want to apply the test with the same or similar purpose but, at the same time, they are completely different; with respect to this idea, Bachman, Davinson and Foulkes (1990) carried out a study which provided information described on ETS reports that emphasizes the typical and general characteristics of TOEFL examinees around the world from the 1970's to the 1980's. The authors state that such examinees, who were more men than women, seemed to be undergraduate and graduate degree planners who have median ages, and have already taken the TOEFL test more than once. Background characteristics, also known as 'personal characteristics' (Bachman, 1990), are an essential part of any testee as a social individual but they are not part of the testee's language ability; however, they may still influence testee's performance in any test (Bachman & Palmer, 1996).

Many authors have provided a similar list of these characteristics. Bachman & Palmer (1990) provide a list which includes age, sex, nationality, resident status, native language, level and type of general education, type and amount of preparation or prior experience with a given test, cultural background, background knowledge and cognitive abilities. It is important to affirm that even when personal characteristics are considered to be either sources of error or test bias, there is evidence that proves the influence of personal characteristics on all the qualities of test's usefulness and, as a consequence, on language test scores.

2.3.4 Topical Knowledge and Affective Schemata

Language use with reference to the world in which the test taker lives may also have an effect on performance; this factor is called topical knowledge, and it is divided into two branches: homogeneous and varied topical knowledge. The first one occurs when test takers use any specific topical knowledge as a basis for demonstrating their language ability, and the second one occurs when the test takers do not control a single area of topical knowledge commonly but many different ones (Bachman & Palmer, 1996). Furthermore, sometimes test takers have already taken a test with bad experiences as a result which will lead to negative affective response to the same test or another one in different situations. This is called affective Schemata and it is defined as the “characteristics of language use task and its setting in terms of past emotional experiences in similar contexts” (Bachman & Palmer, 1996, p. 65), influencing testees’ way of processing and completing tasks.

2.3.5 Motivation, Attitude and Anxiety

Motivation, attitudes and anxiety are also factors that may affect positively and/or negatively on testees’ performance (Kunnan, 1995). On one hand, motivation is what makes a person to perform accurately in a determine test. Davies *et al.* (1999, p.123), recognize that “the higher the test taker’s level of motivation, the truer the reflection of ability shown by the performance, and hence the lower the amount of error (cause, for example, by lack of effort)”. In other words, motivation is important since it may lead to better results when performing; therefore, results obtained by unmotivated students are considered to be less reliable than those from students who are highly motivated.

On the other hand, attitudes may be considered as those predispositions that a test taker has in order to respond to a test positively and/or negatively. Attitudes are sometimes defined in different ways; however, Domino & Domino (2006, p. 127) define attitudes as a “predisposition to respond to a social object, such as a person, group, idea, physical object, etc.” which, according to Cardno (1955, cited by Domino, 2006; p. 127), “in particular situation, the predisposition interacts with other variables to influence the actual behavior of a person”. According to this author, attitudes may have three important purposes which are an emotional, behavioral, and cognitive component. In this research, the emotional and cognitive components are important since these terms imply how strong a test taker feels regarding his/her competences in order to perform accurately.

Apart from motivation and attitude towards a test, anxiety also determines performance (Kunnan, 1995). According to Davies *et al.* (1999), all the factors mentioned before (e.g. motivation, test format and items, time) define the level of anxiety which may cause a beneficial or detrimental effect; however, anxiety tends to be identified as “negative thoughts and perceptions about testing” (Poorman, Mastorovich & Molcan 2007, cited in Oermann & Gaberson, 2009, p. 79) which will be explained in the next section. To illustrate these three factors, a student who has to pass the TOEFL for graduating purposes (extrinsic motivation) and who feels sure about his/her abilities (attitude) with presence of low anxiety definitely may have good results.

2.3.6 Perceptions and Expectations about the TOEFL Test

When it comes to analyze results that test takers obtained in the test, it is important to take into account all the perceptions and expectations that they have about the TOEFL since this also may influence motivation, attitude, level of anxiety, and consequently, performance. On one hand, perception simply refers to somebody's understanding and knowledge of determine event (Tinsley & Wescot 1990, cited in Xie, 2007). On the other hand, an expectation is considered to be a "feeling or belief about the way something should be or how someone should behave" (Longman Dictionary of American English, 2000, p. 265). Thus, a test taker may have a determine idea of how the test is applied, its content and the time allowed to answer it, as well as the possible results s/he might have. From this idea, it can be suggested that testees have different perceptions and expectations when applying the TOEFL which will have an impact on their performance depending on whether they were right or wrong about what they thought the TOEFL is.

2.5 Chapter Conclusions

In this chapter, the necessary theoretical framework related to language proficiency in English, the TOEFL test, test takers' perceptions and characteristics, and the factors that may have positive and/or negative effects on their performance has been explored in order to be able to analyze the obtained data according to the methodology used presented in the following chapter.

CHAPTER III: METHODOLOGY

3.0 Introduction

This chapter will include a clear description of the methodology that will be used in order to undertake this research. Information about the setting, the participants, and the instruments will be provided as well as a gradual description of the data collection and the analysis procedures that will be followed in order to develop this study.

3.1 Describing the Research Setting and Subjects

This study will be carried out in the Faculty of Languages of the BUAP. It is important to mention that in this university there has been a transition between two educational models, from the Proyecto Universitario Fénix (PUF) that started in 2002 to the Modelo Universitario Minerva (MUM) implemented since 2009. LEMO students are under the PUF educational model and LEI students are under the MUM educational model. On one hand, the PUF establishes for LEMO students to take eight English courses best known as Target Language (TL); on the other hand, the MUM stipulates that LEI students have to take five TL courses and 4 workshops that can be considered as English courses.

The participants for this study are 16 students: eight undergraduate LEMO students who have already finished the eight target language courses established in their curriculum, and eight undergraduate LEI students who have already finished the five target language courses and four workshops that are established in their curriculum. All these students are about to finish their curricular courses, to get graduated and to get their degree; consequently, they are considered to have an

advanced level of English proficiency which, in terms of the Common European Framework of Reference for languages, is B2.

3.2 Approach

Since the purpose of the research questions implies the exploration of the topic considering participants in their natural setting (Creswell, 1998), a qualitative research was carried; however, this study is also considered to be quantitative since it makes emphasis on the description and interpretation of data. Moreover, this study was exploratory and descriptive to understand better the issue being studied. A qualitative research can be carried out in different ways. From this idea it follows that this research was a case study since the number of participants are just few members of a big community; therefore, no generalizations will be made.

3.3 Describing Instruments

The Instruments that will be employed for this study are: a diagnostic TOEFL[®] ITP test and a questionnaire regarding participants' characteristics and perceptions about this test. Since this study concerns an issue that deals with a proficiency test applied to a specific people, a test method will be used. This method is considered to be a "powerful method [...] for gathering data of a numerical rather than verbal kind" (Cohein, Manion & Morrison, 2007, p. 414). For that reason, a three-section diagnostic TOEFL[®] ITP test will be applied in order to collect quantitative data about participants' language skills as well as strengths and weaknesses. Such diagnostic test is taken from *The Heinemann ETL TOEFL[®] Preparation Course* book published

and printed by McMillan Heinemann Publisher which resembles the TOEFL® ITP in time, structure, instructions, type and number of tasks, and scoring method.

On the other hand, a survey method will be used in this study due to the fact that this method determines “people’s opinion, attitudes and perceptions about the situation being studied” (Criollo, 2003, p.16) by means of questionnaires or interviews. According to Nardi (2006, p. 67), “a key element in the achievement of reliable and valid information in survey research is the construction of well-written and manageable questionnaires” which are useful instruments that provided structured and numerical data with the characteristics of being administered without the researcher’s presence and relatively direct to be analyzed (Cohein, Manion & Morrison, 2007, p. 317). Therefore, a questionnaire was applied to the participants.

The questionnaire was adapted from the background questionnaire used by Kunan (1995) in his research *Test taker characteristics and test performance: a structural modeling approach*; moreover, some other questions were added with the purpose of gathering information in order to achieve the purpose of this research (See Appendix B). Such questionnaire consists of four parts which have 16 questions in total; the first part focuses on obtaining participants’ personal characteristics, the second part focuses on gathering information regarding their perceptions towards the TOEFL® ITP test, the third one focuses of information concerning their background, and the last one focuses on gathering important information concerning participants’ performance. From the sixteen questions, seven questions are quantitative close-ended questions, five multiple-choice questions are written in charts in which participants have to tick according to what they want to answer and four questions are qualitative open-ended questions. Close-ended and multiple choice questions were

chosen due to the fact that respondents are more likely to answer them, it takes less time and interpretation becomes more reliable; however, these kinds of questions allow fewer variations in participants' responses (Nardi, 2006). The purpose of the questionnaire was to obtain qualitative reliable information about participants' perception towards the TOEFL® ITP test before they have taken the test, and their characteristics and factors that have either a positive and/or negative effect on the TOEFL® ITP scores in order to draw conclusions regarding their influence on the performance.

3.4 Data Collection Procedures

The three-section diagnostic test taken from *The Heinemann ETL TOEFL® Preparation Course* book was applied to participants at one sitting with a duration time of approximately two hours. Each participant was asked to write his/her ID in order to protect their identity. Then, the diagnostic tests were scored. After scoring tests and letting participants know their final score as well as their strong and weak areas, the questionnaire was provided to each participant in order to answer it. Finally, information gathered from questionnaires was analyzed.

3.5 Data Analysis Procedures

The scores obtained from the three-section diagnostic test and the answers obtained from the questionnaire were analyzed. In order to obtain the score, the *Diagnostic Test Answer Key* from *The Heinemann ETL TOEFL® Preparation Course* book was followed on pages 548 and 549. Then, correct answers in each section were counted and converted into a TOEFL score by using the *TOEFL Score Conversion Tables* on

page 505. In order to know the participants' strengths and weaknesses, the *Diagnostic Test Scoring Instructions* on page 503 were followed; correct answers in the first section were divided into the total of number of the questions in that section to know the percentage score of that same section; for the other two sections, it was followed the same procedures. In order to know the final converted score, the scores of the three sections were added, multiplied by ten, and divided into three. After that, such information was written in a chart which later was given to the participants according to their ID. Finally, all information regarding scoring was into computer files and tables.

In order to analyze answers from the questionnaire, the answers to close-ended questions and charts were entered in Excel for processing and analyzing. The answers of each close-ended question and each chart were counted and turned into graphs in order to illustrate findings obtained to make sense of the data collected. Then, open-ended questions were categorized in tables according to a theme. Finally, the results were analyzed, graphs were interpreted, and conclusions were drawn concerning participants' perceptions about the TOEFL® ITP, their performance and the factors that may have either a positive and/or negative effect on performance.

