



# **BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA**

---

---

**FACULTAD DE LENGUAS**

**“Analyzing test-taking reading strategies in TOEFL  
used by university students”**

A thesis submitted to the School of Languages for the  
Degree of

**Maestría en la Enseñanza del Inglés**

By:

Mario Villa Castelán

**Spring 2020**



**“Analyzing test-taking reading strategies in TOEFL used by university students”**

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

**MARIO VILLA CASTELÁN**

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

**MAESTRÍA EN LA ENSEÑANZA DEL INGLÉS**

**Thesis Director**

---

**Dra. Rebeca Elena Tapia Carlín**

**Committee Member**

**Committee Member**

**Committee Member**

---

**Dr. Eliphelet Rivera  
Cuayahuitl**

---

**Dra. Dora María  
Ocampo Herrera**

---

**Mtro. Leonel Ojeda Ruiz**

**Benemérita Universidad Autónoma de Puebla**

**Puebla**

**January 2020**

## Table of Contents

Dedication.....	vi
Acknowledgements.....	vii
Abstract.....	viii

### CHAPTER I: INTRODUCTION

<b>1.0 Introduction.....</b>	<b>1</b>
<b>1.1 Purpose of the study.....</b>	<b>4</b>
<b>1.2 Study significance.....</b>	<b>5</b>
<b>1.3 Research context.....</b>	<b>6</b>
1.3.1 Participants.....	8
<b>1.4 Research questions.....</b>	<b>9</b>
<b>1.5 Methodology.....</b>	<b>9</b>
<b>1.6 Conclusion.....</b>	<b>10</b>

### CHAPTER II: LITERATURE REVIEW

<b>2.0 Introduction.....</b>	<b>11</b>
<b>2.1 Language learning strategies (LLS).....</b>	<b>12</b>
<b>2.2 Test-taking strategies (TTS) versus testwiseness (TW).....</b>	<b>13</b>
<b>2.3 Test-taking strategies background.....</b>	<b>14</b>
<b>2.4 Test-taking strategies (TTS).....</b>	<b>15</b>
<b>2.5 TW taxonomy.....</b>	<b>17</b>
<b>2.6 Theoretical approaches.....</b>	<b>20</b>
2.6.1 Other factors.....	21
2.6.1.1 TTS as (meta) cognitive strategies.....	22
2.6.1.2 TW and language proficiency.....	23
2.6.1.3 Affective filters and other factors.....	24

2.6.2 Instruments to gather TW data.....	26
<b>2.7 Testing.....</b>	<b>27</b>
2.7.1 Multiple-choice tests.....	27
2.7.2 Testing reading.....	28
2.7.2.1 Reading strategies.....	29
2.7.3 Test validity.....	30
2.7.4 TTS instruction to overcome unfairness.....	31
2.7.5 TOEFL.....	34
2.7.5.1 TOEFL description (reading section).....	35
<b>2.8 Critical thinking.....</b>	<b>35</b>
<b>2.9 TW for teachers.....</b>	<b>36</b>
<b>2.10 Conclusion.....</b>	<b>37</b>

### **CHAPTER III: METHODOLOGY**

<b>3.0 Introduction.....</b>	<b>39</b>
<b>3.1 Context of research.....</b>	<b>39</b>
<b>3.2 Participants.....</b>	<b>40</b>
<b>3.3 Instruments.....</b>	<b>40</b>
<b>3.4 Data collection procedure.....</b>	<b>41</b>
3.4.1 Data analysis.....	42
<b>3.5 Conclusion .....</b>	<b>43</b>

### **CHAPTER IV RESULTS AND DISCUSSION**

<b>4.0 introduction.....</b>	<b>44</b>
<b>4.1 Analysis of the TOEFL results.....</b>	<b>44</b>
<b>4.2 Analysis of question classification into Millman’s taxonomy.....</b>	<b>50</b>
<b>4.3 Analysis of the classification of answers according to the used strategies.....</b>	<b>52</b>

**4.4 Analysis of results.....56**

**4.5 Conclusion.....57**

**CHAPTER V: CONCLUSION**

**5.0 Introduction.....58**

**5.1 Answers to the research questions.....59**

**5.2 Recommendations for Further research.....61**

**REFERENCES.....64**

**Appendixes**

Appendix A. Final questionnaire.....68

## **DEDICATION**

I dedicate this thesis to my family who understood and supported me through all the process of this research. To my professors who dedicated part of their time and knowledge to guide me to get where I am now, and to my students whose contributions helped me to increase my understanding on how learning a language works.

## **ACKNOWLEDGEMENTS**

Foremost, I want to express my sincere gratitude to my advisor Dr. Rebeca Elena Tapia Carlin for her valuable guidance and understanding. You definitely provided me with the elements needed to take the right direction.

Furthermore, I would like to thank Mtro. Leonel Ojeda Ruiz, Dr. Eliphelet Rivera Cuayahuitl, Dr. Dora María Ocampo Herrera and every person who in some way contributed to the completion of this thesis. Those people, who with their valuable feedback, contributed to improve my report in this thesis. To all of you, thank you so much!

## ABSTRACT

It is known that students use different strategies according to the task in turn. Besides, the strategies in used depend on the student's ability to use it, how acquainted he is with it and his previous knowledge. The purpose of this research was to analyze what were the different test taking strategies that different students used to increase their scores in the reading comprehension section from the TOEFL exam. This in turn could help to identify those strategies and spread its use among other students to increase their TOEFL scores. In this study, the instruments in used were a pre and post TOEFL test, a reading test-taking strategy taxonomy (from Millman, Bishop & Ebel's (1965) taxonomy on testwiseness), and a questionnaire (adapted from Cohen & Upton (2006), and Kashkouli & Barati (2012) on testing reading on multiple choice test (TOEFL) and reading tasks (FCE) respectively) recollect the different data to be analyzed (under the taxonomy's criteria mentioned before: time-using, error-avoidance, guessing, deductive reasoning, intent consideration, and cue-using). After applying the pre and post tests, and then the test-taking strategies questionnaire, the research findings showed that students made use of certain type of strategies more often than others to increase significantly their reading TOEFL scores. The findings could contribute to a deeper research related to test taking strategies in reading comprehension tests like the TOEFL which in turn will make evaluation more accurate and fair for students who are test-naïve.



## CHAPTER I: INTRODUCTION

### 1.0 Introduction

Over the years, researchers have found that students learn in different ways due to their individual leaning differences (Lightbown & Spada, 2006); students help themselves to learn and use language through the use of different strategies. Clearly these strategies are different from student to student (Cohen, 1998, cited in Amer, 2007). However, there are a number of those strategies that are common among students. Researchers focus on those strategies that students use the most because they may be helpful to other students. Students who do not use certain strategies may not use them because they have not had the chance to be in a situation to bump into them and see how helpful they could be to their language learning. Another reason why students do not get in contact with learning strategies is when tests have a negative backwash on the way classes are taught, so many teachers concentrate more on teaching content than on teaching students how to learn and develop holistic language abilities, and how to learn to pass a test. Therefore, several teachers do not “waste” time showing strategies to students. This way, it is the students’ endeavor to find out what things do help them to learn and pass *the test*. They do not know if certain strategies may be helpful or not, or that there are strategies which are suitable to their learning style.

On the one hand, there are students who, in one way or another, already found strategies suitable to their learning style and they have great academic success in the classroom but somehow they cannot achieve a good score on their language tests or their scores are under the expected average (Ghafournia, 2012).

On the other hand, there are students who are not very noticeable in the classroom but yet they get good scores compared to their language proficiency and sometimes better than those students who teachers consider as having a high language level (Ghafournia & Afghari, 2013). Somehow, either by their own knowledge or other means, they get the right answers. The problem here is if they may get the “right answer by the wrong means” (Cohen & Upton, 2006). That means they find the right answer but not based on their knowledge on the language, but based on some test’s bias. It seems that those students already found and developed strategies that may not help them with their academic linguistic learning but to pass their language tests. This does not mean that some students have found out what is expected from them in a test. As Ghafournia (2012) states “...if test takers cannot identify what is expected from them, they cannot reflect their actual linguistic ability” (p. 2). They may have the linguistic knowledge but may not be familiar with a test format. In other words, it means that some students who get high scores somehow are able to understand what the possible right answers in multiple choice tests are, and they are able to find them under certain amount of time. Some teachers may wonder how these students got a higher score if they have an average linguistic proficiency. Researchers have known some time ago that multiple choice tests have certain faulty items that some students have, consciously or not, already identified and use to pass their language tests (Sarnacki, 1979). One example is that some students have found out that most of the right answers are in the middle of the options or that the longer answer is usually the right answer. So, through experience and observation some students have been able to use the tests structure itself to obtain extra points.

However, researchers affirm that it is not possible to pass a test, even with the right strategies, if the subject does not have at least certain amount of knowledge on the subject. As Cohen (1998) states, success in tests depends on having the necessary linguistic and strategic foundation (cited in Ghafournia, 2012). But, on the other hand, sometimes students with enough second language knowledge are unable to perform well on tests. So again, it may not only depend on enough linguistic foundation, but also on using the correct tools (test-taking strategies). For example, usually students from marginal or far places from the urban areas, where education may not be of good quality (as well as nurture), tend to get the lowest grades, and that is in part of being test naïve. That is, not having enough testwiseness, not only due to lack of linguistic knowledge, but also lack of experience with certain kind of tests. Technically everybody has been a test-taker naïve in a moment of our schooling life. Therefore, to make the tests *fair*, it would be necessary that all students have the same test-taking strategies. Therefore, test-taking instruction is necessary especially to those test-taking naïve students (Ghafournia & Afghari, 2013). However, if a student has the right strategies to pass a test, these strategies will not be useful if s/he does not have certain amount of knowledge on the subject (Rogers & Yang, 1996). Therefore, it is not wild guesses that can help a student to pass a test but academic guesses (Woosley, 1973, cited in Sarnacki, 1979). That means using the test structure, clues and cues, and their knowledge is how students could improve their scores in multiple choice tests. This is what researchers called test-taking strategies. Other researchers have also called them test-wiseness (Cohen, 2007, cited in Kashkouli & Barati, 2012). In this text, this research focuses on test-taking strategies used in the TOEFL reading

section by university students who increased their score by 50 points after an 80-hour course period. As it is explained in the next paragraph, it is believed that by identifying which strategies these students used, they can be instructed to other students so they notice, learn, and use these strategies together with their academic knowledge to improve their TOEFL scores.

### **1.1 Purpose of the study**

Some students have had the experience of working hard to get a good grade, but when the test results are in their hands, they feel disappointed when they notice that the results are not as promising as they expected. Even teachers sometimes are surprised when they review test scores and observe that students who they expected will have a good grade, did not, or that students with the same proficiency level get noticeably different scores. Students that work hard, as well as we as teachers, wonder how that can be possible. Usually we tend to think that our judgment about certain students was wrong. That the low level students are not that low and that our best students are not that good. We may even think that the test validity is in question. That the test did not evaluate what it was supposed to and generates inaccurate scores. However, researchers (Ghafournia & Afghari, 2013) have found that there are strategies that students use in the classroom (and under certain academic tasks at home), and strategies that student use during tests. It is this later kind of strategies that some students take advantage of to get good scores on language tests. This kind of strategies is known as test-taking strategies.

The purpose of this study is to find out what kind of test-taking strategies university students use frequently during a multiple choice test and which strategy is the most frequent. Especially those students who got 50 points or more on a TOEFL post-test after a language course. The study is going to focus on the reading test-taking strategies because one of the materials that the students use during class is a reading comprehension textbook, and one of the university's policies is that students can be promoted to a higher level (or study another language) or receive their major diploma if they pass a certain kind of test (TOEFL). Also, this paper will focus on test-taking strategies because after a pre and post-test students who increased their score 50 points got more of the increase in the reading comprehension section of the TOEFL than on the other two sections. This may be expected because these students took a course where a major emphasis was on reading comprehension; however, how can it be explained that only 4 students, from 4 groups out of 15 to 20 students each, improved their TOEFL scores in 50 points being the reading section the one with the most increase, and the rest of the students did not?

In this study, it is intended to find out what reading test-taking strategies these 4 students use and which one is used most of the time so they could be shared with teachers and taught or passed on to other students. It is expected that this will beneficiate not just other students but also the language department when the percentage of their students who fail to get the required score decreases.

## **1.2 Study significance**

Lately, different universities are requiring students to have certain English language level before graduating. These policies ask students to be able to demonstrate that they have achieved a recommended language level according to the academic level. The university where this study is carried out asks student to get a TOEFL score of 450-550 points (according to the student's major) as a requirement to get their university degree diploma. To these days the number of students who do not obtain the required scored is increasing as reported by the coordinator. The language teachers' academy does its best to make students increase their language proficiency. Some students are taking language tutorials and others are taking English language courses in other institutions to assure their score (based on anecdotal evidence). However, maybe not many teachers involved in this endeavor are observing those students who pass their TOEFL and are asking them how they did it, what strategies they used, or where and which courses they took to succeed. Usually they are seen as one less student to worry about, and that is why nobody inquires more on what these students did to succeed. Sometimes their teachers and classmates assume that they simply study harder. In a closer look at their scores, it was observed that these students got most of their points on the reading section. Even though all students take the same course, which is based on developing reading comprehension and writing, these students are the only ones from their groups that got 50 points more in their TOEFL scores after a language course, resulting the reading section to be the one where they got most of their score increase. This is why the premise that these

students used a different set of strategies is taken in order to carry out this study. Under that premise this study intends to pay close attention to these students and their results on the TOEFL, and find out which test-taking reading strategies these students applied that helped them to increase their scores in 50 points after a course of 80 hrs. Certainly, the results are of significance because the TOEFL test is extensively used as a requirement for students to graduate or not from their majors in the whole university. Also, the results could help the academy teachers to implement these strategies in the already established language programs and extend the use of certain strategies to more students. However, it is also possible that these students increased their scores because they did something outside the classroom to increase their level (which is one of this study's limitations).

