BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE LENGUAS

Evaluating the reading abilities in the CELE departmental English exams against those assessed in the TOEFL

A THESIS SUBMITTED TO THE FACULTY OF LANGUAGES

FOR THE DEGREE OF

MAESTRÍA EN LA ENSEÑANZA DEL INGLÉS

By:

Rosalba Leticia Olguín Díaz

Dr. Michael Witten Ulrich

Thesis Director

PUEBLA, PUE. **NOVEMBER, 2014**

Evaluating the reading abilities in the CELE departmental English exams against those assessed in the TOEFL

This thesis has been read by the members of the thesis committee of

Rosalba Leticia Olguín Díaz

And it is considered worthy of approval in partial fulfillment of the

requirement for the degree of

MAESTRÍA EN LA ENSEÑANZA DEL INGLÉS

_____
Dr. Michael Witten Ulrich
Thesis Director

_____	_____	_____
Dra. Teresa A. Castineira Benítez	Dra. Verónica Sánchez Hernández	Mtra. Fátima Encinas Prudencio
Committee Member	Committee Member	Committee Member

ABSTRACT

CELE English students are struggling to reach a specific TOEFL score to graduate and exit from the university. They require the appropriate reading skills and abilities to become efficient readers. Most of these CELE students are detected until having undergone a previous TOEFL examination. They, most of the time, happen to not reach a satisfactory score. Factors such as the timing, the duration or the format of the test can be blamed for. However, this failure is to be characterized by students' lack of the pertinent reading abilities which in turn can be indicative of academic underachievement and non-strategic methods of reading instruction.

The present study intends to investigate the reading abilities being assessed in both exams: CELE departmental English and TOEFL exams and use the results to make an estimation of which abilities are being assessed and which ones need to be added in the CELE examinations in regard to the reading section. This is the key to help those struggling students break this cycle of failure in reading and making the connections with the academic program to help them acquire those particular reading abilities necessary to achieve success in any future TOEFL examination.

To my parents for their endless love and support

To my sister for her expert advice and guidance
I remain your humble student

ACKNOWLEDGEMENTS

I wish to acknowledge my gratitude to the following people. They were essential in the making and completion of this thesis. I am indebted to them for their help that I cannot do justice to all their contributions.

I thank Dr. Michael Witten, my advisor, for his support and encouragement during the demanding process of writing this thesis.

I should also like to thank my readers: Mtra. Fátima Encinas, Dr. Verónica Sánchez and Dr. Teresa Castineira for their valuable observations and always pertinent comments; their feedback helped me enrich this thesis to conclude this journey successfully.

Finally I must thank my sister, Cinthya, for her guidance and confidence along this venture; she is, undoubtedly, the expert in the field of reading, and Yair Romero, for his advice and willingness to read and revise in such short notice.

'What the Brain Tells the Eye is More Important than What the Eye Tells the Brain'

(Smith, 2004)

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Context

In this fast- growing world, globalization has made standards develop and spread in such a significant mode that they seem imperative for almost everything. Languages are no exception and specific standards have also been developed over the past years to get languages measured, graded and certified. Today's individuals are aware of this global environment and perceive the need to be prepared for the future in which they will have to face all kind of different situations including contacts with other languages and cultures, either in their private lives, school or in their job. The need to be capable of establishing successful communication in another language affects those whose interest focuses on doing business abroad or studying overseas.

This has logically brought a consequence: the interest of governments and, then, schools to provide the standards to demonstrate a language command. Therefore, it is not a surprise that languages are increasing their demand in a wide range of work and school settings where either professionals or students of all kinds need the ability of a language to obtain a job, to get promoted, to enter a university or to just perform effectively in their particular environments. This standardization regarding languages has permeated worldwide and has brought about language tests that can certify the mastery of a language. These language tests are used in education, particularly, as a gateway to either enter or exit school.