3.6 Chapter Conclusions

Although this study seems to be more quantitative than qualitative, it makes emphasis on the description and interpretation of data. Hence, this study is exploratory, descriptive and interpretative. The setting, participants, instruments, data collection and data analyses have already been presented in this chapter in order to give a

clear idea of the process followed to undertake this study. Next chapter will present the findings.

CHAPTER IV: FINDINGS

4.0 Introduction

In this chapter, the findings from this study will be presented and discussed deeply. Such results will be presented based on the research questions; thus, the chapter will be mainly divided into three sections. The first section will show the findings obtained from the three-section diagnostic TOEFL® ITP test. The second section will provide and analyze the results obtained from the questionnaire answered by the participants. The last section will present general conclusions drawn from the analysis.

4.1 Research Questions

The research questions that guide this work are:

1. What are the most common scores that undergraduate LEMO and LEI students get in the TOEFL® ITP examination?
2. What are the LEMO and LEI students' strengths and weaknesses in the TOEFL® ITP examination?
3. What are the student's perceptions about the TOEFL® ITP exam and the scores obtained?
4. What are the factors that influence positively and/or negatively LEMO and LEI students' scores in the TOEFL® ITP test?
5. What conclusions can be drawn about the results obtained by LEMO and LEI students in the TOEFL® ITP exam?

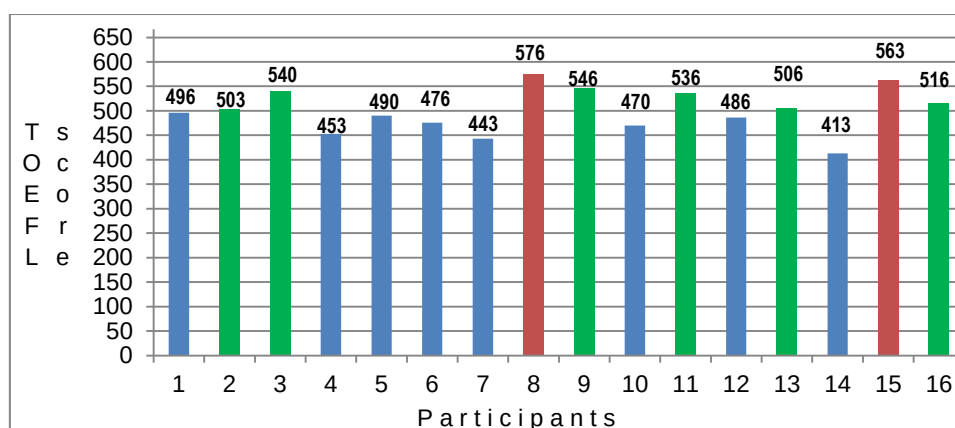
4.2 Participants' Performance

As the three-section diagnostic TOEFL® ITP test was applied to participants, the information obtained was analyzed to know the common scores obtained by the participants as well as their strengths and weaknesses following the next procedures. First of all, the answer sheets were graded as described in the previous chapter and the data gathered from each answer sheet was put into Excel tables and converted into graphs to show results clearly. The results found in the answer sheets will be presented and described in the following section.

4.2.1 Common Scores, Strengths and Weaknesses

First of all, it was essential to identify the scores obtained by participants in the test in order to answer the first research question. In this regard, the participants obtained the following scores, which are shown in Figure 1.

Figure 1: General Scores Obtained in the Diagnostic TOEFL® ITP Test

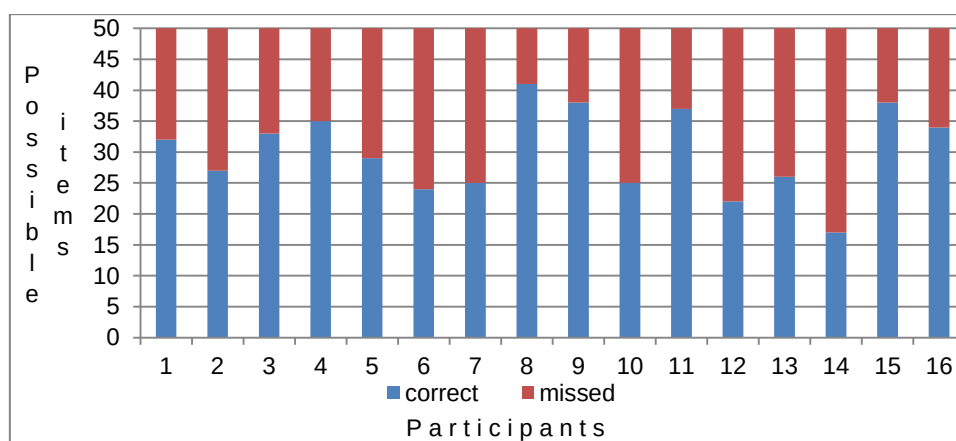


The results show that scores vary from 400 points to 580 points. From the 16 participants, eight participants (blue) obtained between 400 points and 500 points, six participants (green) obtained between 500 points and 546 points, and only two

participants (red) obtained more than 550 points. On one hand, the ETS states that there is not a passing or failing score since institutions were the TOEFL is required set the score required for passing it. Nevertheless, Britt (2009) affirms that getting less than 400 points in the TOEFL is considered to be a low score; furthermore, he says that many institutions, universities and enterprises consider a 550 score to be a good and acceptable score. Therefore, taking into consideration this information, it can be said that only two participants passed the TOEFL[®] ITP test with a good score.

In addition to have a clear idea of participants' common scores, it was also important to identify scores obtained in each TOEFL section to detect participants' strengths and weaknesses. In order to do this, scores obtained in each section, counting correct and missed answers were converted into tables and then into graphs to illustrate results. Figure 2 illustrates results from the Listening Comprehension Section of the test.

Figure 2: Listening Comprehension Section Scores

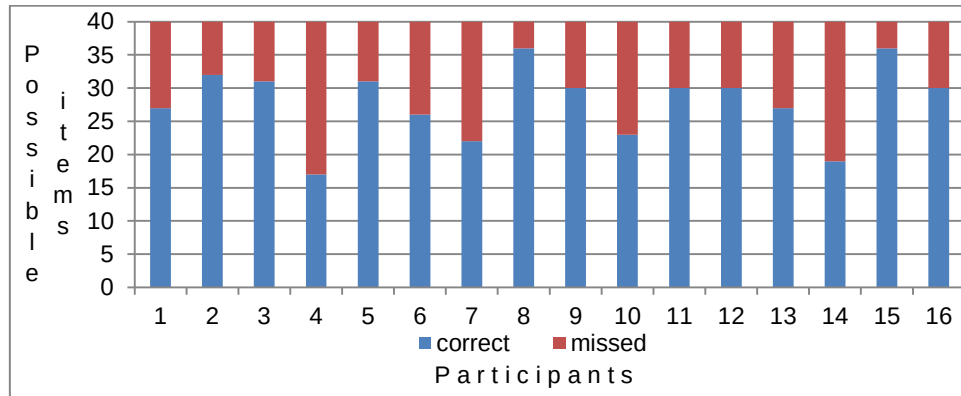


As seen in Figure 2, participants obtained different scores in the Listening Comprehension Section. In chapter two, the first section was said to have 50 possible items divided into 3 parts. In this case, most participants obtained between 25 and 41

correct answers, and only three participants got less than the half of the possible correct answers.

In the same way, Figures 3 presents the results from the Structure and Written Expression Section of the test.

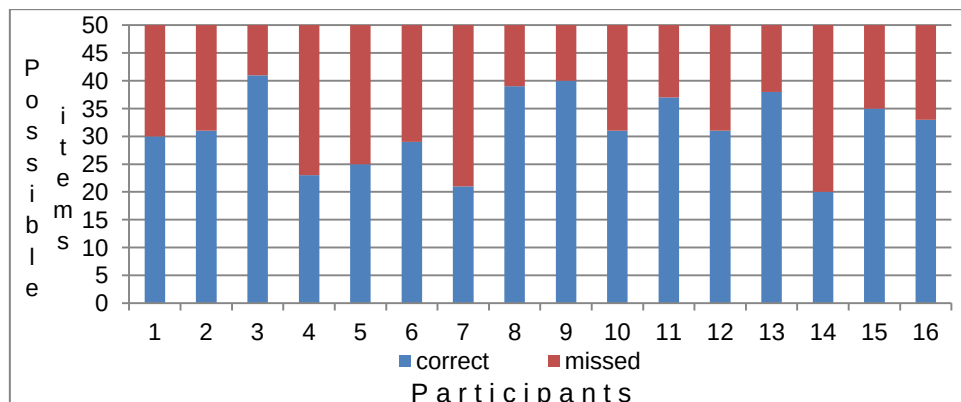
Figure 3: Structure and Written Expression Section Scores



The results show that the correct answers in this section vary from 17 to 36 correct answers from the 40 possible correct answers. It can also be observed that 14 participants out of 16 got more than 20 correct answers which is the half of correct answers.

Figure 4 shows the scores obtained in the Vocabulary and Reading Comprehension Section in the test.

Figure 4: Vocabulary and Reading Comprehension Section Scores



As seen in Figure 4, the scores obtained in the Vocabulary and Reading Comprehension Section are significant similar to those obtained in the Comprehension Section. Correct answers vary from 20 to 41 from the 50 possible items; 11 participants answered correctly between 30 and 41 items while two participants got between 25 and 30 correct answers which is the half of possible correct answers in this section. Only three participants obtained a score lower than 25 items correctly. To summarize this information, it is inferred that participants who obtained a general score equals or higher than 550 points also had a higher score in each TOEFL section.

Considering the previous information presented about results in the TOEFL sections, it is important to remember that the first and third sections of the test have 50 items and the second section has 40 items. In this regard, the strengths and weaknesses were obtained by converting the correct answers of each section into percentages which can be observed in Table 1. It is important to make clear that the following percentages represent only correct answers that each participant obtained in each section; each participant obtained a different percentage of correct answers per each section. The highest percentage is marked in blue and considered the strength of the test taker; on the contrary, lowest percentage is marked in color red and considered the weakness of the test taker, leaving one percentage in color white which is considered to be neither strength nor weakness but still a section to improve.