### **1.3 Research context**

The Universidad Popular Autónoma del Estado de Puebla (UPAEP) is a private university which has its roots in Puebla City. It was founded in 1973 by teachers and students from architecture and Business administration from UAP. Even though it has a religious background, these days, students do not have to belong to a religious group to enter. This university offers different majors, Master programs and PhD degrees in different areas. As many universities in Mexico, UPAEP includes a foreign language subject in all their majors, but it has a policy for each major that has English as a foreign language subject which goes from 450-550 points in the TOEFL test to graduate. This way, the TOEFL test works as a gate-keeper to determine who graduates and who has to keep on working (Fahim et al., 2010). Due to these policies, students, teachers and the language

teacher faculty are working on methods to make students increase their English proficiency and their scores. On the one hand, the teacher faculty adopted different textbooks based on reading and writing (with more emphasis on reading comprehension). Besides this, they have weekly reunions with the director and coordinator to develop different methods and strategies based on the American Council on the Teaching of Foreign Languages (ACTFL) to tackle this problem. However, despite these efforts, the students' TOEFL scores are decreasing as it was mentioned by the (Departamento de Lengua y Cultura) DELC's two representatives. This situation of looking for a manner to increase the students' TOEFL score is the one that motivated this research. This study started by applying a practice test to four groups in the Umbral level (the fifth language level course). After applying the practice TOEFL test to obtain the initial students' scores at the beginning of an 80-session course, their scores were collected to be compared to a second different practice TOEFL test at the end of the course. As expected, it was observed that in each group most of the score increment was on the reading section of the test. However, there were students in each group who increase their scores up to 50 points or more, which is very high for an 80-hour course. This was another motive to this paper. It was assumed that these students did something that helped them to increase their scores that high. Because they were taking the same classes, in different groups with the same teacher and the same content, it was assumed that they may have used some strategies during the test or that they did something after classes that make them increase their TOEFL scores. It was decided to do research on the reading test-taking strategies (testwiseness) that these students may have used during the test, and to identify



which one they used the most. In this endeavor, this paper is trying to look for useful strategies that students may have used during their tests. By finding them, these strategies could later be shared with teachers who could later teach them to their students so that they are able to increase their test scores.

### 1.3.1 Participants

The participants for this study came from different majors and from different years of university that range from sophomore to junior. When they enrolled in the university, they presented an English exam that places all students in different levels. These levels go from Acceso 1 and 2, Plataforma 1 and 2, Umbral, Avanzado, and Conversación according to the DELC department's classification. At this moment all of them just finished the Umbral level. Three boys and a girl from different majors were selected. The female subject is studying Medicine, and the 3 male subjects are taking majors like architecture, marketing, and biotechnology. They were selected, not because they had advanced English language proficiency or because they got high scores in the TOEFL test, but because they were the only ones who increased 50 points or more above their initial scores at the end of the course. Most of the students had an increase and a few had a decrease on their scores on the second TOEFL test, but the increase of 50 points from these 4 students attracted our attention wondering what they did and if that can be transferable or teachable to other students. Most of the 4 participants were average students, except for one of them who had a low proficiency level but was the one who got more than 50 points (80 to be exact).

#### **1.4 Research questions**

1. What test-taking strategies do students use when answering the TOEFL reading section?
2. Which test-taking strategy do students apply more frequently when answering the TOEFL reading section?

#### **1.5 Methodology**

To carry out this study, a TOEFL practice test was applied to 4 groups of about 20 students each one. Then, they received an 80-session English language course which included two textbooks: one for reading comprehension development and the other for writing development. Most of the emphasis was on the reading textbook. At the end of the course, students received a second TOEFL practice test. Then, their scores were compared and from each group there were students who in average increased their scores from 5 to 25 points in this second test; however, there were 4 students from the different groups who increased their score from more than 50 points. This attracted our attention to do a study and find out what strategies these students had used to obtain such an increment. It was hypothesized that these students had used some strategies especially in the reading section because that section was the one that increased significantly. The other areas had increased too, but the reading section was the one that increased the most. Then, it was decided to apply a test-taking strategy questionnaire, adapted from Cohen and Upton (2006), and Kashkouli and Barati (2012) on testing reading on multiple choice test (TOEFL) and reading tasks (FCE) respectively to find out which strategies these students used the most. After getting the results

from the questionnaire, they were classified according to the strategies the students used (time-using, error-avoidance, guessing, deductive reasoning, intent consideration, and cue-using). These strategies were taken from Millman, Bishop and Ebel's (1965) taxonomy on testwiseness (cited in McPhail, 1981). Finally, the results are presented in the conclusion chapter of this paper. It is important to emphasize that even though this taxonomy classifies test-taking strategies on multiple choice test in general, most of the strategies presented are applicable to the reading TOEFL section.

## **1.6 Conclusion**

This research intends to find out what test-taking strategies students used to get a significant increase in their TOEFL score in the reading section and which strategy they used the most. Knowing which strategies are more helpful to students will give teachers an insight on which strategies to teach them and give them the tools to get the best out of their knowledge and maximize their TOEFL score, which is what they need to be able to graduate from the university.

The following chapters will discuss the literature review, the methodology, the results, and the conclusion of this research. In chapter two the literature review will be presented. Chapter three is related to the methodology, the participants, and the instruments used in this research. Chapter four has to do with the results, the analysis and interpretation of those results. Finally, the conclusion of this research will be presented in chapter five.

## CHAPTER II: LITERATURE REVIEW

### 2.0 Introduction

Chapter 2 will cover different aspect related to tasting strategies. These topics range from a background on strategies to test taking strategies. In between, topics like testwiseness taxonomy, theoretical approaches and unfairness testing will be discussed to offer an ample view on the subject matter.

These days it is in vogue to teach our students as many strategies as possible to learn a second language. Those little tricks, that as language students, we once learned and used. Usually, we as teachers try to transmit our students our own experiences with certain strategies hoping they may find out how useful those strategies were for ourselves. We even teach them strategies according to the skill we are teaching. In our repertoire we already classified them: strategies for successful speaking, strategies for vocabulary learning, and so on. That repertoire even includes strategies for taking certain kinds of tests. However, we do not teach our students strategies to present language exams. Some of us, as teachers, may think that doing that is similar to teach our students to cheat on an exam. And cheating is definitely something we do not want our students to learn to do, especially not from us. Sometimes, if we used some strategies to get a good score as students, we do not openly talk about that because we are afraid others may think we got a high score not because of our own knowledge, but because of certain strategies we employed. In this section we are going to talk about those test-taking strategies also known as testwiseness or test management (Cohen,

2007, cited in Kashkouli & Barati, 2012), their taxonomy, components, types of exams they are more effective on, and even why it is necessary (important) to teach them to our students.

## **2.1 Language learning strategies (LLS)**

We cannot talk about test-taking strategies without talking about what a strategy is. Strategy is a word that comes from the ancient world when strategy was similar to tactic in war (Alavi & Bordbar, 2012a). These days, scholars use the term learning strategies to differentiate them from all the different strategies that exist in different fields (Oxford, 1990). Test-taking strategies are seen from the perspective of language learning strategies (LLS). They fall into the category of compensatory strategies or strategic competence (to see them from a communicative view) due to their interrelated nature (Ghafournia, 2012). There are different language learning strategies and everybody uses them according to their convenience and kind of task in hand (in the four language skills). Therefore, students will be able to manage learning related problems when they identify strategies that are suitable to facilitate their learning (Gharbavi & Mousavi, 2012). The first research to talk about LLS academically was Rubin in 1975 in the article named "What The Good Language Learner Can Teach Us" (Cohen & Macaro, 2007). Since then, many researchers have been studying LLS, but there are still some differences about its definition. For example, O'Malley and Chamot (1990) define them as "special ways of processing information that enhance comprehension, learning, or retention of the information" (p. 1) while according to Oxford (1990) strategies are "specific actions taken by the learner to make learning

easier, faster, more enjoyable, more self-directed, more effective, and more transferable to new situations” (p. 8). Besides, Ghafournia (2012) states that “..., each researcher seems to have his/her own definition and classification of language learning strategies” (p. 140). However, many researchers agree that LLS have to have a metacognitive component which implies that a LLS involves a conscious and intentional process (Cohen & Macaro, 2007). Therefore, LLS implied a conscious decision about what works better for our learning or the completion of a task. Even some of the characteristics that Oxford (1990) mentions about strategies are that they can be conscious and teachable. About language learning strategies, different books have been written related to how to teach LLS to our students; however, little has been written about test taking strategies in language testing (Flippo & Borthwick, 1981). As Allan (1992) stated “Although these strategies are extensively studied in L1 literature, they are noticeably neglected in EFL/ESL testing literature (cited in Amer, 2007). Therefore, not many teachers know about them and how beneficial they may be for our students once they have to face the final challenge at the end of the language learning road that is the “final test” (known in many places as TOEFL) which will determine if they are promoted and able to finish their studies or they still have to take remedial courses (Ghafournia & Afghari, 2013). Let us analyze test-taking strategies more closely.

## **2.2 Test-taking strategies (TTS) versus testwiseness (TW)**

First, we have to mention that although test-taking strategies is not a new term, it is still very useful, and it comes from previous terms like testwiseness or test management (Benson, 1988, cited in Rogers & Bateson, 1991). Although

Cohen (1998a, cited in Kashkouli & Barati, 2012) made a differentiation among test-taking strategies and testwiseness saying that TTS include language use strategies and TW includes knowledge on how to take a test. The true is that a student may not be able to answer a test if he does not have partial knowledge of the subject matter and how to answer a test (Amer, 2007). As Rogers and Bateson (199) stated "...effective application of TW strategies is dependent on some partial knowledge of content" (p. 348). Even Cohen himself stated later that test management strategies and TW were known as test taking strategies by previous literature (2007, cited in Kashkouli & Barati, 2012). Also, there are other authors who use the two terms interchangeably (Maspons & Llabre, 1983; McPhail, 1981; Rogers & Bateson, 1991; Ghafournia, 2012). That is why in this paper test-taking strategies, testwiseness and test management are going to be used to connote the same term.

### **2.3 Test-taking Strategies background**

Although language learning strategies have been studied a great deal, we cannot say the same from test-taking strategies (Ghafournia & Afghari, 2013; Cohen, 2011). Sarnacki (1979) explains that the concept of testwiseness was first introduced by Thorndike (1951) as an element affecting reliability and as a general trait of the test taker, but there were problems of interpretation of its specific components because empirical research had not been performed by then. Later, he explains that Ebel and Damrin (1960) mentioned test-wiseness as a component of response variance in objective tests, but they did not carry out any experimentation on the topic. Then, Gibb (1964) proposed an operational definition

of TW and designed an instrument to measure it (Sarnacki, *ibid*). Finally, in 1965, Millman, Bishop and Ebel presented a TW taxonomy which became, in words of Sarnacki, “the classical theoretical work in the area” (1979, p. 252). This taxonomy became the framework for future empirical study. Since then, TW had been mentioned in professional literature. As Cohen and Upton (2006) mentioned that since the late 1970s, researchers started to approach second language testing from the point of view of the strategies that respondents use in the process of performing a language test. These days, some researchers address the topic from the language learning strategies from Oxford (1990) as compensation strategy (Amer, 2007). Others, according to Cohen and Upton (2006), refer to TTS from the point of view of strategic competence from Canal and Swain (1980) and others from the refine strategic competence concept from Bachman and Palmer (1996), which is an ampler framework that includes test-taking strategies. A description of test-taking strategy is presented in the next sections.