The Benemérita Universidad Autónoma de Puebla (BUAP), as most of the majority of educational institutions, requires students to certify a command of any foreign language

in order to graduate and obtain their diploma. BUAP has two Academic Models of study to have students carry out and accomplish their program successfully: FENIX PROGRAM where a foreign language is mandatory and it is included in the program curriculum being later averaged into the students' final Grade Point Average (GPA) or *promedio general*. On the other hand, MINERVA PROGRAM requires students to have a certified foreign language to graduate but it is not included in the curriculum and therefore it doesn't influence their final GPA, but students are required to certify this in order to graduate (known as Requisito de Titulación). In the latter program, the foreign language can be obtained through a valid international test. These tests vary and range in cost, length and availability depending on the language to be tested and the institution that administers them.

Most of BUAP students (MINERVA) opt to study and get their certification in English and BUAP approves a range of international tests where students can get this language certification. Some of these certified tests are issued by Trinity College London, Cambridge Exams, International English Language Testing System (IELTS), English Language Assessment System for Hispanics (ELASH) and Test of English as a Foreign Language (TOEFL). All the above examinations are developed to measure the command of English as a foreign language, the first three; in particular, conform to the Common European Framework of Reference for Languages (CEFR), while ELASH and TOEFL do not follow these regulations. The difference lays on the fact that these last two examinations are administered by the College Board and the Educational Testing Service (ETS) whose offices are in Puerto Rico and the United States of America correspondingly. However, these last two certifications can be translated into CEFR equivalencies as a matter of standardization. Despite no single test can reflect accurately everything a student

knows or is able to do in a given language. Still, these certification exams are well-respected and widely seen as fair and authentic.

MINERVA students have the option to be trained whether at a private institution or inside the university's Centro de Lenguas (CELE), a branch from the Facultad de Lenguas, this last one offers courses at a reasonable cost and with a wide range of different language courses to meet the demand of BUAP students in general but particularly of those in the MINERVA program. CELE's English language programs are designed to assess students under the premise that they will meet the appropriate requirements stated in the Common European Framework of Reference for Languages CEFR; due to the fact that BUAP (MINERVA) students must certify a level A2 as minimum to graduate and obtain their bachelor's diploma or *título*. Nevertheless, a great deal of students chooses to take either ELASH or TOEFL tests rather than the examinations structured in the CEFR because of their low fees and accessibility.

Both ELASH and TOEFL have established dates during the year, five for ELASH and twelve for TOEFL inside BUAP, administered by DEPEA- Departamento de Planeación y Evaluación Académica or students can present them at any other institution; consequently, TOEFL is the most demanded and perhaps the best known in terms of its format, cost and length. Although TOEFL (in any of its modes: paper based- PBT, Institutional Testing Program –IPT or Internet based- IBT) does not meet the CEFR requirements, and has an expiration date of two years after its issuing, BUAP accepts it as a valid test under a specific score (ITP 483 and IBT 57) as a equivalent to an A2 level.

1.2 Problem

CELE as a foreign language training center offers nine levels of several languages, being English the most demanded. After each level, students are required to present a departmental test to validate their passing to the next course. Departmental exams are designed and prepared by a language exam commission (Comisión de Exámenes) inside CELE. These departmental exams, no matter the language they are assessing, have the purpose to evaluate a set of abilities determined by each level as a means to reflect students' language progress.

Unfortunately and even when the departmental exams at CELE are preparing students to pass satisfactorily any communicative examination (those outlined in the CEFR); the reality points out that they fail to prepare and train students adequately for a TOEFL examination; principally in the area of reading skills required by TOEFL. CELE syllabus is not adjusted to this aspect and students' TOEFL scores tend to be low which brings about delays in the process of graduating future professionals and the internal issue whether CELE is really accomplishing its goal as a language trainer which is leading to questioning it in terms of prestige as a University language center.

1.3 Justification

Today, universities make great efforts to have students graduate with adequate knowledge and tools to be prepared for professional development. In this way, universities offer language courses, which hopefully lead students to finish their majors with a better preparation in the professional field they plan to major in. These language courses are part of their academic preparation as a further tool in their professional development. Therefore language teaching aims to suit people needs for language learning. Thus, English becomes

a means to achieve something else, either for occupational or academic purposes.