Table 1: Participants' Strength and Weakness

Participant	1	2	3	4	5	6	7	8
Listening Comprehension	64%	54%	66%	70%	58%	48%	50%	82%
Structure & Written Expression	67%	80%	77%	42%	77%	65%	55%	90%
Vocabulary and Reading	60%	62%	82%	46%	50%	58%	42%	78%
Participant	9	10	11	12	13	14	15	16
Listening Comprehension	76%	50%	74%	44%	52%	34%	76%	68%
Structure & Written Expression	75%	57%	75%	75%	67%	47%	90%	75%
Vocabulary and Reading	80%	62%	74%	62%	76%	40%	70%	66%

Considering the previous information presented about results in each TOEFL section, strengths and weaknesses vary among participant since each participant obtained three different scores in each section; therefore, there is no possible generalization. However, from a wide view, 11 participants out of 16 obtained more correct answers in the Structure and Written Expression Section which is marked in color blue, and it may be considered the mayor strength since the scores proved students' knowledge about structural and grammatical points used in standard written English viewed in the classroom. The Listening Comprehension Section and the Vocabulary and Reading Comprehension Section were observed to have similar low percentages; 8 participants out of 16 got the first section as their weakness while 7 out of 16 got the third section as their weakness. Hence, the first section may be considered as main weakness of the participants even though both sections need improvement.

The ETS states that the combinations of sounds and words, intonation and spoken structures used in the Listening Comprehension Section are usually difficult for nonnative speakers since topics are informal and conversational; in addition, test takers have to listen to mini-dialogues, longer conversations and talks which vary in length. The ETS also declares that in some cases the Vocabulary and Reading

Section becomes difficult since test takers have to develop a large recognition of vocabulary in English used in many different fields. In comparison to the first and third sections, the Structure and Written Expression Section tests language in a more formal way; it measures the test taker's knowledge about structural and grammatical points used in standard written English and topics presented in this sections are of a general academic nature.

It is important to mention that some participants had similar percentages in each TOEFL section; for example, participant 11 was about to get a score of 550 points having a balanced percentage of 74%,75%,74% in each section. In this case, it is difficult to identify whether this participant is doing good or bad in one of the skills.

4.3 Analysis of Results from the Questionnaire

After applying the diagnostic TOEFL[®] ITP test, a questionnaire was given to each participant in order to be answered. The information gathered from each section of the questionnaire was analyzed following a main procedure which consisted on putting information into tables and graphs. The first section provided information about the participants. The second one gathered data about participants' perceptions towards the TOEFL test before they take it. The third section focused on participants' background. Finally, the last section obtained data concerning participants' performance and factors that may have either a positive and/or a negative effect in it.

4.3.1 Personal Characteristics

Personal characteristics play an important role in this research since they may influence performance; therefore, participants were identified with regards to personal

characteristics mentioned in chapter two. From the 16 undergraduate participants, five were women and 11 were men. 80% of the participants ranked between 21 and 24 years old and the rest had between 30 and 36 years old. Furthermore, the half of participants was part of the Modelo Universitario Minerva and the other half was part of the Proyecto Universitario Fénix; this characteristic made participants differ in terms of curriculum. Even though personal characteristics seem not to have affected performance, it is important to keep in mind Bachman & Palmer's (1990) suggestion that there is evidence that not only considers personal characteristics as sources of error or test bias but also proves the influence of personal characteristics on all the qualities of test's usefulness and language test scores.

4.3.2 Participants' Perceptions towards the TOEFL® ITP Test

As it was formerly said, when it comes to analyze results that test takers obtained in a test, it is important to take into account all the perceptions that they have about the test since this also may influence motivation, attitude, level of anxiety, and consequently, performance, which basically refers to participants' understanding and knowledge towards the test (Tinsley & Wescot 1990, cited in Xie, 2007). Hence, it was fundamental to know firstly if participants had already had contact with the TOEFL. Consequently, the first section of the questionnaire, composed by two questions, asks for information about previous contact with this test as well as participants' perceptions.

Firstly, in the Faculty of Languages of the BUAP, a training TOEFL that presents the same type of tasks that a regular TOEFL® ITP does is applied at the DEPEA which is free and available for all students. Therefore, the first question asks

participants to provide the number of times they have taken the training TOEFL and the score they obtained the last time they took the test if their case; this information helps to better understand their perceptions before taking the TOEFL® ITP used for this study.

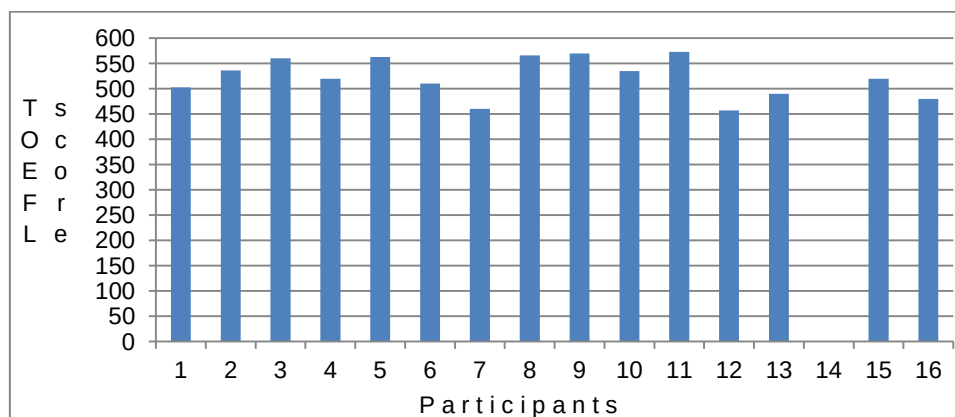
Table 2: Participants Having Taken the Training TOEFL Test

Times	0	1	2	3	4	5
Number of Participants	1	4	6	2	1	2

In Table 2, it can be observed the number of times participants have taken the training TOEFL test. Results vary from zero to five where zero refers to participants who have never had previous contact with the training TOEFL and 5 to participants who have taken it five times. To be more specific, from the total of participants, two participants have taken the test five times, one has taken it four times, other two have taken it three times, six have taken it twice and four have taken it once. Only one participant had never had contact with this test.

Secondly, the first question also asks participants' last score in the training TOEFL® ITP test in case they have taken it; Figure 5 illustrates data obtained.

Figure 5: Test Takers' Last Score in the Training TOEFL Test



In this regard, data obtained show that participants who have taken this training test obtained different scores. Scores vary from 450 to 573 points; however, it can be noticed that from 15 student who had taken the test, 5 participants got a score higher than 550 points, six participants got between 500 and 549 points and only four students obtained a score lower than 500 points. It is important to restate that sometimes the fact that test takers have already taken a test with bad experiences may lead to negative affective response to the same test or another one in different situations influencing testees' way of processing and completing tasks which is known as affective schemata (Bachman & Palmer, 1996),

Information gathered from the second question (Table 3) is interpreted as follow; the question contains six statements presented in a chart that participants had to tick according to what was more applicable to their perceptions. There were five possible answers.

Table 3: Participants' Perceptions Towards the TOEFL® ITP Test

Statement: Your beliefs before taking the TOEFL® ITP exam were that...	strongly agree	agree	undecided	disagree	strongly disagree
it is difficult	0% (0)	56% (9)	38% (6)	6% (1)	0% (0)
it takes less than two hours to answer it	13% (2)	44% (7)	13% (2)	31% (5)	0% (0)
it is applied on a computer	6% (1)	25% (4)	19% (3)	50% (8)	6% (1)
it is taken at one sitting	6% (1)	56% (9)	19% (3)	19% (3)	0% (0)
there is a break between each section	0% (0)	31% (5)	13% (2)	38% (6)	19% (3)
you could go back and forth between sections	13% (2)	19% (3)	25% (4)	25% (4)	19% (3)
<i>Percentage of participants (number of participants)</i>					

The first statement in the chart analyses whether the participants thought the TOEFL® ITP was or not a difficult test. As seen in Table 3, participants have different perceptions about the TOEFL® ITP difficulty; nonetheless, it is clear to see that more

than the half of participants thought this test is difficult. From the total of participants, nine participants agreed that the test is a difficult test while only one participant disagreed with this idea; the rest of participants did not agree nor did they disagree with this idea of test's difficulty.

The second statement in the chart focused on gathering information regarding knowledge of the time required to complete the test. According to the ETS, the TOEFL® ITP is a two-hour test composed by 140 questions; from this idea it follows that this test lasts no less than two hours. In this regard, participants were asked if they thought it would take less than two hours to complete the test to know if they knew the time required for the whole test and each section. Results show that participants' perceptions regarding time require for answering a real TOEFL® IPT noticeably diverge one from the others. On one hand, nine participants agreed with that fact that the test lasts two hours since they had already had contact with this test. On the other hand, five participants disagreed with the statement, and only two participants were not sure about the time required for completing the test.

The third statement aimed to gather information about the assumption of the TOEFL applied on a computer. As seen in the previous table, 6% and 50% of participants *strongly disagree* and *disagree* with this idea since they know that a TOEFL® ITP is a paper-based test. On the other hand, 6% and 25% of participants *completely agree* and *agree* with that statement which might be due to practice with the training TOEFL that implies the use of a computer equipment for answering a test that contains the type of tasks that a real TOEFL® ITP presents. Only 19% of participants were not sure for agreeing or not with this statement.

The fourth statement focused on the fact that this test is taken at one sitting; data demonstrate that most participants thought this statement to be true while three participants disagreed and other three were not sure. Consequently, it can be said that most of them knew that when this test is officially applied it has to be at one sitting; furthermore, this perception might have affected the performance of those who disagree or were not sure.

Since the test is taken at one sitting, the fifth statement concentrated on gathering information concerning participants' perceptions about breaks that could be given between sections when the test is applied. From the total of participants, six and three participants *disagree* and *strongly disagree* with this statement while only five participants *agree* with this idea; the rest of participants had no idea about having breaks when the TOEFL is applied. The ETS states that once test takers starts answering the test, they have to finish it so that the answers may be counted; *The Heinemann ETL TOEFL® Preparation Course* book supports what the ETS states and adds that if it is the case of a training or practice test, breaks are allowed. To summarize this information, it can be assured that for those who disagree or were not sure, the test became difficult and they became tired due to its length.

Finally, the purpose of the last statement from the chart was to collect data referring to going back and forth between sections. In this regard, four and three participants *disagree* and *strongly disagree* with the statement whereas two and three participants *strongly agree* and *agree* with it. Four participants were undecided about the statement. The instructions in the TOEFL® ITP are clear and precise as shown in the diagnostic TOEFL® ITP applied to participants; therefore, participants are not allowed to go back and forth between each section which was important information

that not all of participants knew. Taking into account all the analysis of perceptions, it can be said that results vary; however, those who have already had contact with a TOEFL test were more likely to have a higher score.

4.3.3 Participants' Background

Now that participants' perceptions are known, it is also important to collect information that regards with their background which is asked in the third section of the questionnaire; this section is composed by three questions and two charts. The first question asked participants to provide their age at the time they began to learn English formally, whether they had taken it at school or in a language institute. Results are illustrated in Table 4.