#### **2.4 Test-taking strategies (TTS)**

Test makers have known about test-wiseness from long before, especially in multiple choice tests (Bar-Hillel, Budescu, & Attali, 2005). However, they were unable to define it. One of the first definitions among many (Thorndike, 1951; Ebel & Damrin, 1960; Vernon, 1962; Gibb, 1964; English & English, 1970; Stanley, 1971; Diamond & Evans, 1972; Oakland, 1972; Nilsson & Wedman, 1974, cited in Sarnacki, 1979, and Rogers & Bateson, 1991), and that had been generally adopted, was the one from Millman, Bishop and Ebel (1965) who described TW as “a subjects’ capacity to utilize characteristics and formats of the test and/or test



taking situation to receive a high score" (p. 707, cited in Rogers & Bateson, 1991). This definition was similar to other authors. For example, Nitko (1983, p. 326) defines test-taking strategy as "a student's ability to use the characteristics of both the test and the test situation to attain a higher score" (cited in Alavi & Bordbar, 2012a). Besides, Bar-Hillel et al. (2005) also explained that TW is the ability that the testee uses to take advantage of some characteristics of multiple choice tests to increase the probabilities of obtaining a successful score. However, those definitions have been complemented because they imply that TW is independent from the test taker's knowledge about the subject matter of the test (Flippo & Borthwick, 1981). For example, Rogers and Bateson (1991) asserted that "effective application of TW strategies is dependent on some partial knowledge of content" (p. 348). If the opposite were true, it would imply that the examinee's TW is just about making wild guesses and taking risky decisions. Nevertheless, although guessing and risk taking are often confused with TW, they are not sufficient enough to describe TW. Guessing and risk taking are components of TW but cannot describe TW completely. TW includes different components in which guessing and risk taking are just a small part (Woosley, 1973, cited in Sarnacki, 1979). Thus, test-wiseness includes deductive reasoning which, in order to be successful, needs some partial knowledge (Sarnacki, 1979). Furthermore, Rogers and Bateson (1991 cited in Amer, 2007) states that "TW is a cognitive ability or a set of test-taking strategies a test taker can use to improve a test score" (p. 4). Later, they (ibid, p. 333) stated that the cognition of the skilled test taker is composed of: a- a cognitive monitor that controls which abilities and skills are going to be engaged to answer the item under consideration; b- knowledge,

abilities, and skills relevant to the content or trait being measured; c- knowledge of TW principles; and d- the response (selection and record of choice). Then, Cohen (2007) added the conscious term by suggesting that “test-taking strategies are consciously selected processes the test wise student employs to face language issues and answers in the test-taking activity” (p. 308, cited in Kashkouli & Barati, 2012). Here, Cohen stated that the mixture of test management and test-wiseness is what previous literature named test-taking strategies. Finally, Shaw and Weir (2007, cited in Ghafournia & Afghari, 2013, p. 141) put these entire concepts together in a complete definition of test-taking strategies:

Test-taking strategies are the strategies respondents often select consciously when taking a test. Similar to learning strategies, the element of conscious selection is of essential importance in implementing test-taking strategies. These strategies are diverse by nature. Effective application of test-taking strategies enables less proficient language learners to opt out on language tasks and constitute short cuts to elicit correct responses. In other words, test takers may use test-wiseness to circumvent the use of their actual language knowledge to answer questions. Therefore, it is crucial for test constructors to find out what their tests actually measure during the pilot phase of test development.

On the other hand, Kouzekanani, Llabre and Baldwin (1989) reported that TW is a multidimensional construct with different components making it able to be measured, and taught, and it also increases test performance. These different concepts related to test-wiseness will be addressed in later paragraphs.

To conclude, it is necessary to mention that a test-taking strategy cannot be efficient or inefficient, it depends on the test task at hand. For example, some

learners may use a small number of strategies but efficiently, while some others may use many test-taking strategies but not very efficiently. Another example is that not because a subject applies a TTS frequently, it means that this TTS is going to be successful in the next test (Cohen 1998, cited in Amer, 2007). Another aspect that we have to take into consideration is that there are other variables that may influence the effective use of TTS. Amer (2007) remarked some of those variables like the students' cognitive style, their linguistic proficiency, test-taking style, their repertoire of TTS, their test anxiety, type of test, and type of test-task. He also describes style as "the habitual use of a strategy" (p. 12). These variables will be covered in the paragraphs below.

## **2.5 TW taxonomy**

As it was mentioned before, one of the main problems in the understanding of TW is the unfamiliarity with its components (Sarnacki, 1979). Amer (2007) commented that there have been different TW taxonomies proposals. He mentioned that one of them was Nitko's (2001) who classified TTS into three divisions: time-using, error avoidance, and guessing. Another classification was done by Sarnacki (1979) which is very similar to that proposed by Millman et al. (1965, cited in Sarnacki 1979). They made a division of TTS into six categories: test-using, error avoidance, guessing, deductive reasoning, and intent consideration and cue-using. Then, Amer (2007) mentioned that a taxonomy that is used extensively in general education classified TW into three categories according to Watter and Siebert (1990), and Wenden (1991): strategies used before, during and after answering a test. According to Amer (ibid) there are other

two taxonomies closely related to EFL/ESL test-taking strategies that are proposed by Cohen (1998), and Bachman and Palmer (1996). Cohen explained TTS from the language learner strategies dividing them into L2 learning strategies and L2 use strategies. According to Cohen the latter category also constitutes TTS when their application is into tasks in language tests. The other taxonomy proposed by Bachman and Palmer is based on strategic competence. They classify these Metacognitive language test-taking strategies in three categories: goal setting, assessment, and planning. Nonetheless, Amer stated that these two taxonomies are theoretical conceptualizations which lack experimental validity. For that reason, he mentioned that researchers often refer to Oxford's (1990) taxonomy. She organizes strategies into: cognitive, Metacognitive, affective and social. In a language test, cognitive strategies refer to current mental activities to use the language and world knowledge to carry out a test task. Metacognitive strategies are the examinee's mental operations to guide and manage cognitive strategies to answer a test successfully. Affective strategies are related to emotions such as confidence, stress, and so on, and social strategies are related to increase interaction with the L2.

As we can observe, there are different taxonomies that have been proposed. However, according to Flippo and Borthwick (1981), the one that is generally cited and referred to is the one designed by Millman, Bishop and Ebel (1965). Besides, this taxonomy had been used as a general framework for studies on TW (Rogers & Bateson, 1991). The taxonomy is divided into two categories: strategies independent / dependent of the test constructor or test purpose. In the

strategies first division (called general test-taking strategies) are strategies to use time adequately, rereading answers to avoid mistakes, knowing how and when to make educated guesses, and employing deductive reasoning to obtain the right answer. In the second category we have two subdivisions: strategies to take into account the test maker intention and strategies to use cues in the test (Flippo & Borthwick, 1981). According to Rogers and Bateson (1991), the first group includes elements that are applicable to most of the testing situations and will help the testee to avoid losing points for reasons other than lack of knowledge. The second group of strategies will help the testee when he has knowledge of the purpose of the test or experience from previous tests similar in format and objective. Furthermore, Rogers and Yang (1996) give a similar explanation:

“The first two subdivisions contain elements or strategies which, if employed, will help examinees avoid losing points for reasons other than lack of knowledge of the content tested. The last two subdivisions in Part I contain elements which allow examinees to gain points beyond what they would have received on the basis of sure and full knowledge of what is being tested. The principles listed in Part II of the taxonomy may prove beneficial when the test taker has knowledge of particular test making behaviors or knowledge of particular testing practices gained from past experiences with tests similar in purpose and format. As in Part 1, the elements in the first subdivision will help examinees avoid losing points, while the elements in the second subdivision will help examinees gain points” (p. 249).

Moreover, McPhail (1981) remarked that this taxonomy was integrated from the test construction principles and the respondent's problem solving styles. He also reported that this taxonomy was designed as a framework for empirical research. Finally, according to Sarnacki (1979), this taxonomy is a complete work

on TW; however, it is necessary to take into account the other taxonomies to have a better understanding on the subject.

## **2.6 Theoretical approaches**

According to Ghafournia and Afghari (2013), most of the studies on strategic aspects on TTS do not have a base theory of cognition. They explained that it may be due to the complex interrelation between test-taking strategies (TTS) and language learning strategies (LLS). To separate TTS from LLS is a highly complex procedure because it is difficult for the researcher to isolate strategies used only in language learning from strategies used only in test taking. Furthermore, Ghafournia (2012) as well as Fahim, Bagherkazemi and Alemi (2010) pointed out that because of the interrelated nature of learning strategies and test taking strategies, both concepts should be studied collectively. On the other hand, Sarnacki (1979) claimed that TW should be evaluated from two theoretical approaches. The first one asserts that TW exists due to faulty test-item construction. Here, TW is not seen as the examinee's traits, but as a consequence of poorly test-item development; therefore variability in test performance is due to this characteristic on tests. The other theory stated that TW is a constant quality of the testee. This is not related to psychometric aspects of the test, but to the subject's ability to employ TW; therefore, the variability in test performance is due to the amount of TW each test-taker has. Because of these different theories, a more holistic theory about TTS is necessary to understand this concept. A theory that probably should include test characteristics, and individual's abilities, holding

testing and language learning strategies as their bases to describe test-taking strategies is necessary.

### 2.6.1 Other factors

There are other interesting aspects to take into account to understand TTS. One of them is that even though TTS is usually linked to LLS as a strategic process, it has not been studied extensively (Ghafournia & Afghari, 2013). Another aspect that Rogers and Yang (1996) mentioned is that the conditions that generate a higher score on tests are if the test-taker possesses TW and relevant knowledge and if the test contains flawed items. This should be taken into consideration by testers to obtain fair and accurate scores. Therefore, some may think that test-wiseness is similar to guessing. As it was explained before, guessing and risk-taking are elements of TW but are not appropriate labels of TW. In other words, it will not be wise to make a wild guess or take an eeny, meeny, miny, moe risk to choose and answer. TW includes guessing and risk-taking using partial knowledge and deductive reasoning. As Sarnacki (1973) stated [...although the test-wise individual may be a risk taker, it does not conversely follow that the risk taker is necessarily test-wise. This is demonstrated in the earlier contention that risk taking is a component of TW, not a synonymous or parallel construct] (p. 257). TW is neither answering in certain style or bias like choosing the answers that are in the middle. Actually, doing so may lower one's score. To end up, Kashkouli and Barati (2012) motivated by Phakiti (2003) and Cohen (1998b) explained that TTS are conscious and selective actions which can be affected by the kind of test taker, his

proficiency level, the settings, and the nature of the task. Following, some other factors related to TTS are going to be analyzed.

#### 2.6.1.1 TTS as (meta) cognitive strategies

The concept of cognitive theory was coined in the 1970s and since then the learner was perceived as an active constructor of his own knowledge using metacognitive skills. This view of the language learner, as an active constructor, motivated some scholars to work on the strategic process of learning and other variables influencing learning such as language proficiency, learning styles, gender, motivation, anxiety, beliefs, learning tasks, and cultural background (Ghafournia & Afghari, 2013). As mentioned before, TTS is seen from the strategic competence perspective. Nonetheless, studies on this field are relatively new. It is believed that this situation is due to difficulty in separating LLS from TTS. For example, Bachman, Cushing and Purpura (1993, cited in Ghafournia & Afghari, 2013) stated that the effect of TTS on test performance is direct in some moments and indirect in others. They continue observing that this variation in the use of TTS is caused by nonlinguistic variables more often than linguistic variables. Therefore, the interpretation of test results is a difficult endeavor. Bachman (1990, cited in Ghafournia & Afghari, 2013) stated that this difference is due to true score (language proficiency) and error score (nonlinguistic factors). The latter are divided into systematic and unsystematic. Systematic factors include personal aspects which include strategic competence. Strategic competence involves metacognitive strategies which assist the learner to process language in different situations of language use including test-taking settings. Before this explanation, other authors



were already linking TTS with cognitive strategies instead of metacognitive strategies; for example, Sarnacki (1979) pointed out that TW suggested cognitive abilities that the test-taker can use in a variety of tests, independent of the subject content. This is complemented by Phakiti (2003) who stated that most of the cognitive strategies come about together with metacognitive strategies, and that a test-taker ought to be metacognitive to employ cognitive strategies (cited in Amer, 2007). Amer (2007) later remarked that the use of cognitive and metacognitive strategies may be influenced by the task and the testee's characteristic. For example, he commented that unsuccessful examinees use more metacognitive strategies to retrieve information, while successful examinees employ metacognitive strategies to understand and remember. Finally, Amer (2007) influenced by Purpura (1997, p. 311) added that "...we can no longer operationalize Bachman and Palmer's (1996) notion of "strategic competence" solely in terms of a metacognitive component; rather, "strategic competence" should minimally include cognitive and metacognitive processes". Finally, BIÇAK (2013) added that the student's affective factors are components of the student's metacognitive attributes which assist students to be confident on testing situations. Consequently, we can observe that TTS are related to cognitive and metacognitive strategies because linguistic and nonlinguistic factors influence the test results making it difficult to determine what a test score really represents.

#### 2.6.1.2 TW and language proficiency

When the TW concept first appeared, scholars thought that TW was independent of the examinee's knowledge about the subject matter (Millman et al.,

1965 cited in Sarnacki, 1979). Thus, they thought that multiple choice tests were vulnerable to wild guessing (Sarnacki, *ibid*). Later, Rogers and Bateson (1991a cited in Rogers & Yang, 1996) questioned that TW and content knowledge were independent. In a study developed by them, where strategies used by high and low academic students were examined, they found out that students who were considered academically talented were classified as test-wise, and students with low academic level were more test-naïve. In other words, students with partial knowledge on the topic to be tested and with knowledge about TW scored better than students with low academic knowledge and low knowledge about TW. This study implies that partial knowledge about the topic been tested and TW are interrelated to influence positively in a final score. In their study they cited Garner (1990, p. 520) who said that “Students with low content knowledge but test-wise knowledge and students with partial knowledge but low test-wise knowledge will perform less well than students who possess both on such items”. Then in 1996, Rogers and Yang emphasized that the correct use of TW reasoning strategies are dependent on some partial knowledge, which alone may not be enough to respond successfully either. Other authors like Cohen and Upton (2006) confirmed the same reasoning, reporting that proficiency has an effect on the use of strategies. Ghafournia (2012), on the other hand, noted that the level of language proficiency in students determine the frequency and the kind of TTS used. Besides, more proficiency students use TTS more often which differ from LLS which have showed different results (Park, 2010). Finally, Cohen (1998) asserted that for students to be successful in a language test, they not only need enough linguistic foundation but also an appropriate use of TTS (cited in Ghafournia & Afghari, 2013). As a

conclusion, we can determine that TTS to be successful need partial knowledge on the subject and vice versa, partial knowledge is not very useful if the test-taker possesses no TTS.