Considering that CELE programs represent the guide to the design of the departmental exams and that these are the only institutional means to assess language performance; it becomes of great importance to synchronize what the language programs state in terms of format and abilities with the departmental exams. If these two aspects consider that students are most likely to take a TOEFL examination then the adequate merge result will bring about departmental exams not only under a communicative approach but also that would be assessing the reading abilities under a format of a TOEFL examination that students will eventually carry out afterwards. CELE language programs aim to instruct students to be communicatively successful no matter the ability but the examinations that assess these skills as a whole fail to be available to the future graduates in terms of money and time, especially Cambridge or Trinity. As a result, it is of prime importance to adjust the program to the examination most likely to be used, in this case TOEFL and revise the skills, foremost in the reading area, to be evaluated so they are included and tested appropriately in the departmental exams at CELE.

The prime justification for this project is the evidence that a considerable number of BUAP students have been struggling to graduate not because of a lack of knowledge related to their language learning but for the primary reason of not developing the adequate reading abilities during their attending to a language course, in this case CELE, to perform successfully in the examination established by the BUAP to grant their diploma. The literacy ability in a foreign language, besides the academic achievement, to exit university continues to be an issue concerning teachers, students and the BUAP administration.

There are two major factors underpinning the significance of this study. The first is that reading problems in a foreign language in university students are reflected in their TOEFL scores and future exit of their university program. The second is that CELE departmental exams are lacking the format and the assessment of those abilities examined in TOEFL tests which without no doubt affect students, in first place, and then the institution as a whole due to the rates of graduate students per year or cohort.

1.4 The Present Study

The present study sets out to investigate which reading abilities are really considered and assessed in two different examinations: the local CELE departmental English exams (level 4) at BUAP and the TOEFL ones. The resulting reading abilities in both exams will lead to, first; have a comparison between them and then a consideration in terms of redesigning the layout of the former ones to match the requirements of the latter just in regard to the reading section they both assess.

The data to be analyzed consists of six CELE departmental English exams (there are three exams per academic year covering spring, summer and fall) from the years 2012 and 2013. The level to go under study is level four because in terms of the nine levels CELE offers, this level goes somehow in the middle of the process of language learning and it works as a gauge to measure students' progress thus far. Also, in light of the fact that level five had previously undergone piloting to change textbook and therefore its exams and programs it was not considered for this study because of these ongoing changes. The six departmental exams will be broken down so that only the reading section will be considered and further examples will be used to have a better picture of the analysis.

These exams were retrieved from DEPEA prior request for this study and since the electronic data specific to these years was not available they were handed out in a hard copy. As for the TOEFL exams, the examples will be retrieved and examined from those contained in the Heinemann's TOEFL course book (available at any bookstore) and that is widely used among language institutions along other publishers. The choice of this particular TOEFL book is that the LEMO, apart from CELE, also offers other courses, aiming to the general public and the university community, that can range from general and conversational English to Cambridge and TOEFL preparation courses.

As indicated earlier in this chapter, this study addresses the next research questions:

- ✓ Are the reading abilities in CELE departmental exams of English (level 4) assessed in accordance with the reading abilities in TOEFL?
- ✓ Which reading abilities in CELE departmental exams of English (level 4) are being assessed?
- ✓ Which reading abilities in TOEFL are being assessed?
- ✓ Which abilities assessed in TOEFL are missing in the CELE departmental exams of English (level 4), if any?

The first research question addresses the reading abilities assessed in both CELE departmental exams of English (level 4) and TOEFL. Clearly this means that the reading abilities needed to be identified through the use of selected techniques. In this study, it was considered important to analyze the reading section in the CELE departmental exams with the use of the pertinent taxonomy to measure which abilities these exams' reading sections were assessing or testing as stated in the second question. Consequently the same was done with the TOEFL reading section as outlined in the third question of

this study that had to do with the reading abilities assessed in TOEFL. This approach enabled the analysis of the reading abilities under assessment in both exams. Finally, as described in the fourth question, in order to gain insight into which abilities matched or diverged in both exams the necessary comparison between the resulting abilities was carried out to finally proceed to the appropriate diagnostic and conclusions.