Table 4: Test Takers' Age to Start Learning English

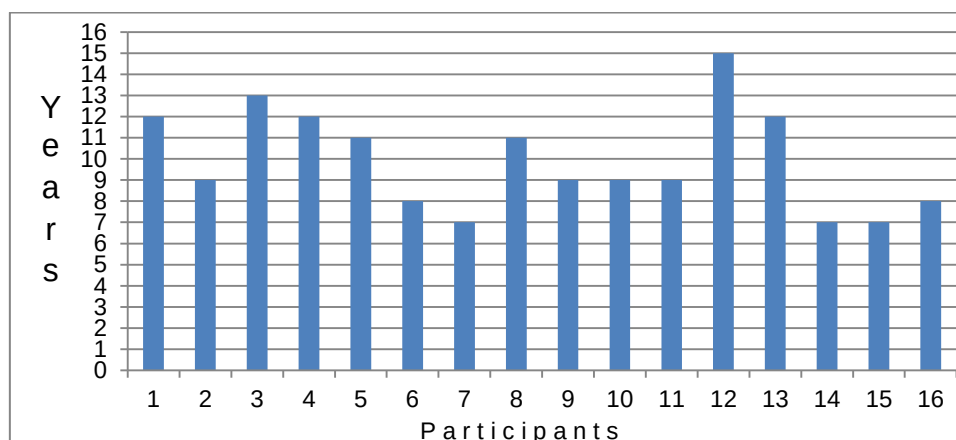
Ages	from 1 to 5 years	from 6 to 9 years	from 10 to 14 years	15 years or more
Participants	0	3	7	6

Despite the fact that there were four variables concerning participants' age to start learning English, two variables prevailed. It is easy to identify seven participants having between 10 and 14 years old at the time they started learning English; on the contrary, six participants affirmed to be 15 years old or older when they started learning English whose scores were lower than 500 points. Only three participants said to be between six and nine years old when they started learning English but only one from these students got a score higher than 550 points.

In addition to identify participants' age to begin learning English, it is important to know the time they have been in contact with the language; that is why question

number two from this section asked participants about the time they have formally studied this language which is illustrated in Figure 6.

Figure 6: Number of Years Participants Have Studied English



As seen, the number of years that participants claimed to have studied English formally vary from six to fifteen years; nevertheless, such information indicates that all participants have studied it for at least six years but it shows no negative effects on students' performance since most of those participants who claimed to have been studying English more than five years got lower scores.

In order to have a narrow view of the time they have spent studying English, it was necessary to ask participants about the hours they used to spend in English class during different levels of education they have gone through. Furthermore, five variables were given with respect to the amount of hours used in English class per week. In Table 4 results are presented.

Table 5: Participants Learning English in Different Educational Levels

Statement: The hours you used to spend in English class per week...	none	1 – 3	4 – 6	7 – 9	10 or more
in elementary school	63% (10)	19% (3)	6% (1)	6% (1)	0% (0)
in middle school	19% (3)	63% (10)	19% (3)	0% (0)	0% (0)
in high school	6% (1)	56% (9)	38% (6)	0% (0)	0% (0)
*in a language institute	50% (8)	19% (3)	13% (2)	6% (1)	13% (2)
<i>Percentage of participants (Number of participants)</i>					

According to information provided in Table 5, it is inferred that most participants had no contact with English until they enter high school. Only few participants had English classes in elementary school with different amount of time used per week. Moreover, it can be observed that as they advanced in the different educational levels, the hours used for English class increased; most participants said to have had classes in middle school, and the ones who denied it obtained a lower score. It is important to mention that the participant with the lowest score claimed not to have had classes of English at any educational level which clearly affected his/her performance.

Participants were also given the option to provide information about attending to a language institute in which they could have contact with English. Data obtained show that 50% of participants attended to a language institute with a different amount of time exposure to language per week; it is curious to see that this factor had a positive effect on the performance of the two students with the highest scores but not on the rest whose scores were lower than 500 points.

Contact with English in the past provides a notion of knowledge of English; however, usage of English used daily at the present time gives a clear idea about the development of the communicative ability. The following table presents data that

represents English as being used in participants' daily life in four different contexts where they spend most of their time: school, home, friends and in the street (e.i. in a store, park or restaurant).

Table 6: English as Being Used in Participants' Daily Life

Statement: The hours you really use English per day (<i>whether you speak, read, write or listen to it</i>)...	1- 2 hrs	3 - 4 hrs	5 - 6 hrs	7 - 8 hrs	9 - 10 hrs	11 or more
at school	13% (2)	50% (8)	19% (3)	19% (3)	0% (0)	0% (0)
at home	56% (9)	31% (5)	0% (0)	6% (1)	6% (1)	0% (0)
with your friends	81% (13)	13% (2)	0% (0)	0% (0)	6% (1)	0% (1)
in the street (i.e. in a store or other place)	81% (13)	19% (3)	0% (0)	0% (0)	0% (0)	0% (0)
Percentage of participants (Number of participants)						

Based on the information presented on Table 6, the amount of time spent by participants speaking, reading, writing or listening to English vary in each context. At school, 50% of the participants were said to spend between three and four hours while 19% of participants claimed to spent from five to six and another 19% from seven to eight hours using English. Only one or two hours were spent in English by 13% of participants. It is important to mention that participants counted the hours that they usually spend inside and outside of the classroom from which it can be inferred that most of them do not implement the usage of English at school as much as they could which might be due to different circumstances. However, this factor affected differently in each case since some participants with lower scores stated to use a lot of English and vice versa.

Describing the next context, it can be observed that 56% of the total of participants seem to spend between one and two hours using English at home, 31%

seems to spend from three to four hours and only 6% of participants spend from seven to eight and another 6% from nine to ten hours which seems to be not very favorable in one of these cases. In the next two contexts, most participants claim to spend from one to two hours using English which does not seem to have affected participants' performance comparing one among the rest

4.3.4 Participants' Performance

After analyzing participants' personal characteristics, perceptions and background, the analysis of the second section regarding performance is presented in order to identify factors that could have had either a positive and/or negative effect on their performance. As previously described, most of these factors are known as test method facets (Bachman, 1990) which have to do with the way the test is planned to be administered; such facet methods are facets of testing environment, facets of rubric, and facets of input and expected response.

The same author says that test environment are related to the characteristics of the testing environment which may make testees perform differently; such facets are familiarity of the place and equipment, personnel, time of testing, and physical conditions. That is the reason why the first question from this section focuses on the familiarization that participants had with the test once they had taken it. This question contains six statements from which it is inferred that most students seem to be familiarized with the test.

Table 7: Familiarization with the TOEFL® ITP Test

Statement: You were familiarized with the TOEFL® ITP's...	strongly agree	agree	undecided	disagree	strongly disagree
Sections	6% (1)	69% (11)	0% (0)	19% (3)	6% (1)
Instructions	0% (0)	75% (12)	13% (2)	6% (1)	6% (1)
Vocabulary	0% (0)	25% (4)	44% (7)	19% (3)	13% (2)
equipment (paper)	19% (3)	44% (7)	13% (2)	25% (4)	0% (0)
answer sheet	25% (4)	56% (9)	0% (0)	19% (3)	0% (0)
Time	6% (1)	38% (6)	6% (1)	38% (6)	13% (2)
type of tasks	0% (0)	63% (10)	19% (3)	13% (2)	6% (1)
<i>Percentage of participants (number of participants)</i>					

First, from the statement that emphasizes familiarization with the TOEFL sections, it can be observed that 11 participants out of 16 claimed to be familiarized with the sections of the test; in addition, another one strongly agreed with the fact of knowing the sections. Conversely, four participants denied being familiar with the sections presented on the test.

The next statement focused on familiarization with the instructions on each TOEFL section. In this regard, 75% of the total of participants agreed on being familiar with the instructions; on the contrary, 12% of participants said to be unfamiliar with them. Only 13% of participants were not sure about such familiarization.

After that, participants were asked to say whether they were familiar or not with the vocabulary used in the test. In this case, seven participants said to be not sure whether they were or not familiarized with the vocabulary used in the test, and three and two participants *agreed* and strongly agreed that they lacked of knowledge of vocabulary used. It can be observed that only 4 participants were familiarized with the vocabulary.

Then, participants provided information with respect to familiarization with the equipment used for the application of the test; 63% of the total of participants assumed to know what equipment would be used and 13% said not to be sure. Consequently, 25% of participants showed lack of familiarization with the equipment to be used. In Bachman's (1990, p.118) words, "test takers tested with familiar equipment such as paper and pencil or tape recordings might perform better than those tested with unfamiliar equipment, such as computer". The importance of familiarization with the equipment is as important as the answer sheet, which was asked in the fifth statement. As seen in Table 4, the agreement of four and nine participants shows that most of them were familiarized with the answer sheet while only three participants disagreed.

Finally, the last statement provides data about participants' familiarization with the type of task they have to do in the test. It can be perceived that 63% of the total of participants said to be familiar with the type of tasks used in the test whereas 38% and 13% of participants presented lack of familiarization with them. Only 1 participant who represents the 6% was not sure about knowing the tasks to be done.

The second question in this section had to do with the personnel in charge of applying the test giving directions; Figure 8 demonstrates results obtained regarding participants' opinion toward the instructions that were given by the personal in charge of applying the test.

Table 8: Clear Instructions Before the TOEFL® ITP Test Started

Clear instructions	strongly agree	agree	Undecided	disagree	Strongly Disagree
Participants	5	9	0	2	0

It can be seen that 5 participants out of 16 strongly agree that the instructions before taking the test were clear; in addition to this 9 more participants agree that the instructions were clear and concise. On the contrary, just 2 test takers disagree with the idea that the instructions were clear. Instructions are very important due to the fact that performance may reflect the testees' understanding of procedures to be followed as well as the kind of tasks to perform in the test; unclear or inaccurate instructions are reflected in test performance (Madsen 1982, cited in Bachman, 1990). To summarize this information, it can be stated that the instructions were clear enough thereby it is implied that this factor did not affect participants' performance in the test.

In the same way, the presence of the personnel administering the test could affect performance. For that reason, participants were asked about presence of anxiety produced because of the personnel's presence.

Table 9: Presence of Personnel Applying the Diagnostic TOEFL® ITP Test Causes Anxiety

Anxiety	strongly agree	agree	undecided	disagree	strongly disagree
Participants	0	3	3	10	0

As observed, Figure 9 demonstrates the analysis about the presence of personnel applying the test made test takers feel anxiety or not is shown. Ten test takers disagree that the presence of personnel made them be anxious and three participants are undecided about it. In contrast, three participants agree that the presence of personnel while they were taking the test made them feel anxiety. In concern to

personnel, Bachman (1990) states that examinee may perform differently when being administered by familiar or unfamiliar personnel as well as with a different number of individuals. In test centers, test takers get in touch with supervisor or proctors, who are instructed to exercise extreme vigilance during a test administration to prevent examinees from giving or receiving assistance in any way (ETS, 1997). In contrast, the TOEFL in colleges, universities and institutions is administered most of the times by personnel known by the testees who sometimes are students from those institutions as in this study. Finally, it can be said that for most of the participants the presence of personnel while they were answering the test did not have a negative influence in the performance of the test.