### 2.6.1.3 Affective filters and other factors

There are other factors affecting test scores and test taking strategies (Pour-Mohammadi & Abidin, 2011). In the last years there have been studies to try to identify what causes variation in test performance and language test (Cohen & Upton, 2006). Among these factors are bias in response, test-taking experience, intelligence, time (related to memory), test format and method, age, text anxiety, examinee characteristics, and linguistic competence. Most of them are nonlinguistic. The first factor is characterized by testees opting to choose the third option or by certain style in responding. According to Sarnacki (1979) bias in response selection does not reflect TW. He even remarks that such bias may lower the examinee score. Another factor is test-taking experience which may be related to the testee's school grade. It was found that the more advanced in grade the students are, the more TW they showed (Slakter et al., 1970a; Crehan et al., 1974, cited in Sarnacki, 1979). However, test experience by itself does not imply that a student is a skilled test-taker, nor does it guarantee success in test-taking (Sarnacki, *ibid*). On the other hand, different authors found that TW is a characteristic presented in all age groups from preschool to college students and adults (Gaines & Jongsma, 1974; Nilsson, 1975; Gross, 1976; Pryczak, 1973; Bajtelsmit, 1975ab, all cited in Sarnacki, 1979). However, other authors found out that after some time subjects tend to forget those TTS that they manage during

school years due to lack of exposure (Woodley, 1973; and Bajtelsmit, 1975a, cited in Sarnacki, 1979). Furthermore, even though researchers found that intelligence has no correlation to TW (Flippo & Borthwick, 1981; Rogers & Yang, 1996), the test format does. It was found that usually teacher-made tests present flaw items because of unfamiliarity with TW; nonetheless, even standardized tests developed by professional test makers sometimes present flaw items, as was shown by Metfessel and Sax (1958, cited in Sarnacki, *ibid.*). Two more factors related to TW and test performance are verbal achievement and test anxiety. Verbal achievement implies that the testee will have some advantages by being able to identify grammar, vocabulary and sentences structure. Conversely, test anxiety may influence both TW use and test performance due to changes in motivation, self-confidence (BIÇAK, 2013) and mental attitude (Rogers & Yang, 1996). Finally, there is another variable that involves a bigger group of aspects. Here we refer to the testee's characteristics. Phakiti (2003) reported that test-taker's characteristics include: age, native language, culture, gender, background knowledge and cognitive, psychological and social characteristics, motivation, attitude, intelligence, anxiety, and socio-economic status (cited in Cohen & Upton, 2006). As it can be observed, some of these factors have already been analyzed above. However, culture and socio-economic factors will be analyzed later in the further paragraphs related to minority groups.

#### 2.6.2 Instruments to gather TW data

Even though it may seem amazing that there are tests that measure TW, there are and they showed accurate results. One of the most common methods to

gather data even though it is not a test, but it is worth to mention, is thinking aloud introspectively and retrospectively. This method is also known as verbalization. The disadvantage with retrospection is that the subjects have to retrieve information from the long term memory into the short term memory and this process may risk accuracy (Kormos, 1998, cited in Salehi, 2011). Cohen (1994) also mentions a list of limitations using this method and he was the first author to mention the practicability of gathering information through this method (Cohen, 1984, cited in Salehi, 2011). When this method is used, usually test wise students need less time than test naïve students (Rogers & Bateson 1991). Besides verbalizations, there is a list where Rogers and Yang (1996) mention different scholars that have developed diverse TW tests to measure different elements of test-taking strategies in students at different ages (Gibb, 1964; Millman, 1966; Rogers & Wilson, 1993; Woodley, 1973; Bajtelsmit, 1975; Slakter et al., 1970; Diamond & Evans, 1972; Allan, 1992). Besides, Amer (1993) cited Allan (1992) who developed a multiple-choice test of TW exclusively for ESL students. Allan also maintained that a valid measure of TW in ESL testees could identify students with this disadvantage in a psychometric test. This topic about fairness in objective tests will be addressed later when we talk about minorities.

## **2.7 Testing**

Testing is a useful and necessary tool. It is not only used to evaluate the students' knowledge level, but also to examine the quality of the teaching, schools, and teachers; and believe it or not all of them try to make use of test-wiseness (Rogers & Yang, 1996). As Cohen (1998) stated, success in language tests

depends on sufficient linguistic and strategic knowledge (cited in Ghafournia, 2012). Because of the use of tests like TOEFL at university level it is necessary to train students on TTS use. It is possible that students who fail may not only be due to lack of linguistic competence, but also lack of TTS competence. One may think that teacher-made exams are more susceptible to present flaw items; however, there are studies showing that high-stakes standardized exams also present items susceptible to TW (Bar-Hillel et al., 2005). This is due to the increasing use of psychometric/multiple choice test. The use of these tests is in vogue these days because they are easy to grade and practical to be applied to large groups of students. Therefore, because this kind of tests may have TW components, it is a disadvantage to be TW naïve, or as Sarnacki put it, “a test-wise handicap” (1979, cited in Rogers & Bateson, 1991). This is one of the reasons why teaching TTS is vital nowadays.

### 2.7.1 Multiple-choice tests

As it was stated before, multiple-choice tests are very popular in ESL/EFL, particularly when it is necessary to evaluate large number of students, because they are efficient, economic, reliable, and easy to grade (Kesselman-Turkel & Peterson, 1981). However, this kind of tests is more susceptible to flaw items (Geiger, 1997; Katalin, 2000, cited in Ghafournia, 2012). Even though it was stated that wild guessing is not part of the TW student, and that it may lower one’s score, the true is that some students in fact make wild guesses. Therefore, these objective tests cannot show which answers were lucky guesses and which were based on knowledge. In order to avoid this situation, professional test-makers use

different techniques in designing tests. For example, they apply key balancing (to equal the right answers in the letters options) to resist bias like choosing most of the answers in the middle position or like choosing an answer based on one's own knowledge rather than on the reading passage (Fagley, 1987). However, this may not stop wild guessing completely; therefore, some authors suggest formula scoring which means taking out points for every incorrect answer and less point for omissions. Nonetheless, there are students who leave some options without trying to use deductive reasoning due to risk aversion. In this case, the multiple-choice test continues being unfair for those students. Thus, to be fair, first, it should be informed to test-takers that certain formula is being used to evaluate them, and second, to encourage risk aversion students to use deductive reasoning; it will be necessary to tell them to risk answering. This will be contradictory if the plan is to discourage guessing (Bar-Hillel et al., 2005). Finally, other researchers suggest that it will be fair to instruct test naïve students on TTS to balance students than to find a way to avoid wild guessing.

### 2.7.2 Testing reading

Evaluating receptive abilities has been a more difficult challenge than evaluating productive skills because they are not clearly observable (Ko, 2010). Therefore, there have been different concerns about testing reading. The main concern is to make a reading test to truly measure the reading ability and not any other unrelated factor. For example, Preston (1964), Tainman (1973), and Keets (1978) demonstrated that some test-takers score better when they do not read the passage in the test (cited in McPhail, 1981). This is what Cohen and Upton (2006)

call “selecting the right answers for the wrong reasons” (p. 9). For example, Ghafournia and Afghari (2013) found that test-takers take a reading test as any other test task where they did not learn anything from the passage. On the other hand, in their study they found that high proficiency reading students use more comprehending test-taking strategies more often than did students at the intermediate or low level. For Sheorey and Mokhtari (2001), reading is not a one-way process but one where the reader formulates hypotheses, tests prediction and uses world knowledge and the language to form meaning (cited in Fahim et al., 2010). Nonetheless, it seems that students depend more on the use of TTS rather than reading strategies when they are in a test situation. Supporting this, in Ghafournia and Afghari’s (2013) study it was found an interaction of the participants’ reading ability and the use of cognitive test-taking strategies in reading comprehension tests.

#### 2.7.2.1 Reading strategies

According to Pressley and Afflerbach (1995), reading strategies are divided into planning, monitoring, and evaluating (cited in Cohen & Upton, 2006). Besides, for Alavi and Bordbar (2012) reading strategies are “processes used by test-takers to enhance reading comprehension and overcome comprehension failures” (p. 3). As it was stated before, TTS are studied from the LLS perspective. However, there are strategies that students use when taking a test that they do not use when reading at home. That is why it is necessary to have a theory that includes TTS in the different four general ESL/EFL skills. For the moment, researchers analyze



reading TTS from the perspective of reading strategies. A complete definition of reading strategies is given by Carrell and Grabe (2002), which goes as follow:

“... it is clear that when reading, a reader engages in processing at the phonological, morphological, syntactic, and semantic and discourse levels, as well as engages in goal setting, text-summary building, interpretive elaborating from knowledge resources, monitoring and assessment of goal achievement, making various adjustments to enhance comprehension, and making repairs to comprehension processing as needed” (p. 234, cited in Cohen & Upton, 2006).

Besides, reading comprehension outlines three main perspectives to understand the elements of reading: the task, processing and reader purpose perspectives (Enright et al., 2000, cited in Cohen & Upton, 2006). Fahim et al. (2010), on the other hand, advice reading instructor to develop critical reading in their students. He proposes different critical reading strategies. Finally, one of the most helpful strategies that we can find in reading TTS is that test-takers should read a few questions before starting to read the passage. This way the testees could develop a prereading organizer which will help them to understand the text better. Also, usually the first two or three questions are the ones that involve the central idea.

### 2.7.3 Test validity

Even though test-takers and exams preparation teacher may appreciate the value of TTS, test makers do not do so that much because it decreases their tests validity. When students are able to answer an item correctly without using their knowledge is when the validity of the test comes into question. However, Maspons

and Llabre (1983) made the reasoning that if a student answers incorrectly not because of lack of knowledge, but lack of familiarity with the test format, does that increase the test validity? Thus, a possible validity problem may exist when it is a challenge to deduce the meaning of a test score (Rogers & Bateson, 1991). On the other hand, there are researchers that thank the contribution of TTS to build valid language tests (Bachman, 1990; Cohen, 1998, cited in Ghafournia, 2012). These researchers take TTS as an advantage to construct more valid and reliable language tests. This may work on the premise that if we can construct a flawed item, we can also identify one. This brings into question teacher-made tests. Sarnacki (1979) stated that not many teachers have the need, desire time or knowledge to construct reliable and valid tests. Usually once an item is written, it is barely evaluated. Thus instructing teacher in TTS may help them to be more aware in identifying flawed items in their exams. However, as it was mentioned before, it does not matter if the test is teacher-made or standardized, flawed items are usually present (Sarnacki, *ibid*). According to Cohen (2011), the use of TTS data to validate language tests is relatively new. As it is going to be discussed below, if students, low in TTS, are instructed in these skills, then TTS will be a constant. Therefore, it may be easier to make valid scores than valid tests.

#### 2.7.4 TTS instruction to overcome unfairness.

How many of us have had the misfortune of losing points from our score not because of lack of knowledge but due to unfamiliarity with a test format, especially when you present the TOEFL or FCE without any previous training or experience with these kinds of tests formats. One of the most common situations of losing

points among students in the TOEFL test is time. Students find themselves having the sensation that they are doing well on the test and suddenly the time is up, and they still have several items unanswered. Another common situation is when students choose the correct option but in the wrong item number, and usually they realize they have made that mistake only when they arrived to the last item and noticed that they have extra items but have run out of answers to choose from. For sure, as it was mentioned above, this is more probable to happen if students do not have any experience with psychometric tests. Therefore, as some authors will state in this paragraph usually this happen to minorities: students that come from neglected groups to whom education is not of good quality or good conditions, as well as other factors. For example, in the USA it has been noticed that Hispanic students' TTS is low compared to other similar population groups in the USA (Arroyo, 1982; Ginther, 1978, cited in Maspons & Llabre, 1983). Today, the use of standard selection exams by universities is common (BIÇAK, 2013). An example of this can be observed in some state university entrance tests where thousands of students (many of them from small towns) have to look for another institution or go back home because they were not able to obtain an acceptable score on a test format they have not faced before. Many of these students are the TW naïve who probably have not had much experience in taking a standardized exam with these psychometric characteristics even though they may have the knowledge and capability to succeed in a university major. McPhail (1981) asserts that:

Most minority students score lower on achievement and aptitude tests than the population at large (Brill, 1974; Slakter, Koehler & Hampton, 1970; Stanley, 1965; Stanley & Porter, 1967). Consequently, tests have taken on a sinister quality in

certain minority communities in the areas of education, employment and promotion.

Many suspect that test results are used to keep minorities from attaining higher education and good jobs (Green, 1975; Stone, 1971; Williams, 1972; Word, 1974).