The research questions arose from a number of issues which are highlighted in the literature review. Firstly, it appears that CELE students show to be having reading problems in regard to the Departmental English exam (level 4) they present after they finish their course. Secondly, for most of these students the reading abilities they do not develop over their CELE course affect their future performance in a TOEFL examination. Thirdly, many struggling readers (either in Spanish or English) are not detected until they come to face the fact that to graduate they need to obtain a determine TOEFL score to exit university. In some cases such individuals have been unable to cope with the huge amount of reading that TOEFL requires as they come to realize their inability to successfully pass it. Fourthly, for students who have struggled with reading up to level four, little improvement would occur as they progress into the next level since the CELE departmental English exams do not promote or help in their adequate reading ability development.

1.5 Summary

In summary, the present study was intended to investigate the reading abilities being assessed in both exams: CELE departmental English and TOEFL exams and use the results to make an estimation of which abilities are being assessed and which ones need to be added in the CELE examinations in regard to the reading

section. The focus was firstly to investigate which reading abilities both exams assessed and then compare them to have a list of those abilities being assessed against the ones which were not considered. The second focus was to investigate the characteristics in both format and abilities to have a valid match for future CELE departmental English exams and contribute to a change that will eventually benefit students who wish to graduate and obtain the required score in a TOEFL examination. Chapter two of this thesis addresses the review of relevant literature, while the research method is covered in Chapter three. Chapter four is devoted to data analysis and results to finally outline in Chapter five the conclusion and implications for the present study.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

Reading is a complex process and undeniably necessary to carry out adequate performance or functioning in the majority of societal communities. Having a successful reading entails an amalgam of processes, starting off by the simplest one: the recognition and identification of letters, then how these letters become sounds to open the gate to words and sentence structures or syntax. However, the ultimate objective at the end of the day for any reader is: to be capable of recalling facts and events outlined in the reading that make sense among one another by activating previous knowledge lodged in his brain to end up remembering the results or consequences from that given reading. At the finish line of this race, the reading success is measured in terms of how this reader uses the newly fresh knowledge to construct new meaning (Alderson, 2000; Grabe, 1991; Solé, 2000; Stanovich, 1980) in his brain for future problem solving situations.

This goal to be accomplished as such, takes on different lower and higher level processes (Alderson, 1990; Grabe, 1991; Grabe & Stoller, 2002; Stanovich, 1980), such as decoding, reasoning, inferences, and even predictions– low and high order processes will be further discussed in the literature review-. A central aim of education, in general – beginning with elementary school- is to provide students with the pertinent reading skills to perform these essential basic processes whenever necessary either to undergo any institutional examination, as to assist readers when exercising reading for other purposes rather than those only connected to school like contexts.

This paper, indirectly but yet related, brings into question an existing problem in the system of education: individuals trapped in a situation of reading failure. This problem does not only permeate the local context but worldwide. Although the efforts in different places around the world to lessen this situation, what remains true is that because of the nature of reading and the administration of education, the number of individuals with poor reading skills increases despite those efforts. If the process of reading is really understood in conjunction with the different ways to improve it, there would be fewer individuals facing difficulties, like being increasingly challenged academically, owing to their weak reading abilities (Snow, 2002).

Along this statement is also relevant to mention that not enough attention or help has been devoted to teachers to promote and encourage reading comprehension as a way to close the everyday wider gap between those with appropriate reading skills and those who lack them (Snow, 2002). Snow (2002), as well, elaborates on discussing how this also brings into questions government policies and programs regarding education, reading and literacy whose foundations are most of the time brought from the outside and adapted but rarely based on local and empirical information which leads to uncertainty and a resulting inappropriate evaluation. Due to this interest, Snow (2002) also provides a well-articulated model of reading comprehension in the Reading Study Group's report (RAND) where she defines reading comprehension as "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (p. 11). She proposes that reading comprehension includes three elements: "the reader who is doing the comprehending, the text that is to be comprehended, and the activity in which comprehension is embedded.