Anxiety tends to be identified as “negative thoughts and perceptions about testing” (Poorman, Mastorovich & Molcan 2007, cited in Oermann & Gaberson, 2009, p. 79); nonetheless, not only may anxiety influence the participants’ performance positively and/or negatively, but also external or internal factors present at the moment that participants are trying to concentrate in order to answer as accurately as they can. In the questionnaire, the tenth item was a chart focused on gathering information of such external or internal factors to identify if their presence influenced the results.

Table 10: Factors Presented at the Moment Participants Were Answering the Test

Statement: At the moment you answered the TOEFL® ITP test...	strongly agree	agree	undecided	disagree	strongly disagree
you were hungry	6% (1)	50% (8)	6% (1)	38% (6)	0% (0)
you were going through personal problems	6% (1)	13% (2)	19% (3)	56% (9)	6% (1)
you wanted to go to the restroom	0% (0)	31% (5)	13% (2)	56% (9)	0% (0)
you felt under pressure because of the time	38% (6)	25% (4)	19% (3)	13% (2)	6% (1)
there was a lot of noise	19% (3)	31% (5)	6% (1)	25% (4)	19% (3)
there were technical problems	6% (1)	19% (3)	13% (2)	50% (8)	13% (2)
you were nervous	13% (2)	6% (1)	31% (5)	44% (7)	6% (1)
<i>Percentage of participants (number of participants)</i>					

Firstly, participants were asked to provide information concerning hunger at the moment they were answering the test. According to Bachman (1990), test takers' performance may also differ from each other due to the time of testing, i.e. a test may be administered early or late in the day, after having eaten or without having eaten, etc. In this regard, it was found that nine participants claimed to be hungry at that moment while six participants were not; one participant said not to be sure about it. This factor was definitely reflected in scores of those who felt the sensation of being hungry.

Secondly, participants were asked whether they were going through personal problems or not, which might have influenced performance. Results clearly show that from the total of participants, 6% and 56% strongly disagreed and disagreed with the idea of going through personal problems while they were answering the test, 19% were not sure and 6% and 13% of participants strongly agreed and agreed.

The following statement asked participant whether they wanted or not to go to the bathroom. 56% of the total of participants denied wanting to go to the restroom and 13% were not sure; in contrast, 31% of participants affirmed that they actually wanted to go to the restroom.

Time is also a factor that may influence performance; that is why participants were asked whether the amount of time provided for completing each task and each section of the test made them feel under pressure or not. Results demonstrate that 10 participants agreed to have felt under pressure at the moment of answering due to the time provided; three participants disagreed and other three were undecided.

Noise is a factor that disturbs participants and deletes concentration, and consequently, most of times it affects performance. Bachman (1990) affirms that physical conditions also interfere with the expected performance which can be noise, temperature, humidity, seating arrangement, and lighting. Concerning this factor, eight participants state that there were a lot of noise and one participant was not sure. The rest of participants agreed that there was a lot of noise which affect their performance mainly in the Listening Comprehension Section.

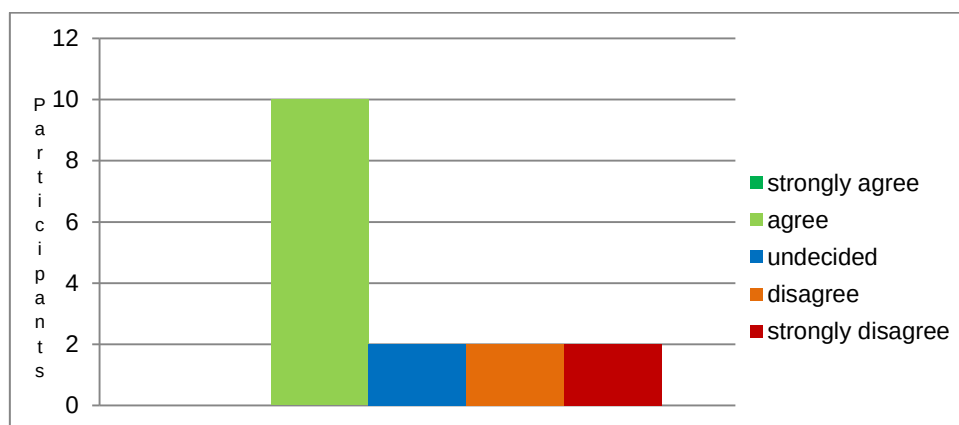
Regarding technical problems faced at the moment of answering the test, 10 participants denied facing those kinds of problems while four participants declared to have faced such problems.

Finally, Kunnan (1995) states that anxiety also determines performance; factors such as test format and time define the level of anxiety which may cause a beneficial or detrimental effect (Davies et al., 1999). Participants provided data that demonstrate that only three of them were really nervous at the moment they were answering the test and five were not sure about it. The rest presented no sign of

anxiety at the moment of answering the test from which it can be inferred that the ones who were nervous were more likely to be affected by this factor.

Another important factor that may affect performance is time. As seen previously, 67% of participants claimed to know that it takes no less than two hours to complete the test while only 44% of participants claimed to be familiar with this perception; these two variables were asked in previous sections. The following data was obtained from the fifth question from this section of the questionnaire which aimed to gather data about appropriateness of time used to answer each item in each TOEFL section.

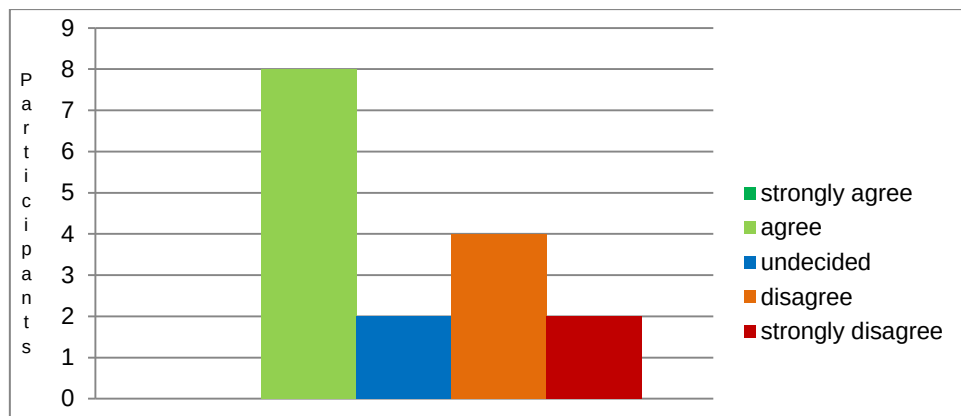
Figure 7: Appropriate Time for Answering Each Item in the TOEFL® ITP Test



It is plain to see in Figure 7 that most of participants strongly agree that the time was good enough to answer each item on the test and two participants were undecided about it. Conversely, two participants disagree and two more strongly disagree that the time was enough. Generally speaking, the majority of the participants considered that the time given to answer each item was suitable; therefore, the time for answering each item of the test was not a problem and it cannot be considered as a factor that could have a negative effect in the final score of the test.

In the same way, participants were asked about appropriateness of the volume of the audio in the Listening Comprehension Section which is illustrated in the following figure.

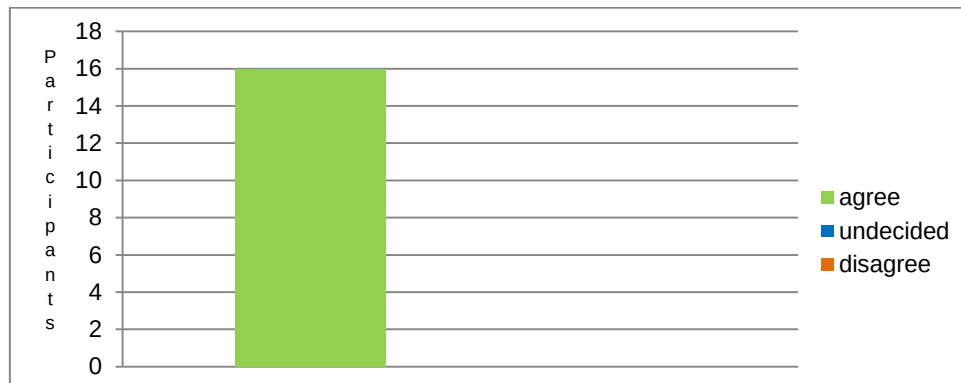
Figure 8: The Volume of the Audio in the First Section Was Appropriate (Loud Enough)



From the total of participants, a half agrees that the volume of the audio was good enough, and two participants are undecided about this issue. However, 4 participants disagree and two more strongly disagree that the volume of the audio was suitable for answering the Listening Comprehension Section. It is important to take into account that the volume of the audio also may be affected by noise.

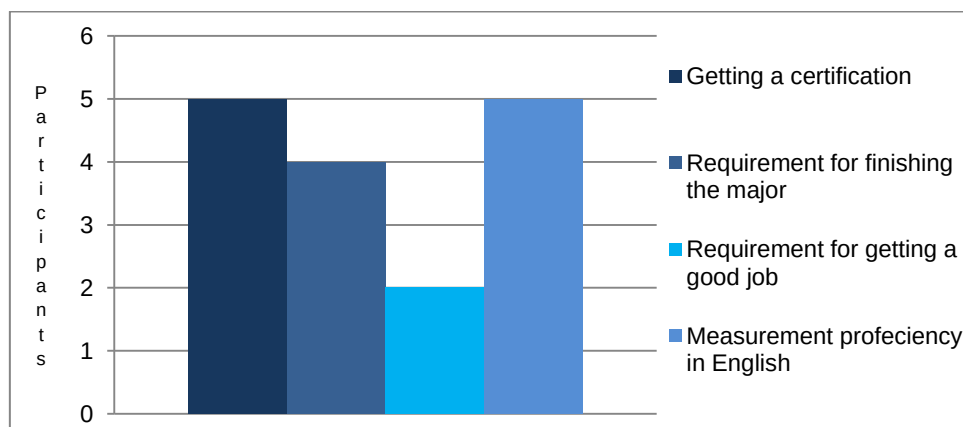
Another important factor to mention is motivation which is what makes a person to perform accurately in a determine test. Figure 9 presents data obtained from question thirteen that asked participants whether they are motivated or not to take the TOEFL[®] ITP in the future and Figure 10 specifies the reasons.