Even though it may sound exaggerated that there is a conspiracy, it is difficult to deny that the educational system in Mexico is different for minorities (and usually the best minds prefer a position in a city than to teach in a far small town). Then, having lack of test experience and being test naïve is characteristic of certain groups of students. Another example is that much of the literature found for this research was carried out in developing countries from the Middle East, Latin America and Asia. Therefore, it can be assumed that when students are tested, they are evaluated on two aspects: their knowledge on the subject and their knowledge on taking psychometric tests (Kesselman-Turkel & Peterson, 1981). Consequently, being TW naïve these days is a disadvantage. However, many scholars agree on the multiple and great advantages that TTS instruction may have not just on test-takers' test performance, but also in the validity of scores and the affective filters on the test-taker (Sarnacki, 1979; McPhail, 1981; Maspons & Llabre, 1983; Fagley, 1987; Kouzekanani et al., 1989; Rogers & Yang, 1996; Douglas & Richard, 1998; Amer, 2007; BIÇAK, 2013; Fahim et al., 2010; Pour-Mohammadi & Abidin, 2011; Ghafournia, 2012; Ghafournia & Afghari, 2013). Fahim et al., (2010) for example, give one main reason to teach TTS: to improve validity on the test results. Ghafournia (2012) goes further stating that TTS should not only be taught but also implemented in future language teachers' curricula. Amer (2007) emphasizes that instruction in TW will help lower examinees' affective filters; therefore, they will become more relaxed, confident and enthusiastic about

taking tests. Fagley (1987) suggests that TTS instruction could contribute to adjust test scores. Even BIÇAK (2013) found positive correlation on test performance after a two week course on TTS. Kouzekanani et al. (1989) declare that many testees receive low grades because they lack TTS competence and that students are not going to become test wise just because they take many multiple-choice exams, they need to learn about it! Finally, Douglas and Richard (1998) declare that teaching TTS to students does in fact improve scores in exams like TOEIC and TOEFL. In this regard, only a few researchers found little or no evidence of the influence of the teachability of TTS on test performance (Flippo & Borthwick, 1981). Thus, we can be optimistic that TTS instruction will create more fair results among test-takers. As it was mentioned before, it is easier to obtain valid scores than to design valid tests.

#### 2.7.5 TOEFL

As it was mentioned before, TOEFL (Test of English as a Foreign Language) is a multiple choice test. Nevertheless, it is more than that. In some universities and workplaces, it is a gate keeper or filter for students to graduate or workers to be promoted. That is why every year an astronomical number of students enroll in preparatory courses to take this high-stake test for vocational or academic purposes. Originally, TOEFL was created to evaluate the English proficiency from international students whose language was not English and who wanted to study in a university in the USA or Canada. However, these days a number of academic institutions in different countries as well as independent organizations, agencies and foreign governments use it as a selecting instrument

(Douglas & Richards, 1998). Another TOEFL characteristic is that it only measures academic general English, not abilities like math, science or academic aptitude. For this endeavor, another kind of test should be administered by the institution. Therefore, a student who gets a high score in this test, does not necessarily mean that that student will be successful in a given program. As Douglas and Richard (ibid) stated, it is necessary to take into account previous academic records and other level and experience of the field of study. Next we are going to observe a description of the different section of the TOEFL with a special emphasis on the reading section.

#### 2.7.5.1 TOEFL description (reading section)

In general, TOEFL is divided into 3 sections, each one with its own subdivisions. The listening comprehension section is divided into three parts. Part A tests the ability to understand specific questions about what was said in short conversations. Part B tests the ability to answer general questions about what was said in short conversations. Finally, part C tests the ability to answer general questions about information in longer conversations and short lectures. The second section is called structure and written expressions. This section is divided into two parts. Part A tests the ability to complete sentences in grammatically correct English. Part B, on the other hand, tests the ability to identify errors in sentences. The last section of the three is reading comprehension which is not divided. It is composed of only one part that tests what is stated and implied in the readings (Douglas & Richards, 1998). However, it contains five readings followed by 10 questions each. Generally, there are only three types of questions in this

section: vocabulary, cohesion and reading comprehension items, which include questions about facts and main idea of a paragraph or the whole passage (Matthiesen, 1999, cited in Fahim et al., 2010). For Fahim et al. (2010), to answer these kinds of questions is linked in some way to thinking critically to be able to employ TTS.

## **2.8 Critical thinking**

Vaughn (2008) stated that critical thinking is “the systematic evaluation or formulation of beliefs, or statements by rational standards” (p. 830, cited in Fahim et al., 2010). Furthermore, critical thinking is defined by Paul (1990) as the ability to discipline the thinking development through efficient information processing. He also stated that critical thinking is learning how to ask and answer questions of analysis, synthesis, and evaluation (1985, p. 37, cited in Fahim et al., *ibid*). Following Paul's definitions, it is believed that training in test-wiseness helps students to develop critical thinking in the reading section of the TOEFL (McPhail, 1981). Therefore, if students are trained in TTS we will not only get more accurate scores and more confident students, but also we will develop an additional tool, which is critical thinking. Besides, this critical thinking process is much more demanding than other traditional processes like memorization and uncritical knowledge absorption. Finally, according to Fahim et al. (2010), there is a significant relation between testee's critical thinking and the TOEFL reading section performance.

## **2.9 TW for teachers**

Besides the different advantages that students may get from being instructed in TTS, there is another side of the coin that is necessary to take into account. Teachers may be willing to teach TW strategies to their students, but they could not know how to do it, and there may be some strategies that they may ignore and which could be critical in examinations. To tackle this deficiency, it would be necessary to include TTS instruction in the pre-service language teachers' programs. As Amer (2007) suggests "teachers should be well prepared, in theory and practice, to train their students to be test-wise, and to help them develop positive attitudes towards language tests" (p. 16). Besides, pre-service teachers may get some benefits from being instructed in TTS. For example, in a study carried out in Iranian EFL undergraduates, pre-service teachers showed a positive attitude towards teaching and learning TTS for multiple-choice English reading comprehension tests. This instruction could help them to increase their scores in their subject tests and in standard tests that they have to present to graduate (Pour-Mohammadi & Abidin, 2011). Moreover, to obtain positive results on student teachers training on TTS, it is necessary to take into account different aspects like the mastery of the TW teaching, the different effects that certain approaches may have on student teachers, the amount of training, among others (Flippo & Borthwick, 1981). Furthermore, it may not be enough to train pre-service teachers on TTS, but also some studies would be necessary to analyze how they transmit this knowledge to their students.



## 2.10 Conclusion

It has been showed in different studies that many students demonstrate poor test performance due to lack of TTS. Mistakes like poor use of time, anxiety and confusion due to unfamiliarity with the test format and procedure are related to lack of test-taking strategies competence. Therefore, we can conclude that when students present a test, they are really being tested on two aspects: how much they know about the subject and how much they know about taking a test (Kesselman-Turkel & Peterson, 1981). Also, other studies showed that the majority of the students lack test taking strategies in minority groups. However, teaching students to be testwise has showed to have different advantages. The first one is that TW has a positive impact on test performance. It helps students to solve problems and difficulties while taking a test. Second, TTS can be instructed from childhood to adulthood. Third, it reduces anxiety and improves students' attitude toward language tests. Fourth, TTS can be measurable and therefore, transmitted to testwise-naïve students because it is not a set of general abilities but a set of specific skills. Fifth, it makes tests to be fair with minorities (Flippo & Borthwick, 1981, Rogers & Yang, 1996, Pour-Mohammadi & Abibin, 2011). Finally, teachers can also get different benefits from TTS instruction. First, if these strategies are implemented in pre-service teachers' programs, they will not only know how to instruct their students to be testwise, but also, they will be able to use these strategies to do better on their own tests and, because they will identify these strategies, they could be able to write TW-proof items, which in turn will produce more accurate and valid scores on their language tests.

## **CHAPTER III: METHODOLOGY**

### **3.0 Introduction**

The purpose of this study was to find out which test-taking strategies were used by students who increased their TOEFL score significantly in only one semester. To carry out this study a number of steps were taken. They will be specified in the paragraphs below. Some of these steps were the selection of the participants, the questionnaires to be applied to them, which intended to extract information about the test-taking strategies they used, the taxonomy used to analyze the results and finally the findings. The first part will start with a brief description of the context of the study.

### **3.1 Context of research**

The TOEFL test is a standard exam that many universities in Mexico employ as a filter to students' graduation. The institution where this research was carried out is not an exception. It implemented different TOEFL scores requisites for different majors according to the language level they may need in their future careers. "Some majors may demand more English language knowledge than others" is usually the reasoning to set a required TOEFL score. That is the reason why some majors like medicine require students to get a 550 score while majors like Law need 450 points. This not only sets certain level of pressure on students with low English language competence, but also on the language department which has to give a solution to the number of students with low language level and to the university which expect a high number of students to graduate without

having the obstacle of not being able to pass a language test. The language department DELC (Departamento de Enseñanza de Lengua y Cultura) is responsible for achieving this goal. However, even though the language department has implemented a methodology (ACTFL) and textbooks, the TOEFL score level seems to be decreasing instead of increasing. The language department implemented different language level courses (acceso 1 and 2, plataforma 1 and 2, umbral, avanzado and conversación) as well as tutorial classes to students with low language level. It also implemented reading comprehension and writing textbooks to increase the TOEFL scores in the advanced level. The courses that use the reading and writing textbooks are the umbral and avanzado. In the last level is where students have to do the TOEFL test and present a required score.

### **3.2 Participants**

The participants for this study were students in the umbral level and who were in a semester previous to present the TOEFL. These students came from different majors and from different semesters. Three boys and a girl were selected as it will be indicated below. The majors they were studying were Architecture, Marketing, Biotechnology and Medicine, respectively. Each student came from a different umbral group.

### **3.3 Instruments**

To gather data a questionnaire was adapted from two questionnaires used in previous studies for the purpose of this study. One was a questionnaire used by

Cohen and Upton (2006) who attempt to identify which strategies examiners use in the new TOEFL reading tasks. The other questionnaire was from Kashkouli and Barati (2012) who were trying to analyze the different test-taking strategies used by testees in task-based reading assessments. None of these two questionnaires were appropriate for this study because of the following reasons. The first questionnaire included some items that were appropriate only for the iBTTOEFL test which includes dragging. This of course is not an activity included in the PBT TOEFL reading section. On the other hand, the questionnaire used by Kashkouli and Barati was not completely appropriate for this study because it was designed to measure the strategies used in the FCE test which includes task-based readings as well as reading comprehension, so some of the items include activities that are not common in the PBT TOEFL reading section. These are the reasons why the questionnaire used in this study had to be adapted from these two questionnaires. Moreover, the questionnaire employed Likert type scale options that were from 1 meaning never to 5 meaning always. In appendix A, the final questionnaire can be observed. Furthermore, the questionnaire was translated into Spanish (the students' native language) to avoid any ambiguity in their understanding (Kashkouli & Barati, 2012). Finally, the questionnaire was reviewed by two professors in the TEFL (Teaching English as a Foreign Language) area to validate its objectivity and clarity of the items. Some questions were modified to improve their clarity and comprehension according to the professors' suggestions.

### **3.4 Data collection procedure**

There were different moments to collect data from the participants. It started with a selection of the participants. First there was a sample TOEFL test at the beginning of the semester. This test was applied in three moments, one each day due to shortness of time and students' availability. All the students from 4 groups applied this sample test. Each group had between 15 and 20 members. All students were at the umbral level using the two textbooks (reading and writing). Because of the program, they only study until half of each book. The reading text includes readings to develop critical thinking which, according to the Language Department, will help students develop reading comprehension. The classes were taught by the same teacher in the four groups. At the end of the semester, all the groups presented a second sample TOEFL test following the same procedure as the first one. It may be proper to emphasize that because the students did not present the complete test in one day, they may not be tired or exhausted as usually happened in the standard test. However, both the first and the second test were applied following the same procedure. At this moment their scores were analyzed and only the students who increased 50 points in the second sample test were selected resulting in four students. It is important to emphasize that these students were not the ones with the higher language level. Actually those students increased their score from 7 to 20 points and some decreased from 10 up to 40. The students who were selected achieved 50 points after only one semester and most of the improvement was in the TOEFL reading section. We could infer that these students increased their scores in that section because they study a course

where reading comprehension was emphasized. This is true in the fact that most of the score increase from all the students was in the reading section, but these particular students were the only ones who increased significantly compared to the rest of the other students. Finally, the subjects answered a test-taking strategy questionnaire adapted from Cohen and Upton (2006), and Kashkouli and Barati (2012).

#### 3.4.1 Data analysis

When all the four participants' responses were gathered, they were analyzed by classifying them according to Millman's et al. (1965) testwiseness taxonomy. There are other taxonomies (Watter & Siebert, 1990; Wensen, 1991) but according to Amer (2007), Millman's et al. taxonomy is the one that has been more cited in different studies. Millman's taxonomy is divided into elements independent of the test constructor and elements dependent upon the test constructor or purpose. Each of these categories contains different subcategories. The first one is divided into four subcategories: time-using strategies, error-avoidance strategies, guessing strategies, and deductive reasoning strategies. The second category is divided into two subcategories: intent consideration strategies and cue-using strategies. The items in the questionnaire that was applied to students were classified according to Millman's taxonomy. This way each question belongs to one of the Taxonomy subcategories (time using, error-avoidance, etc.). Then, the students' answers in the questionnaire were placed in each of the classified questions. Finally, the strategies with the higher numbers were observable. This method allowed to observe which strategies were used more

often by students when they were responding to the reading section of the TOEFL test.