Figure 9: Being Motivated to Take the TOEFL® ITP Test in the Future



In the figure above, results show that most participants are motivated to take the TOEFL® ITP in the future while only one participant claimed not to be motivated to take the test while t. Davies *et al.* (1999, p.123), recognize that “the higher the test taker’s level of motivation, the truer the reflection of ability shown by the performance, and hence the lower the amount of error (cause, for example, by lack of effort)”. With respect to participants’ reasons for taking the TOEFL® ITP in the future, four variables were found: getting a certification, requirement for finishing the major, requirement for getting a good job and measurement of English proficiency.

Figure 10: Factors that Motivates Test Takers to Take the TOEFL® ITP in the Future

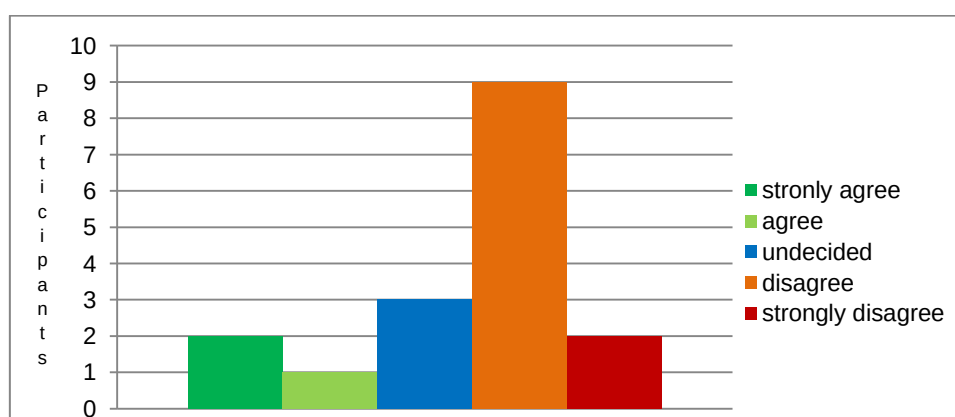


As seen, Figure 10 illustrates that five participants would take the test because they think that it is important for them to have a certification of the language, four

participants agreed that they would have to take the TOEFL® ITP because it is a requirement for finishing the major, two participants think that taking the TOEFL® ITP in the future will help them to find a good job and finally, five participants would take the test because this will help them realize about their proficiency in English. To summarized information presented in these two graphs, it can be concluded that participants were positively motivated which affected positively scores obtained. It is important to mention that three participants out of 16 were about to start a TOEFL preparation course which may lead to obtain high scores in the future.

As it was said throughout the study, the TOEFL is considered to be a good test that proves nonnative speaker's level of proficiency in English. However, some authors (Griffin, 2004) say this proficiency test seems to be tricky and therefore, it does not show test takers' proficiency. For this reason, participants were asked whether they agreed or disagreed with the fact that the TOEFL is the only way to demonstrate the English proficiency in the faculty of languages of the BUAP.

Figure 11: The TOEFL® ITP Test as the Only Way to Demonstrate English Proficiency

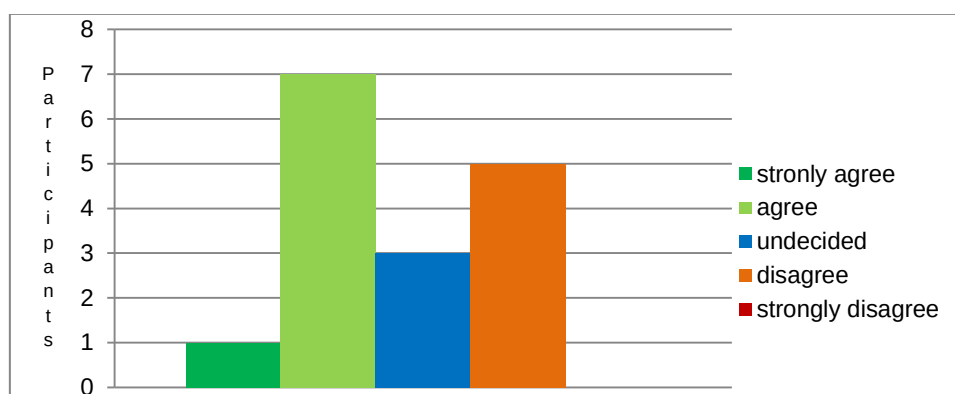


On one hand, only two participants strongly agree and one agrees with this fact, and three more are undecided about it. On the other hand, two participants strongly

disagree and nine disagree with the fact that the TOEFL is the only way to demonstrate their English proficiency. In a broader perspective, most of the test takers agree that the TOEFL should not be the only test of English to assess the English Language Proficiency.

Then, participants were asked to conclude by saying whether the final score of the test gives or not a clear idea of their English proficiency.

Figure 12: The final TOEFL® ITP Test Score Gives a Clear Idea of English Proficiency



Even when most participants were against with the idea of having the TOEFL as the only way to prove their proficiency level, one and seven participants strongly agreed and agreed that their final score in the test clearly shows their English proficiency. In contrast, three participants were undecided about this issue and five disagreed.

4.4 Chapter Conclusions

Having applied the two instruments in order to answer the research questions, the data gathered was analyzed following quantitative and qualitative procedures in order to display the results clearly to draw conclusions which will be presented broadly in the last chapter of this study.

CHAPTER V: CONCLUSIONS

5.0 Introduction

After having presented the findings of the research, important conclusions and significance of such findings are going to be presented; furthermore, limitations of the research and suggestions for further research are also going to be provided.

5.1 Findings

The present research is a case study that recounts the performance and perceptions of 16 undergraduate students from the Faculty of Languages of the BUAP; hence, no generalizations are expected to be done, only contributions to the area of evaluation and certification of English as a foreign language. As the main purpose of this study was to identify undergraduate students' most common scores in the TOEFL® ITP, perceptions and the factors that had either positive and/or negative effects on their performance, a report of the main findings based on participants' perceptions and performance is presented.

5.1.1 LEMO and LEI Undergraduate Students' Performance

Firstly, LEMO and LEI undergraduate students' common scores were identified and most students' performance was not the one expected since they are supposed to have a B2 level of proficiency in English according to the CEFR, which would be scores equals or higher than 543 (Tannenbaum & Baron, 2011). Scores obtained from the diagnostic TOEFL® ITP test vary from 400 points to 580 points. Hence, taking into account Britt's (2009) beliefs about considering a low score getting less than 400 points and a high score getting more than 550 points, it is concluded that

none of the participants got a low score in the diagnostic TOEFL test. However, from sixteen participants, only three got a B2 level of proficiency in English and two obtained a score higher than 550 points which is the score required for a variety of purposes such as studying abroad, getting a good job, getting a professional certification, practicing a profession or getting graduated which participants may be required in the future (Griffin 2004; Boyd *et al.* 2007, Hernandez *et al.* 2005, Bachman *et al.* 1990). In this regard, in order to get graduated from the Bachelor degree at the Faculty of Languages, LEMO and LEI students are given two options: defending a thesis project or presenting the TOEFL® ITP test administered at the DEPEA in case they have an average score of 8.5 at the end of their curricular courses; therefore, if these students decided to get graduated by presenting the test, only two participants would get their Bachelor degree.

Secondly, LEMO and LEI undergraduate students' strengths and weaknesses in the TOEFL® ITP test were also identified on the basis of their correct answers. It was observed that the majority of LEMO and LEI students obtained higher scores in the Structure and Written Structure section whereas more than the half of the participants obtained low scores in the Listening Comprehension Section and Vocabulary and Reading Comprehension Section; scores in the third section were similar to those obtained in the first section. Considering the results it is concluded that, in this case study, the Listening Comprehension Section and the Vocabulary and Reading Comprehension Sections were observed to be the most difficult part of the test, and therefore, students' weakness. As stated before by the ETS, this was due to combinations of sounds and words, intonation, spoken structures and audios' length used in the Listening Comprehension Section, which are usually difficult for nonnative

speakers since topics are informal and conversational. In contrast, the Structure and Written Structure section was the section in which students obtained more correct answers since this section tests language in a more formal way, measuring structural and grammatical knowledge of a general academic nature. Thus, it can be inferred that this was LEMO and LEI students' strength and that they have developed a strong basis of English grammar that will not interfere with their performance on a test taken in the future.

5.1.2 Students' Perceptions Towards the TOEFL® ITP Test

Furthermore, student's perceptions concerning the TOEFL® ITP test were determined on the basis of information concerning the times students have taken the training TOEFL applied at the DEPEA. Taking into consideration that perceptions in the field of testing refer to previous knowledge that a testee has to have before taking a test (Tinsley & Wescot 1990, cited in Xie, 2007), it is concluded that previous experiences could affect students' perceptions and performance since most of them had already taken a TOEFL test. From this idea it follows that LEMO and LEI students certainly know what a TOEFL® ITP consists of, its level of difficulty, the time provided to answer it, and its format and way of administration. In the cases in which perceptions about breaks during the administration of the test as well as permission to go back and forth between sections were found to be false, those perceptions are thought to have affected only to two or three students due to the fact that they obtained a score lower than 500 points.

5.1.3 Factors that Affected LEMO and LEI Students' Performance

As said before, there are many factors that may have either a positive and/or negative effect on test takers' performance. It was observed that some factors affected negatively participants' performance in the TOEFL whereas others influenced on performance positively. For instance, in this case study and in some other cases (Hernandez, R. et al., 2005) communicative language ability, the capacity of using language competence appropriately in a test (Candlin 1986, cited in Bachman 1990), was the first aspect that predominantly influenced students' performance; such factor was proved to have affected LEMO and LEI undergraduate students' performance when their background was analyzed. Their late contact with the language affected their performance; in addition, the lack of usage of English in their everyday life reflects their proficiency in English.

Furthermore, some methods facets of testing environment, test rubric and test input also affected negatively students' performance; those methods facets are time of testing and physical conditions. The organization regarding silence, sequence and relative importance of parts, time allocation, lengthiness and speediness are considered to be methods facets of test rubric and input (Bachman, 1990). In contrast, familiarity of the place and equipment, personnel, instructions and specification of procedures were factors that had a positive effect on students' performance.

5.2 Significance of Findings

The present study contributes with the field of evaluation and certification of languages. The data collected confirms the fact that many LEMO and LEI

undergraduate students are going through this issue. Hence, the study shows that many undergraduate students who are studying English in high levels have difficulties when they take the TOEFL® ITP test, since their perceptions regarding this test vary. Furthermore, the results reveal the importance of awareness of the factors that had positive and/or negative effects on testees' performance in students as well as in teachers and the institution. On one hand, students may overcome negative factors and improve their performance. On the other hand, the institution may implement a curricular course which may lead students to have a certification of the English language before they graduate.

5.3 Limitations of the Research

First of all, this study is considered to be a case of study since the number of cases studied was not enough in order to make any generalization about the issue been researched taking into account the whole community of the setting where it was carried out; the fact that the diagnostic test used as instrument lasts two hours made some prospective participants decline to contribute with this research. Secondly, it is important to recognize that it is unknown to what extend the results provided are reliable in comparison to students' reality; therefore, this study may be taken as the base for further research in the field of evaluation and certification of languages.

5.4 Suggestions for Further Research

The present study was just a part of a complete research that may be carried out in the future regarding the evaluation and certification of languages at the Faculty of Languages of the BUAP. On one hand, further research may be developed from

different perspectives that were not taken into account in this study following similar or different procedures which may lead to contributions in other areas; on the other hand, further research may be carried out by analyzing deeply one of the issues that were considered in the present research to make results be more reliable. For instance, a research concerning specific strategies (i.e. skimming and scanning in the reading section) used by students when performing may lead to know other factors that might affect students' performance when applying the TOEFL® ITP test. Another one may lead to look for the implementation of a curricular course in undergraduate students' curriculum to avoid the presence of negative factors in students' performance (i.e. familiarization). Another study may focus on analyzing the process in which low-score students obtain a required score. Finally, even though the ETS has carried out research for proving test usefulness features in the TOEFL, action research could be done by comparing students' proficiency using the TOEFL and another test to see to what extent students' proficiency may be revealed on the scores obtained in each test considering Griffin's (2004) research.

5.5 Reflective Account

The development of this study required a lot of work; nevertheless, the contributions reward the time devoted to research. The delimitation of the area to research was very difficult to choose since I was mainly interested in different areas such as curriculum design and learning styles and strategies. As I was about to get graduated, I got interested on taking the TOEFL® ITP test and that was how I got interested on the evaluation and certification of EFL. Furthermore, I attended to a MEXTESOL convention where I was given relevant information that confirmed the delimitation of

the research and helped me with the development of such study (See Appendix A). As stated in chapter one, passing a TOEFL® ITP with a score equals of higher than 550 points seemed to be a difficult task for many students currently studying a degree at the Faculty of Languages of the BUAP and, since I was still part of this community who also wanted to present this test, I was interested on knowing what were the reasons why this issue was affecting students which would may also affect me. Therefore, in order to better understand the issue being investigated, it was necessary to take a TOEFL preparation course which made me realized of many things that were useful for developing this study.

Developing the research questions was not easy due to lack of enough knowledge concerning evaluation and certification of languages. From this idea it follows that selecting and organizing the theoretical framework was very hard and exhausting; nevertheless, reading a lot of information made me be more interested on developing this study as well as learning more about the evaluation area.

After organizing a lot of information, determining the research methodology was easy but designing and applying the instruments were time consuming and complicated due to students' willingness to contribute with the study. Processing information using quantitative procedures and analyzing them was quite easy due to previous experience; however, presenting findings was difficult. Concluding was the last but not the least chapter to be presented which was not difficult since writing became an easy task. Revision and edition was also important in this process; feedback from my thesis advisor, committee advisors and friends was important for improving the whole research in format and content.

Finally, I could realized about the factors that may affect my performance at the moment of taking the TOEFL[®] ITP test; in addition, our perceptions with respect to this proficiency test changed completely since we were immersed in the study. Also, feedback given to all students that contributed with this study will help them be prepared in order to take the test and, at the same time, be aware of what may influence their performance.

REFERENCES

- Alderson, J. C., Krahne, K. J. & Stansfield, C. W. (1987) *Reviews of English language proficiency tests*. USA: TESOL
- Allison, D. (1999) *Language testing & evaluation: an introductory course*. Singapore: Singapore University Press.
- Angoff, W. H. & Sharon, A. T. (Jun 1971) A comparison of scores earned on the test of English as a foreign language by native American college students and foreign applicants to U.S. colleges. *TESOL Quarterly*, 5 (2), pp. 129-136. [Online]. Retrieved Oct 5th, 2012 from <http://www.jstor.org/stable/3585995>
- Bachman, L. (1997) *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Bachman, L., Davidson, F. & Foulkes, J. (1990) A comparison of the abilities measured by the cambridge and educational testing service EFL test batteries. In Douglas, D. & Chapelle, C. (Eds.), *A new decade of language testing research: selected papers from the 1990 language testing research colloquium* (pp. 25-45). Alexandria, VA: TESOL.
- Bachman, L., Davidson, F., Ryan, K. & Choi, I. (1995) *An investigation into the comparability of two tests of English as a foreign language*. Cambridge: UK.
- Bachman, L. & Palmer, A. (1996) *Language testing in practice*. Oxford, UK: Oxford University Press.
- Baker, D. (1989) *Language testing: a critical survey and practical guide*. Great Britain: Edward Arnold.

- Bastable, S. (2008) *Nurse as educator: principles of teaching and learning for nursing practice*. USA: Jones and Bartlett Publishers, Inc.
- Baxter, A. (1997) *Evaluating your students*. London: Richmond Publishing.
- Boyd, F., Byer, B. & Farre-Rigau, Y. (2007) *Getting to know standardized tests of English language: CELE UNAM language tests, ELASH, TOEFL® iBT*. ANUPI, A.C. [Online] Retrieved Noviembre 1st, 2012, from http://www.anupi.org.mx/PDF/07019_Panel.pdf
- Brindley, G. (1989) *Assessing achievement in the learner-centred curriculum*. Sydney: Macquarie University/NCELTR.
- Britt, G. (2009) TOEFL prep for Spanish speakers. [Online]. Retrived September 27, 2012. From: [http://books.google.com.mx/books?id=yQUZ7FPhq1AC&printsec=frontcover&dq=Britt+\(2009\)+Types+of+TOEFL&hl=es-419&sa=X&ei=10NiUePnIIXk2wXMioEQ&ved=0CC0Q6AEwAA#v=onepage&q&f=false](http://books.google.com.mx/books?id=yQUZ7FPhq1AC&printsec=frontcover&dq=Britt+(2009)+Types+of+TOEFL&hl=es-419&sa=X&ei=10NiUePnIIXk2wXMioEQ&ved=0CC0Q6AEwAA#v=onepage&q&f=false)
- Clark, J. L. D. (1977) *The performance of native speakers of English on the test of English as a foreign language* (TOEFL Research Report 1). Princeton, NJ: Educational Testing Service. [Online] Retrieved September 26th, 2012 from <http://www.ets.org/Media/Research/pdf/TOEFL-RR-01.pdf>
- Cohen, A. D. (1994) *Assessing language ability in the classroom*. New York: Heinle and Heinle.
- Cohein, L., Manion, L. & Morrison, K. (2007) *Research methods in education*. 5th Edition. NY: Routledge/Falmer.

Cook, G. (2003) *Applied linguistics*. Oxford: Oxford University Press.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press

Criollo, R. (2003) *The ultimate guide to writing a thesis in TESOL/AL*. Mexico: ACD.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. & McNamara, T. (2002) *Dictionary of language testing*. Cambridge, UK: Cambridge University Press.

Domino, G. & Domino, M. L. (2006) *Psychological testing: an introduction*. Cambridge: Cambridge University Press.

Douglas, K. S. (n. d.) Test of English as a foreign language. In Alderson, J. C., Krahnke, K. J. & Stansfield, C. W. (Eds.), *Reviews of English language proficiency tests* (p. 79). USA: TESOL.

Duran, R. P., Canale, M., Penfield, J., Stansfield, C. W. & Liskin-Gasparro, J. E. (1985) *TOEFL from a communicative viewpoint on language proficiency: a working paper* (TOEFL Research Reports 17). Princeton, NJ: Educational Testing Service. [Online] Retrieved November 2nd, 2012, from <http://www.ets.org/Media/Research/pdf/RR-85-08.pdf>

Educational Testing Service (1986, 1989) *Understanding TOEFL*. Princeton, NJ: Author.

Educational Testing Service (1997) *TOEFL test and score manual, 1997-98 edition*. Princeton, NJ: Author.

Educational Testing Service (June 2005, June 2006) *Test and score data summary for TOEFL® computer-based and paper-based tests*. Princeton, NJ: Author.
[Online] Retrieved November 1st 2012 from
<http://www.ets.org/Media/Research/pdf/TOEFL-SUM-0506-CBT.pdf>

Educational Testing Service (2012) [Online] Retrieved November 1st, 2012, from
www.ets.org

Galloway, V. (1987) From defining to developing proficiency: a look at the decisions.
In Byrnes, H. & Canale, M. (Eds.), *Defining and developing proficiency: guidelines, implementations and concepts* (pp. 25, 27). Lincolnwood, IL: National Textbook Co.

Gebhard, J. G. (1996) *Teaching English as a foreign or second language: a teacher self-development and methodology guide*. USA: The University Of Michigan Press.

Griffin, D. (2004) *The TOEFL exam: how reliable is it?* Paper presented at ANUPI, A. C. [Online] Retrieved November 2nd, 2012, from
http://www.anupi.org.mx/PDF/04005_DenisGriffin.pdf

Harris, M. & McCann, P. (1994) *Assessment*. Oxford: Heinemann Publishers.

Henning, G. (1987) *A guide to language testing: development, evaluation, research*. USA: Heinle & Heinle Publishers.

Hughes, A. (1989) *Testing for language teachers*. UK: Cambridge University Press.

Hernández, R., Salazar, J. F., Arellano, M.A., Rosales, L. & Garza, D. P. (2005, November 15) *TOEFL® institucional: análisis de los contenidos de aprendizaje*

evaluados en la prueba y análisis del desempeño de los alumnos del curso HI95-805 inglés académico. [Online] Retrived September 27th, 2012 from [http://www.mty.itesm.mx/rectoria/dda/rieee/pdf-05/33\(DHCS\).R.Hdz.J.F.SalazarA.ArellanoL.RosalesD.Garza.pdf](http://www.mty.itesm.mx/rectoria/dda/rieee/pdf-05/33(DHCS).R.Hdz.J.F.SalazarA.ArellanoL.RosalesD.Garza.pdf)

Ingram, D. E. & Wylie, E. (1993) Assesing speaking proficiency in the international English language testing system. In Douglas, D. & Chapelle, C. (Eds.), *A new decade of language testing research: selected papers from the 1990 language testing research colloquium* (pp. 220-234). Alexandria, VA: TESOL.

Institute of International Education (2011) Retrieved from www.iielatinamerica.org

Kim, M. H. & Lee, H. H. (2006) Determinants of TOELF score: a comparison of linguistics and economic factors. In *Korean and the World Economy V Conference, 7-8 July 2006*. Seoul, Korea: Korea University.

Kunnan, A. J. (1995) *Test taker characteristics and test performance: a structural modeling approach*. In Milanovic, M (Ed.) *Studies in Language Testing 2: Series*. Cambridge: Cambridge University Press.