### **3.5 Conclusion**

The attempt of this chapter was to explain the different steps followed in the methodology of this study. It was explained and justified why a questionnaire was adapted and presented in Spanish to the students, how the different students were classified to apply the strategy questionnaire and how Millman's taxonomy helped to classify and find out which strategies these students used the most. Now, it is the turn of chapter four to explain the results and conclusion of the present study.

## **CHAPTER IV: RESULTS AND DISCUSSION**

### **4.0 Introduction**

In order to understand this study, it is necessary to go deeply into the results obtained during the research. The result of a questionnaire adapted to the purpose of the study as well as the use of a taxonomy to classify the test-taking strategy is a process that needs to be explained in order to make the process valid. Therefore, the purpose of this study is to analyze and interpret these results as closely as possible to draw an accurate conclusion. Below, this analysis is presented in detail.

### **4.1 Analysis of the TOEFL results**

First of all, it is necessary to analyze the different results obtained in the TOEFL test. As it was explained in the literature review, the TOEFL test is a high-stake test that measures academic English language. It does not measure academic content; therefore, a student with high language proficiency does not necessarily represent high academic proficiency. Other tests and records are necessary to determine academic proficiency in other areas than languages (Douglas & Richards, 1998). Therefore, the next charts present the four groups of students' TOEFL scores in the different sections like listening comprehension, structure and written expression, and reading comprehension in the first and second sample test.



Table 1. Group 1 first and second TOEFL scores.

student	Major	First TOEFL Scores				Second TOEFL Scores				Diff
		Sec 1	Sec 2	Sec 3	TOTAL	Sec 1	Sec 2	Sec 3	TOTAL	
S1	Ing. Química	40	34	31	350	41	38	37	387	37
S2	Medicina	42	38	42	407	43	37	45	417	10
S3	Medicina	43	34	42	397	43	42	45	433	37
S4	Medicina	55	45	49	497	54	42	48	480	-17
S5	Medicina	58	46	54	527	58	45	49	507	-20
S6	Psicología	60	45	52	523	57	46	59	540	17
S7	Odontología	42	37	45	413	40	40	46	420	7
S8	Odontología	45	34	37	387	40	40	43	410	23
S9	Odontología	53	41	46	467	49	43	48	467	0
S10	Ing. Mecatrónica	39	40	39	393	36	32	43	370	-23
S11	Ing. Agronomía	42	34	35	370	41	29	37	357	-13
S12	Diseño y Produc. Pub.	46	28	33	357	42	37	36	383	27
S13	Mercadotecnia	47	39	48	447	52	34	52	460	13
S14	Mercadotecnia	55	48	54	523	50	42	52	480	-43
S15	Nutrición	51	44	47	473	49	44	39	440	-33
S16	Nutrición	49	42	47	460	49	41	47	457	-3
S17	Comunicación	39	34	38	370	37	38	37	373	3
S18	Psicología	61	43	52	520	57	48	55	533	13
AVERAGE		48	39	44	438	47	40	45	440	2

As it can be observed in Table 1, the students that are marked in yellow were the students that the teacher perceived as high proficiency. This can be

corroborated in their scores. However, their improvement was between 13 and 17 points. As it was stated before, this study was looking for those students who improved above 50 points or more. Observing the differences in scores between the first and second tests, it can be observed that only four students (S1, S3, S8 and S12) obtained a significant increase between 23 and 37 points in this group. That is the reason why none of these subjects were candidates for this study. Nevertheless, it is worth noticing that as there were students who improved, there were also a number of students who lost points. 11 students increased their scores from 3 to 37 points, but 7 students lost between 3 to 43 points. In general, as it can be seen in Table 1, the group had a total average increase of 2 points which was very low. Besides the emphasis on reading in class, this group only presented 1 point of increase in the reading section

The second group presented an increase in the reading section of 4 points in average which can be seen in Table 2. As it was presented before, most of the class was focused on reading comprehension, so it was already expected, although not this low. The different results obtained by students in the first and second test can be observed in Table 2 below.

Table 2. Group 2 first and second TOEFL scores.

stud ent	Major	First TOEFL Scores				Second TOEFL Scores				Diff
		Sec 1	Sec 2	Sec 3	TOTAL	Sec 1	Sec 2	Sec 3	TOTAL	
S1	Veterinaria	39	30	48	390	37	38	43	393	3
S2	Admo. De Inst. Edu.	44	42	35	403	47	41	42	433	30

S3	Fisioterapia	42	42	39	410	43	42	43	427	17
S4	Diseño y produccion Pub.	48	36	48	440	49	39	48	453	13
S5	Medicina	43	34	37	380	42	42	46	433	53
S6	Veterinaria	46	44	42	440	47	39	48	447	7
S7	Diseño y produccion Pub.	42	32	38	373	41	37	40	393	20
S8	Fisioterapia	35	38	40	377	36	37	36	363	-13
S9	Nutrición	39	39	39	390	35	30	41	353	-37
S10	Ing. Mecatronica	36	42	30	360	40	38	33	370	10
S11	Admon. De Instituciones	41	45	41	423	44	39	46	430	7
S12	Nutrición	49	46	43	460	53	48	43	480	20
S13	Admon. De Instituciones	41	34	36	370	40	37	35	373	3
S15	Diseño y produccion Pub.	46	37	38	403	50	40	44	447	43
S16	Ing. Mecatronica	44	37	28	363	39	38	40	390	27
	AVERAGE	42	39	38	399	43	39	42	412	13

In the above chart, four students (S2, S5, S15 and S16) had an important increase in their scores which go from 27 to 53 points between the first and second test as we can see in the last column. Again, the students marked in yellow are the students that presented a higher language competence during the course compared to the rest of the group as their scores confirmed. Nonetheless, their improvement ranges from 7 to 20 points. Besides, there were students (S8, and S9) who lost points which ranges from 13 to 37 points as can be observed in the last column. In this group there was a total improvement of 13 points in average and 4 point in the reading section. From group 2 in Table 2, students 5 and 15

obtained a significant improvement; however, only student 5 was taken into consideration into the study.

In group 3 Table 3 below some unexpected results were obtained. As before, the students that the teacher reported as having a high level of proficiency are marked in yellow. The point here is that three (S1, S5 and S6) of the four students marked in yellow presented a decrease in their scores. This was not expected because in most of the groups the high language proficiency students usually present an increase. This decrease goes from 17 to 40 points, which is very significant. Another analysis would be necessary to determine what caused such a decline. On the other hand, here the score increase had an impact in the group's average. The scores rose from 10 to 80 points as can be seen in the last column in Table 3. In general, the group had an increase of 19 points as it can be observed below.

Table 3. Group 3 first and second TOEFL scores.

stud ent	Major	First TOEFL Scores				Second TOEFL Scores				Diff
		Sec 1	Sec 2	Sec 3	TOTAL	Sec 1	Sec 2	Sec 3	TOTAL	
S1	Derecho	56	51	54	537	58	36	56	500	-37
S2	Derecho	40	32	34	353	41	34	35	367	13
S3	Fisioterapia	43	30	31	347	35	34	43	373	27
S4	Diseño y produccion Pub.	40	36	42	393	40	38	46	413	20
S5	Psicología	58	43	48	497	53	41	50	480	-17
S6	Medicina	50	51	54	517	52	48	43	477	-40

S7	Medicina	42	30	42	380	44	41	43	427	47
S8	Arquitectura	46	37	40	410	51	41	47	463	53
S9	Gastronomía	43	29	39	370	34	30	45	363	-7
S10	Mercadotécnia	39	32	30	337	44	38	43	417	80
S11	Admo. De Inst. Edu.	43	39	31	377	37	32	41	367	-10
S12	Diseño Automotriz	43	41	35	397	47	43	43	443	47
S13	Gastronomía	34	38	31	343	36	32	48	387	43
S14	Diseño y produccion Pub.	53	45	41	463	50	46	50	487	23
S15	Fisioterapia	39	34	37	367	39	38	41	393	27
S16	Derecho	39	36	34	363	41	34	46	403	40
S17	Derecho	57	49	49	517	61	40	57	527	10
	AVERAGE	45	38	40	410	45	38	46	429	19

This is the group which presented two different unexpected scores. One is the decrease in the TOEFL final score from 17 to 40 points (as can be seen in S5 and S6), and the other one is the high increase that the low proficiency students had, ranging from 13 to 80 points (S2 and S10) as can be observed in Table 3 above. Another observation, as in the previous groups (Table 1 and 2), is the increase in the reading section (6 pts) which, as before, was expected because of the reading emphasis in class. However, this is the group which presented the highest score among the four groups. From this group, two students were selected (S8 and S10) due to their high score increase of more than 50 points, which is of importance to this research.

Finally, there was a fourth group (Table 4) that was analyzed. This group also presented some unexpected results as can be observed in the table below.

Table 4. Group 4 first and second TOEFL scores.

student	Major	First TOEFL Scores				Second TOEFL Scores				Diff
		Sec 1	Sec 2	Sec 3	TOTAL	Sec 1	Sec 2	Sec 3	TOTAL	
S1	Medicina	51	44	46	470	51	41	44	453	-17
S2	Gastronomía	34	41	41	387	37	38	38	377	-10
S3	Gastronomía	57	47	46	500	55	37	48	467	-33
S4	Ing. Diseño Auto.	60	47	53	533	55	50	52	523	-10
S5	Medicina	48	43	39	433	47	38	46	437	3
S6	Esp y Est Latinoamer.	47	43	44	447	47	47	46	467	20
S7	Medicina	41	39	45	417	35	26	41	340	-77
S8	Admon de Inst. Edu.	51	54	49	513	51	49	46	487	-27
S9	Medicina	41	44	39	413	40	40	46	420	7
S10	Derecho	39	34	33	353	35	36	39	367	13
S11	Biotecnología	40	41	35	387	43	45	43	437	50
S12	Administración	44	38	39	403	48	36	44	427	23
S13	Medicina	46	37	44	423	40	40	44	413	-10
AVERAGE		46	42	43	437	45	40	44	432	-5

Many of the students in this group had very low language proficiency. Even the students that the teacher reported as having a high proficiency, in relation to the rest of the group, had a low level compared to the other groups. As can be seen in Table 4, they reported also a decrease in their scores from 10 to 33 points

(from S2, S4, S13 to S3). On the other hand, low language level students had an increase from 3 to 50 points (S5 and S11). However, that did not increase the group average which had a decrease of 5 points but an increase of 1 point in the reading section (Table 4). As in the previous group, an in depth analysis would be necessary to determine the different factors that cause that those high proficiency students lower their scores. In this group, only one student could increase his score in 50 points despite this group was the one that showed not only the lowest TOEFL score, but also a decrease of the four chosen groups.

The above tables presented how the four students were selected from the four groups of students who presented the TOEFL test. Now, an analysis of the question classification will be done in the next paragraph.

#### **4.2 Analysis of question classification into Millman's taxonomy**

As it was stated in the previous chapter, a questionnaire was adapted taking appropriate parts from other two questionnaires. One was a questionnaire developed by Kashkouli and Barati (2012) whose intention was to analyze the different test-taking strategies that subjects employed in reading tasks in the FCE (First Certificate). The second questionnaire was taken from Cohen and Upton (2006) whose study intended to determine which reading and test-taking strategies students use when answering reading tasks from the new iBT TOEFL reading section. After analyzing these two questionnaires, a new questionnaire was developed for the purpose of this study (see Appendix A) which included 40

questions with Likert type scale options intending to obtain information from the different strategies students used during the test.

Later the questions in the resulting questionnaire were classified according to Millman's et al. (1965) taxonomy. As it was stated in chapter 3, there are other taxonomies (Watter & Siebert, 1990; Wensen, 1991) but according to Amer (2007), Millman's et al. taxonomy is the one that has been more cited in different studies and has provided a general framework for further studies of TW (Rogers & Bateson, 1991; Rogers & Yang, 1996). This taxonomy is divided into two broad categories: I. Elements independent of the test constructor or purpose and II. Elements dependent upon the test constructor or purpose. In the first category there are four subcategories which are: A. Time using strategies, B. Error-avoidance strategies, C. Guessing strategies, and D. Deductive reasoning strategies. The second category also includes some subcategories: A. Intent consideration strategies and B. Cue-using strategies. The different questions from the questionnaire were classified according to this taxonomy as it can be observed in the next chart.

Table 5. Millman's et al. (1965) taxonomy and classification of questions.