McMillan Heinemann (1996) *The Heinemann ETL TOEFL® Preparation Course*. UK: Author.

McDonald, M. E. (2002) *Systematic assessment of learning outcomes: developing multiple-choice exams*. USA: Jones & Bartlett Publishers, Inc.

McNamara, T. (2000) *Language testing*. Oxford, UK :Oxford University Press.

Nation, I. S. P. & Macalister, J. (2010) *Language curriculum design*. USA: Routledge, Taylor and Francis Group.

- Nardi, P. M. (2006) *Doing survey research: a guide to quantitative methods*. 2nd Edition. USA: Pearson Education, Inc.
- Longman Dictionary of American English (2000) New Ed. Longman: USA
- Oermann, M. H. & Gaberson, K. B. (2009) *Evaluation and Testing in Nursing Education*. Third Edition. NY: Springer Publishing Company.
- Pachler, N. & Field, K. (2001) *Learning to teach modern foreign language in the secondary school: a comparison to school experience*. USA: Routledge, Taylor and Francis Group.
- Srivastava, D. S. & Kumar, S. (2005) *Education: assessment, evaluation and remedial*. India: Isha Books.
- Tannenbaum, R. J. & P. A. Baron (2011) *Mapping TOEFL® ITP Scores Onto the Common European Framework of Reference*. Princeton, NJ: Educational Testing Service. [Online] Retrieved November 10th, 2012 from <http://www.ets.org/Media/Research/pdf/RM-11-33.pdf>
- Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I. (1998) *The relationship between computer familiarity and performance on computer-based TOEFL test tasks* (TOEFL Research Report 61). Princeton, NJ: Educational Testing Service. [Online] Retrieved November 10th, 2012 from <http://www.ets.org/Media/Research/pdf/RR-98-08.pdf>
- Weir, C. J. (1993) *Understanding & developing language tests*. UK: Prentice Hall International.
- Xie, B. (2007) *Chinese international students' perceptions and attitudes toward seeking psychological counseling services: a qualitative exploration*.

Minnesota: University of Minnesota. [Online] Retrieved January 28th, 2013 from
<http://books.google.com.mx/books?id=h6qNRN-5o-AC&pg=PA13&dq=perceptions+about+the+TOEFL&hl=es&sa=X&ei=XMUGUZb0Nc2vygGb7YHADw&ved=0CEcQ6AEwAw#v=onepage&q=perceptions%20about%20the%20TOEFL&f=false>

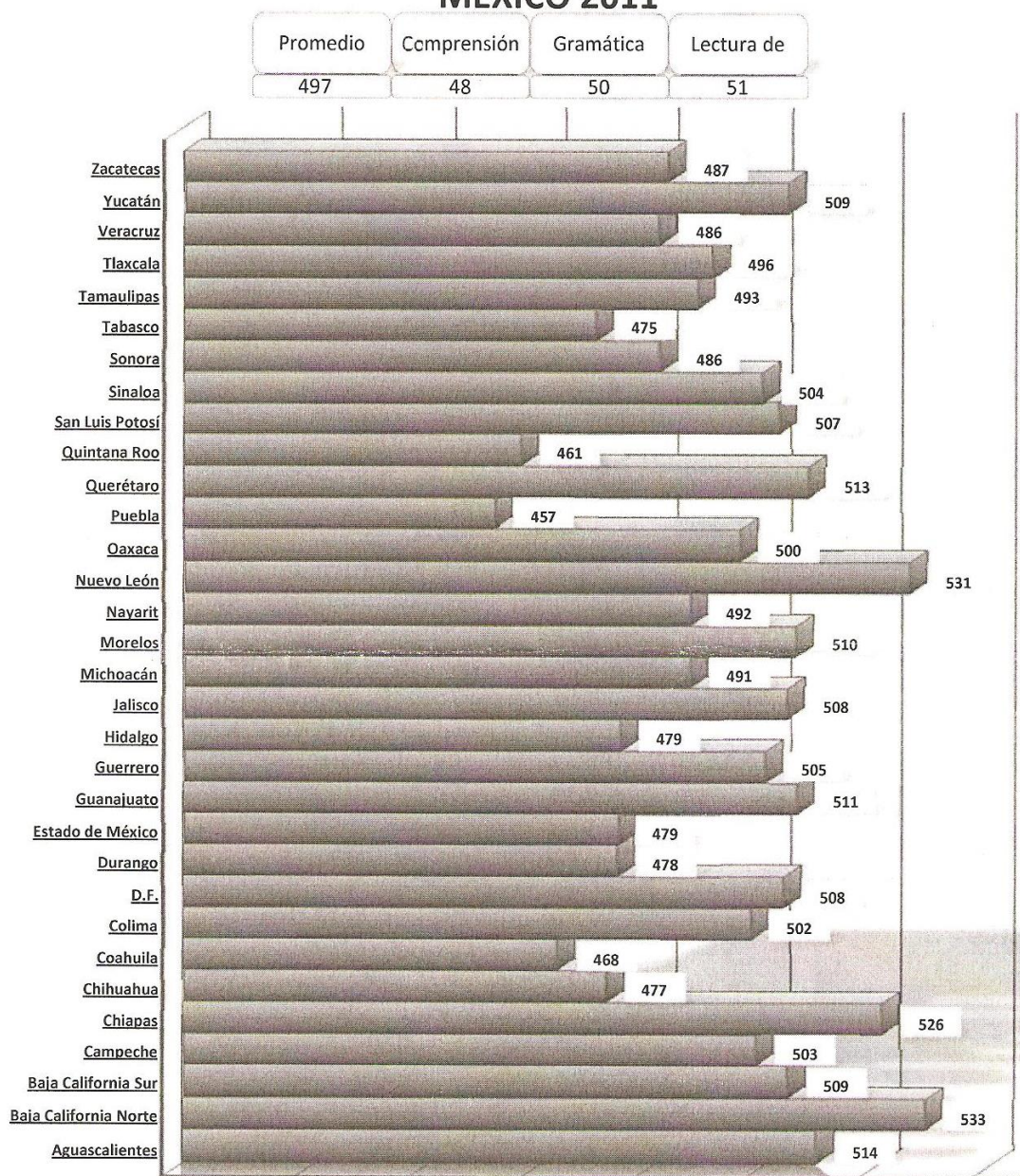
APPENDIX A:

TOEFL[®] ITP
ASSESSMENT SERIES

INSTITUTE OF
INTERNATIONAL
EDUCATION

Estadísticas Nacionales

MEXICO 2011



Porcentaje de alumnos por calificaciones obtenidas

Calificación	310-347	350-397	400-447	450-497	500-547	550-597	600-647	650-677
Porcentaje	1%	7%	15%	22%	27%	18%	8%	1%

Estadísticas generadas de + 145,918 exámenes TOEFL ITP administrados en México 2011.
Institute of International Education www.iielatinamerica.org

APPENDIX B:

TOEFL Test Taker's Questionnaire

This questionnaire is designed to provide us with information about participants' perceptions towards the TOEFL® ITP test as well as the factors that had either a positive or negative effect on their performance. Your answers to these questions will be kept strictly confidential. Please answer each question as accurately as you can. Thank you for your cooperation.

General Direction: Answer the following questions by ticking (✓), underlining, or providing specific information if the case.

Part I. Test taker's information

Age: _____ years old.

BUAP ID: _____ Current term you are in:

Gender: Male / Female

7th ☐ 8th ☐ 9th ☐ 10th ☐

Part II. Your perceptions towards the TOEFL® ITP

1. If you have taken the training TOEFL® ITP test:

A) Write the number of times you have taken this test: _____.

B) Write the last score you got: _____.

2. Your beliefs before taking the TOEFL® ITP exam were that...	strongly agree	agree	undecided	disagree	strongly disagree
it is difficult					
it takes less than two hours to answer it					
it is applied on a computer					
it is taken at one sitting					
there is a break between each section					
you could go back and forth between sections					

Part III. Background

3. When you first began to learn English formally (*whether you took English at school or in a language institute*), you were between...

A) 1-5 years old

B) 6-9 years old

C) 10-14 years old

D) 15 or more years old

4. The number of years you have studied English (*whether you took English at school or in a language institute – without taking into account the time you have studied English in the university*) is _____.

5. Choose and tick (✓) the hours you used to spend in English class per week...	none	1 – 3	4 – 6	7 – 9	10 or more
in elementary school					
in middle school					
in high school					
*in a language institute					

**just in case you have attended to a language institute.*

6. Choose and tick (✓) the hours you really use English per day (<i>whether you speak, read, write or listen to it</i>)...	Less than 2 hrs	3 to 4 hrs	5 to 6 hrs	7 to 8 hrs	9 to 10 hrs	11 or more than 12 hrs
at school						
at home						
with your friends						
in the street (<i>e.i. in a store or other place</i>)						

Part IV. Performance

7. You were familiarized with the TOEFL® ITP's...	strongly agree	agree	undecided	disagree	strongly disagree
sections					
instructions					
vocabulary					
equipment (paper)					
answer sheet					
type of tasks					

8. Before the test started, the personal in charge of applying the TOEFL® ITP test gave clear directions and specifications of procedures to be followed.

A) strongly agree B) agree C) undecided D) disagree E) strongly disagree

9. At the moment of taking the TOEFL® ITP test, the presence of the personnel who were in charge of applying the test made you be anxious.

A) strongly agree B) agree C) undecided D) disagree E) strongly disagree

10. At the moment you answered the TOEFL® ITP test...	strongly agree	agree	undecided	disagree	strongly disagree
you were hungry					
you were going through personal problems					
you wanted to go to the restroom					
you felt under pressure because of the time					
there was a lot of noise					
there were technical problems					
you were nervous					

11. You consider that the time used to answer each item was appropriate (enough).

A) strongly agree B) agree C) undecided D) disagree E) strongly disagree

12. You consider that the volume of the audio used in the Listening Comprehension Section was appropriate (high enough to be understood)?

- A) strongly agree B) agree C) undecided D) disagree E) strongly disagree

13. You are motivated to take the TOEFL® ITP test in the future (*answer only one option*).

A) If you agree, what could be **a** factor that might motivate you? _____

B) If you do not agree, what could be **a** factor that might demotivate you? _____

C) If undecided, why? _____

14. Have you ever or are you currently taking a preparation course to take the TOEFL® ITP?

- A) Yes B) No

15. Do you agree with the fact that the TOEFL is the only official way to demonstrate your English proficiency in this university?

- A) strongly agree B) agree C) undecided D) disagree E) strongly disagree

16. The score obtained in the TOEFL® ITP test gives you a clear idea of your English proficiency.

- A) strongly agree B) agree C) undecided D) disagree E) strongly disagree

Thank you for completing this questionnaire.

We will treat the information you have provided with the strictest confidence.