Millman's taxonomy	Questionnaire's items classification	Total
I. Elements independent of the test constructor or purpose		
A. Time using strategies (IA)	1, 2, 3, 5, 6, 7, 27, 28, 29, 36, 37, 39.	12
B. Error-avoidance strategies (IB)	4, 19, 20, 21, 30, 35, 38, 40.	8
C. Guessing strategies (IC)	17, 31, 32, 33.	4



D. Deductive reasoning strategies (ID)	18, 25, 26, 34.	4
II. Elements dependent upon the test constructor or purpose		
A. Intent consideration strategies(IIA)	12, 13, 14, 15, 16.	5
B. Cue-using strategies (IIB)	8, 9, 10, 11, 22, 23, 24	7

From this chart, it can be observed that most of the items in the questionnaire corresponded to time using strategies (with 12 items). This was taken into consideration to assure that this fact was not going to alter or give faulty results. A bigger number of questions fell into IA strategy due to some questions seemed to be repeated like question 5 and 6: “En cada texto determiné que preguntas eran más difíciles y las respondí antes que las fáciles”, “En cada texto determiné que preguntas eran más fáciles y las respondí antes que las difíciles” respectively. Here we could have asked only one question and assumed that if the students’s response was negative, the other was positive. However, it was intended not to leave some answers to assumptions. In the next chapter, it is going to be analyzed the total points each strategy obtained according to the Likert scale responses.

### **4.3 Analysis of the classification of answers according to the used strategies**

After the scrutiny above, the different answers points from the Likert type scale options were analyzed as it is presented in the following tables.

Table 6. Total points of time-using strategies.

Question	S1	S2	S3	S4	Total	Strategy
Q1	5	4	4	4	17	IA
Q2	4	3	3	3	13	IA
Q3	5		3	4	12	IA
Q5	3	2	1	3	9	IA
Q6	3	3	1	3	10	IA
Q7	3	4	5	4	16	IA
Q27	2	1	3	4	10	IA
Q28	2	1	3	2	8	IA
Q29	4	5	5	3	17	IA
Q36	5	4	4	3	16	IA
Q37	3	4	4	2	13	IA
Q39	2	1	4	3	10	IA
					151	TOTAL

In Table 6 it can be observed that questions 1 and 29 showed the use of time-using strategy the most. Q1 explained that students had a plan and followed it. Q29 showed that besides that plan, in time using strategies, students do not look for the easy readings, they answer them according to how they appear in the test. This in fact illustrates that students did not use the strategy. Instead of looking for the easier readings (the ones that they may be more familiarized to) and let the difficult ones at the end to economize time, they answered them as they appeared. Q5 and 28 are on close acquaintance with the previous questions in the fact that in each reading they did not use time strategically and they did not look for the easiest readings.

Table 7. Total points of error-avoidance strategies.

Question	S1	S2	S3	S4	Total	Strategy
Q4	1	5	4	4	14	IB
Q19	4	3	5	3	15	IB
Q20	5	5	5	4	19	IB
Q21	3	4	3	4	14	IB
Q30	2	4	4	1	11	IB
Q35	2	3	1	2	8	IB
Q38	3	3	5	4	15	IB
Q40	5	3	5	3	16	IB
					112	TOTAL

Q20 presents a tendency to avoid errors by paying careful attention to the selected option. Q35 may present two interpretations. One is that students did not make use of note taking strategy or that they followed the TOEFL rule which states that students should not take notes or write on the test.

Table 8. Total points of guessing strategies.

Question	S1	S2	S3	S4	Total	Strategy
Q17	1	5	1	2	9	IC
Q31	1	3	3	2	9	IC
Q32	2	5	2	3	12	IC
Q33	1	3	2	3	9	IC
					39	TOTAL

In Table 8 it can be observed that the guessing strategy is one of the strategies used the least. This indicates that students barely left some questions to

randomization. Q32 indicates that sometimes students choose some answers according to their previous knowledge on the topic. This may be also interpreted as trying to economize time.

Table 9. Total points of deductive reasoning strategies.

Question	S1	S2	S3	S4	Total	Strategy
Q18	2	4	5	3	14	ID
Q25	5	4	2	3	14	ID
Q26	5	5	5	3	18	ID
Q34	3	5	4	4	16	ID
					62	TOTAL

Here students try to deduct the answer by using logic and discrimination. Obviously, to do so they had to make use of their previous knowledge. That is the reason why Q26 had a high score.

Table 10. Total points of intent consideration strategies.

Question	S1	S2	S3	S4	Total	Strategy
Q12	4	3	3	2	12	IIA
Q13	4	5	5	4	18	IIA
Q14	4	4	4	3	15	IIA
Q15	2	3	2	4	11	IIA
Q16	4	5	5	4	18	IIA
					74	TOTAL

On the one hand, intent consideration strategy showed that it was one of the strategies used most of the time. Q13 and 16 confirm this by students trying to

understand the meaning of the question. On the other hand, Q15 showed a low frequency; however, it may be interpreted as positive since the question implies that students most of the time did not try to guess.

Table 11. Total points of cue-using strategies.

Question	S1	S2	S3	S4	Total	Strategy
Q8	4	2	5	3	14	IIB
Q9	5	3	5	4	17	IIB
Q10	1	3	1	4	9	IIB
Q11	3	4	4	4	15	IIB
Q22	4	5	4	4	17	IIB
Q23	2	3	3	2	10	IIB
Q24	3	3	4	4	14	IIB
					96	TOTAL

Finally, cue-using strategy was frequently used in questions Q9 and Q22. Here students made use of the different cues and clues that the text and the vocabulary present in order to arrive to a correct answer. Q10 indicates that the subjects did the opposite, in other words, they look for clues in the passage and questions before choosing and answer. In the next paragraph the final results are going to be analyzed to interpret the research.

#### 4.4 Analysis of results

At first sight it seems that the strategies that were used the most were time using (IA) and error-avoidance (IB). However, it has to be taken into consideration that strategy IA and IB were the ones that had most of the number of questions. In

order to obtain accurate results, it is necessary to divide the number of questions into the total points. Thus, then the next table presents the final results.

Table 12. Final results

Strategy	Total of Answers	Number of Questions	Average
Time-using (IA)	151	12	12.6
Error avoidance (IB)	112	8	14.0
Cue-using (IIB)	96	7	13.7
Intent consideration (IIA)	74	5	14.8
Deductive reasoning (ID)	62	4	15.5
Guessing (IC)	39	4	9.8

In this table it can be detected that deductive reasoning was the strategy used the most, followed by intent consideration. Error-avoidance and cue-using are very close in use with 14 and 13.7 points in average followed by time-using strategies. The strategy that seems to be used the least is guessing with 9.8 used in average.

#### 4.5 Conclusion

As it is presented here, the strategy used the most was deductive reasoning. The ones that students used more frequently were intent consideration, cue-using, and error avoidance strategies. In order to arrive to this, average frequency in use was the parameter that was employed in order to get conclusions. In the next chapter, these conclusions are going to be seen in depth to determine

why these students use some strategies more than others and how these findings could be beneficial for other students despite its limitations.

## CHAPTER V: CONCLUSIONS

### 5.0 Introduction

Standardized language tests are examinations that determine a promotion in many workplaces and universities. Students and teachers look for different learning methods and strategies to succeed on these examinations. According to Bar-Hillel et al. (2005), even high-stake tests are also susceptible to testwiseness. As it was mentioned in chapter two, in order to apply certain test-taking strategies, the testee needs to possess certain balance between TTS and knowledge about the subject being tested. If one of these components is low, the success in the test will be reduced. As Kesselman-Turkel and Peterson (1981) stated, because of the increased use of psychometric tests, these days we are evaluated on two aspects: how much we know about a subject matter, and our ability to take tests. Because of this, it is necessary to prepare not just our students, but also future language teachers to promote the TTS learning in the classroom. As many researchers have stated, it is easier to obtain valid scores than valid tests. Then TTS will be a constant, not a matter that can invalidate a test. For these reasons, this research was carried out. The search for those TTS that some students use, besides their knowledge, to succeed in language tests, will facilitate students and teachers to obtain better scores in the TOEFL test. This in turn will aid students to graduate and to promote more studies on TTS which according to Allan (1992, cited in Amer, 2007) has been a neglected area. Finally, teaching TTS to our students will provide them with confidence and the chance to be evaluated on their knowledge and not on how much they know about taking tests.



An analysis of the different findings in this research is what chapter five is about, and it will attempt to answer the research questions.

### **5.1 Answers to the research questions**

Looking for an answer is not an easy endeavor, especially when that answer is hidden in the decisions and knowledge of our students when taking a test. There are different methods to obtain information from learners when they perform a task, all of them with their advantages and disadvantages. Before answering the research questions, here are some findings. First, not always the high proficiency students were the ones who got most of the points increase. Besides, low level students proved that they can present a very high improvement, even more notorious than the high level students. Another finding was that after an 80 hour course mainly based on reading, the increased in the TOEFL reading section was very low (from 1 to 6 points). Now let's move on to the research questions.

In this study a practice TOEFL test and an adapted questionnaire from Cohen and Upton (2006), and Kashkouli and Barati (2012) as well as Millman, Bishop and Ebel's (1965) taxonomy were utilized to answer the research questions: what test-taking strategies do students make use of when answering the TOEFL reading section? And from these strategies, which one do they apply more frequently? It would be interesting to mention that these research questions appear after applying the second practice test and analyzing the results. There was an idea in the air, but it was clearer only after this study. The general question was how only these students improved 50 points in just one 80-hour course? After applying them a questionnaire about TW, it could be observed which TTS they applied the most.

Taking frequency of use as the measure to determine which strategies were used the most, the chart below presents this average strategy use.

Table 1 Strategy Use Average

Strategy	Total of Answers	Number of Questions	Average use
Time-using (IA)	151	12	12.6
Error avoidance (IB)	112	8	14.0
Cue-using (IIB)	96	7	13.7
Intent consideration (IIA)	74	5	14.8
Deductive reasoning (ID)	62	4	15.5
Guessing (IC)	39	4	9.8

In this study frequency of use was determinant to draw some conclusions. This helps us to determine which strategies were used more frequently and which one was used the most (Salehi, 2011). It can be observed that strategies like intent consideration, cue-using, error avoidance and time-using were the strategies that learners used the most to increase their results in this test. Guessing was the strategy that students use the least and we may speculate that it was carried out only when any of the other strategies were not enough to obtain the correct option and/or because of lack of knowledge about the topic. However, making use of their background knowledge and using deductions was the tactic that probably was the responsible for such a high increase in these test takers' scores, mainly in the reading section.

It can be observed that the strategy students used most of the time was deductive reasoning. Therefore, as it was said, even high-stake standardized tests

are susceptible to TW. That is why some researchers state that it is easier to get reliable scores than reliable tests.

As it was mentioned before, these students made use of different strategies to achieve high scores in the TOEFL test, but the one they used the most was the deductive reasoning strategy with all its principles. Besides, the other strategies that the subjects used very frequently were intent consideration, error avoidance, and cue-using. This leads us to think of what can be done with this information and how important it can be to future research.

## **5.2 Recommendations for further research**

If an institution uses a high-stake test as the TOEFL as a gate keeper to determine which students are prepared enough to graduate, then it would be fair to prepare these students for this kind of test. Teaching them a second language is not enough to achieve good scores. As it was mentioned before, these days we are not just evaluated on how much we know about a subject, but also on how much we know about test taking strategies, especially with tests whose formats are usually different from what we are used to. Also, it is well known that our evaluations must be as close as what we practice in the classroom. Therefore, one first step would be to have our students to practice TOEFL-like evaluations in our classroom. Because scores do not only show the linguistic abilities that a subject possesses, but also the capability to benefit from the characteristics and structure of a test (Belcher, 1985 cited in Pour-Mohammadi and Abidin, 2011), the second step would be to prepare them to use some TTS. As it had been mentioned before, even high-stake tests are susceptible to TW, so preparing our students in this

matter could result in more reliable scores. But first, it would be necessary to know which strategies could be more appropriate and identify which strategies our students already use. In this study, it could be observed which strategies students who got a remarkable score used more frequently and which strategy they used the most. This could help teachers, administrators and curricula developers to make some decisions on what else to teach in the language classroom. However, there are still many questions that are still unanswered. Therefore, further research could be done to complement the limitations of this study which are described below.

Because the research questions to do the study on this topic appear after the second practice TOEFL test was administered, some aspects were not considered and they could be taken into consideration to obtain more accurate results in future studies. For example, even though the questionnaire was reviewed by two other colleagues, it could also be piloted with another group of students before it is applied to the research group. This could help to determine if there were some items that were not clear for students specifically, so some items could be restated or more questions could be added. Besides, the same questionnaire can be applied twice to all the students. One, after the first TOEFL practice test, and then at the end of the course, after the second TOEFL practice test to observe which strategies the subjects, who got higher scores, used before and after. This way we could also determine two more aspects. One, if the study subjects used the same strategies in both tests. Second, we could compare the strategies these subjects used to the rest of the strategies the rest of the group used. This way we could get to more accurate conclusions.

Furthermore, if time and resources are available, we might apply the same study to different proficiency groups to determine first which strategies each level used, and then determine which strategies, according to their level, students may need or be able to use. It has been observed in other studies that students with different proficiency levels use different strategies (Amer, 2007; Kashkouli & Barati, 2012). But it would be necessary to determine which TTS each level uses so students could be instructed gradually, and get acquainted to them.

To conclude with this study, it can be said that it helped to obtain a wider understanding of what a score means and not to neglect TW naïve students, or assume that every students is familiar with high-stake test formats like the TOEFL, particularly when it could have a long term effect in the future of the learner (Baleghizadeh & Yousefian, 2012). It is important to remember that general instruction does not include testwiseness and that taking several multiple choice tests does not guarantee that a student will become test wise.

## References

- Alavi, S. M., & Bordbar, S. (2012). A closer look at reading strategy use in reading section of toeflibt. *Theory and practice in language studies* 2, 3, 450-460. Retrieved from [www.academypublication.com/issues/past/tpls/vol02/03/04.pdf](http://www.academypublication.com/issues/past/tpls/vol02/03/04.pdf)
- Alavi, S. M., & Bordbar, S. (2012a). Are they right participants for the right strategies? A case study in the role of levels of language ability in strategy use in reading section of toeflibt. *Theory and practice in language studies* 2, 5, 877-886. Retrieved from [www.academypublication.com/issues/past/tpls/vol02/05/02.pdf](http://www.academypublication.com/issues/past/tpls/vol02/05/02.pdf)
- Amer, A. A. (2007). Efl/esl test-wiseness and test-taking strategies. Retrieved from [files.eric.ed.gov/fulltext/ED497399.pdf](http://files.eric.ed.gov/fulltext/ED497399.pdf)
- Baleghizadeh, S., & Yousefian, Y. (2012). The Relationship Between Test-Takers' Rate of strategy use and their reading comprehension performance. *The NERA journal* 48, 1, 73-77. Retrieved from <http://connection.ebscohost.com/c/articles/81281411/relationship-between-test-takers-rate-strategy-use-their-reading-comprehension-performance>
- Bar-Hillel, M., Budescu, D., & Attali, Y. (2005). Scoring and keying multiple choice tests: a case study in irrationality. *Mind and society*, 4, 3-12. Retrieved from <https://ideas.repec.org/p/huj/dispa/dp370.html>
- BIÇAK, B. (2013). Scale for test preparation and test taking strategies. *Educational sciences: Theory and practice*, 13, 279-289. Retrieved from <https://eric.ed.gov/?id=EJ1016653>
- Cohen, A. D. (2011). *Strategies in learning and using a second language (Second edition)*. London, UK: Longman.
- Cohen, A. D., & Upton, T. A. (2006). Strategies to respond to the new toefl reading tasks. *Monograph series*, 33. Retrieved from <https://www.ets.org/Media/Research/pdf/RR-06-06.pdf>
- Cohen, A. D., & Macaro, E. (2007). *Language learner strategies : thirty years of research and practice*. Oxford ;. Oxford University Press.
- Douglas, E. F., & Richard, K. (1998). Teaching toeic/toefl test-taking strategies. *TESOL*, 32. Retrieved from <https://eric.ed.gov/?id=ED427543>

- Fagley, N. S. (1987). Positional response bias in multiple-choice tests of learning: Its relation to testwiseness and guessing strategy. *Journal of education psychology, 79*, 95-97. Retrieve from <https://eric.ed.gov/?id=EJ348476>
- Fahim, M., Bagherkazemi, M. & Alemi, M. (2010). The relationship between test takers' critical thinking ability and their performance on the reading section of toefl. *Journal of language teaching and research 1, 6*, 830-837. Retrieved from [www.academypublication.com/issues/past/jltr/vol01/06/11.pdf](http://www.academypublication.com/issues/past/jltr/vol01/06/11.pdf)
- Flippo, R. F., & Borthwick, P. (1981). Should testwiseness curriculum be a part of undergraduate teacher's education. *American reading forum, 2*. Retrieved from <https://eric.ed.gov/?id=ED218591>
- Ghafournia, N., & Afghari, A. (2013). The interaction between cognitive test-taking strategies, reading ability, and reading comprehension test performance of iranianefl learners. *English language teaching 6, 8*, 139-150. Retrieved from URL: <http://dx.doi.org/10.5539/elt.v6n8p139>
- Ghafournia, N. (2012). The relationship between using multiple-choice test-taking strategies and general language proficiency levels. *Procedia-social and behavioral sciences 70*, 9-94. Retrieved from <http://www.sciencedirect.com/science/article/pii/S187704281300044X>
- Gharbavi, A., & Mousavi, S. A. (2012) Do Language Proficiency Levels Correspond to Language Learning Strategy Adoption? *English language teaching 5, 7*, 110-122. Retrieved from <http://dx.doi.org/10.5539/elt.v5n7p110>
- Kashkouli, Z., & Barati, H. (2012). Type of test-taking strategies and task-based reading assessment: A case in Iranian EFL. *Social and Behavioral Sciences 70*, 1580-1589. doi: 10.1016/j.sbspro.2013.01.226
- Kesselman-Turkel, J & Peterson, F. (1981). *Test-taking Strategies*. USA: The University of Wisconsin Press.
- Ko, M. H. (2010). A comparison of reading comprehension tests: Multiple-choice vs. open-ended. *English Teaching, 65, 1*, 137-159. Retrieved from [journal.kate.or.kr/wp-content/uploads/2015/05/kate\\_65\\_1\\_5수정.pdf](http://journal.kate.or.kr/wp-content/uploads/2015/05/kate_65_1_5수정.pdf)
- Kouzekanani, K., Llabre, M. M., & Baldwin, R. S. (1989). Analysis of Testwiseness

Components: A quasi-experimental approach. *American education research association*, 73. Retrieved from <https://eric.ed.gov/?id=ED306239>

Lightbown, P. M. & Spada, N. (2006). *How languages are learned. Third edition.* China: Oxford University Press.

Maspons, M. M., & Llabre, M. M. (1983). Effect of pretraining hispanic students on test-taking strategies on the reliability and predictive validity of a mathematics predictor test. *American education research association*, 67. Retrieved from <https://eric.ed.gov/?id=ED229431>

McPhail, I. P. (1981). why teach test wiseness? *Journal of reading* 25, 1, 32-38. Retrieved from <http://www.jstor.org/stable/40030492>

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition.* NY, USA: Cambridge University Press.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know.* 31 Boston: Heinle & Heinle

Park, G. P. (2010). Investigation into Learning Strategies used by effective and less effective learners in Korea. *Asian social science* 6, 8, 3-13. Retrieved from <http://www.ccsenet.org/journal/index.php/ass/article/view/6839/5354>

Pour-Mohammadi, M., & Abidin, M. J. Z. (2011). Attitudes towards teaching and learning test-taking strategies for reading comprehension tests: The case. *Journal of studies in education*, 1. Retrieved from [https://scholar.google.com.mx/scholar?q=Attitudes+towards+teaching+and+learning+test-taking+strategies+for+reading+comprehension+tests:+The+case&hl=en&as\\_sdt=0&as\\_vis=1&oi=scholar&sa=X&ved=0ahUKEwimlcmWqq7TAhUQ22MKHdVsAWkQgQMILjAA](https://scholar.google.com.mx/scholar?q=Attitudes+towards+teaching+and+learning+test-taking+strategies+for+reading+comprehension+tests:+The+case&hl=en&as_sdt=0&as_vis=1&oi=scholar&sa=X&ved=0ahUKEwimlcmWqq7TAhUQ22MKHdVsAWkQgQMILjAA)

Rogers, W. T., & Bateson, D. J. (1991). Verification of a model of test-taking behavior of high school seniors. *The Journal of Experimental Education*, 59, 4, 331-350. Retrieved from <http://www.jstor.org/stable/20152301>.

Rogers, W. T., & Yang, P. (1996). Test-Wisness: Its Nature and Application. *European journal of psychological assessment*, 12, 3, 247-259. Retrieved from [https://www.researchgate.net/publication/232570706\\_Test-Wisness\\_Its\\_Nature\\_and\\_Application](https://www.researchgate.net/publication/232570706_Test-Wisness_Its_Nature_and_Application)



Salehi, M. (2011). Test Taking Strategies Implications for Test validation. *Journal of language teaching and research*, 2, 4, 850-858. Retrieved from [https://www.researchgate.net/publication/267919907\\_Test\\_Taking\\_Strategies\\_Implications\\_for\\_Test\\_Validation](https://www.researchgate.net/publication/267919907_Test_Taking_Strategies_Implications_for_Test_Validation)

Sarnacki, R. (1979). An Examination of Test-Wiseness in the Cognitive Test Domain. *Review of Educational Research*, 49(2), 252-279. Retrieved from <http://www.jstor.org/stable/1169961>

## APPENDIX A: Final questionnaire

Nombre: \_\_\_\_\_ Edad: \_\_\_\_\_  
Semestre: \_\_\_\_\_ Veces que has tomado el examen TOEFL: \_\_\_\_\_

### Estimado participante:

Gracias de antemano por tu contribución a esta investigación. Enseguida aparece un cuestionario sobre las estrategias que utilizaste durante el tiempo en que respondiste la sección de Reading en el examen TOEFL. Por favor, lee cuidadosamente las preguntas y selecciona una de las siguientes 5 opciones: 1 (nunca), 2 (generalmente no), 3 (a veces), 4 (generalmente sí), 5 (siempre).

1. Cuando empecé a contestar el examen, planeé como contestarlo y me apegue al plan.  
1      2      3      4      5
2. Dividí mi tiempo para cada lectura.      1      2      3      4      5
3. Me salté las instrucciones y fui directo al examen.  
1      2      3      4      5
4. Leí cuidadosamente las instrucciones para asegurarme qué tenía que hacer y como lo tenía que hacer.      1      2      3      4      5
5. En cada texto determiné que preguntas eran más difíciles y las respondí antes que las fáciles.      1      2      3      4      5
6. En cada texto determiné que preguntas eran más fáciles y las respondí antes que las difíciles.      1      2      3      4      5
7. Contesté las preguntas en el orden en el que aparecían en el texto y deje las preguntas generales para el final.      1      2      3      4      5
8. Antes de leer las preguntas leí todo el texto.      1      2      3      4      5
9. Leí la pregunta, consideré las opciones y después fui al texto/párrafo.  
1      2      3      4      5
10. Leí la pregunta pero no consideré las opciones, y después fui al texto/párrafo.      1      2      3      4      5
11. Leí la pregunta y luego el texto para encontrar pistas de la respuesta.  
1      2      3      4      5
12. Parafraseé la pregunta.      1      2      3      4      5
13. Me esforcé en entender el significado de la pregunta.  
1      2      3      4      5
14. Predije o di mi propia respuesta después de leer la porción del texto a la que se refería la pregunta.      1      2      3      4      5
15. Predije o di mi propia respuesta después de leer la pregunta y las opciones (antes de regresar al texto).      1      2      3      4      5
16. Leí las preguntas y/o las respuestas varias veces para entenderlas mejor.  
1      2      3      4      5

17. Consideré las opciones y elegí una opción porque tenía una palabra desconocida.
- 1 2 3 4 5
18. Consideré las opciones y me enfoqué en una opción conocida/familiar.
- 1 2 3 4 5
19. Consideré las opciones y pospuse mi decisión para releer el texto y asegurar la respuesta correcta.
- 1 2 3 4 5
20. Seleccioné una opción y analicé las demás para asegurarme que elegí la correcta.
- 1 2 3 4 5
21. Tan pronto como encontré una opción dejé de analizar las demás opciones.
- 1 2 3 4 5
22. Vi el vocabulario de la opción y lo localice en el texto.
- 1 2 3 4 5
23. Si no entendía las opciones o tenían vocabulario que no conocía, no las seleccionaba.
- 1 2 3 4 5
24. Si entendía las opciones o tenían vocabulario que conocía, entonces las seleccionaba.
- 1 2 3 4 5
25. Seleccioné opciones eliminando otras opciones que no tienen sentido basándome en el contenido del texto.
- 1 2 3 4 5
26. Use mi conocimiento previo para entender el texto.
- 1 2 3 4 5
27. Deje las lecturas difíciles para el final.
- 1 2 3 4 5
28. Contesté las lecturas difíciles primero.
- 1 2 3 4 5
29. Contesté las lecturas en el orden que aparecían.
- 1 2 3 4 5
30. Traduje el texto/las preguntas/respuestas al español.
- 1 2 3 4 5
31. Use mi conocimiento previo para seleccionar la respuesta correcta. Sin ver el texto.
- 1 2 3 4 5
32. Use mi conocimiento previo para seleccionar la respuesta correcta. Por que no la encontré en el texto.
- 1 2 3 4 5
33. Respondí algunas preguntas difíciles al azar (adivinando) con la esperanza de que fuera la correcta.
- 1 2 3 4 5
34. Respondí algunas preguntas difíciles haciendo aserciones académicamente (traté de adivinar usando mi conocimiento previo o eliminando las respuestas menos lógicas de lo que entendí del texto).
- 1 2 3 4 5
35. Hice pequeñas notas o subraye las ideas principales durante el examen.
- 1 2 3 4 5
36. Monitoreaba mi propio progreso para terminar el examen a tiempo.
- 1 2 3 4 5
37. Las respuestas en las que tuve duda, las marque para revisarlas posteriormente.
- 1 2 3 4 5
38. Examiné las respuestas cuidadosamente antes de entregar el examen.
- 1 2 3 4 5
39. Me sobró tiempo al final y lo utilicé para analizar algunas de mis respuestas.

1      2      3      4      5  
40. El estar familiarizado con el formato del examen me ayudó a responder más rápido.

1      2      3      4      5  
41. Escribe otra(s) estrategia(s) que hayas utilizado al responder la sección de Reading cuando tomaste el examen TOEFL.

---

---

---

Gracias por tus respuestas. Son muy valiosas.

Adaptado de Barati & Kashkouli (2005) y Cohen & Upton (2006).