Langer (1995) stated, back then, in the National Assessment of Educational Progress' report (NAEP); in an effort to improve reading, that literacy enhances active thinkers to use literacy skills to think, and interpret knowledge, the world, and themselves. If students, as active thinkers, engage in their reading they will learn to reflect and develop deeper understanding. Despite the decade between these two reports it is evident that they share the opinion of reading as a way to develop better understanding by knowing when to read, how to read and how to reflect on what has been read, both reports target reading betterment.

In light of this study's objective, the questions that guide this research and the purpose of this thesis, the literature falls into 4 main sections. The first section is concerned with the nature of reading and the reading process in general. The second section revolves around cognitive models that attempt to explain the reading comprehension process. The third section addresses reading skills and strategies and the assessment of the reading ability. The fourth section examines the patterns used to assess reading comprehension in testing. These four areas are of particular importance to this study. The evidence shows quite clearly that.

2.1 The Nature of Reading

Reading, apart from spoken language as a representation of the written, is one of the many unique human and yet complex cognitive processes that differentiates us from other mammals in the animal world. As it was previously mentioned, success in the reading process translates not only from the correct process of identification of letters and subsequent decoding of those letters in words to sentences and syntax but also the construction of meaning (Solé, 2000). The reading ability in words of Solé (2000) is the

cognitive process through which an individual interacts with a given text to meet the expectations or objectives that this individual has regarding that text. In order to carry out a reading comprehension activity it is imperative to always have an objective in mind. The reading process, defined in a very rough way, takes place when the reader constructs meaning after that text to which Alderson (2000) concurs by stating that reading is, in essence, a purposeful activity.

Moreover, Alderson and Banerjee (2002) add to this conception that within this interaction – text and reader- , the reader's background, e.g. knowledge, affect, language abilities, etc…; comes into the picture along with the text's characteristics (topic, genre, vocabulary, syntax, etc.) to build not only comprehension but also negotiate meaning where the reader also express both a social and individual identity. Therefore, reading becomes a multi-component process with many different layers that overlap as part of a whole. It starts from the decoding of written words or patterns into their linguistic form, to access meanings stemming from those words within phrases and sentences (Grabe, 1991). The next step is to analyze these semantically, to finally put together the general meaning of those sentences and paragraphs to make sense of a text and own it.

Following this same idea, Carrell (1984) and Anderson and Pearson (1984) consider that a text does not carry any meaning per se until encounters a reader. Then, the text guides the reader on how to retrieve previous knowledge to construct meaning; this is called background knowledge or schemata. Schemata theory insists that reading is a selective and interactive process that takes place between a reader and a text where various aspects become part of this process such as prior knowledge and language knowledge whose interaction with the information contained in the text contribute to comprehension. Schemata theory also adds that reading for a purpose provides a motivation to get engaged

in the process of reading. Noting that any reading encounter means having a purpose, it is relevant to mention which are the most common purposes we engage in to carry out a text.

Grabe and Stoller (2002) outline seven purposes common to reading which range from searching for information and skimming to critiquing and writing. They also argue that reading for comprehension is the primary and most basic purpose that embeds the other six ones being the most complex and complicated to treat. Grabe and Stoller (2002) also go beyond by defining what reading skills and strategies are by noting the former ones as abilities acquired gradually through time and practice as a result of a task making their application become automatic while the latter ones are controlled abilities carried out consciously and intentionally and can be open to exercise reflection upon them. Grabe and Stoller (2002) also rush to clarify that these distinctions do not have necessarily a clear cut boundary due to the nature of reading as a cognitive process per se rather than by a definitional issue.

Drawing on the idea of reading comprehension, Smith (2004) moves further this idea of reading conformed by skills and strategies, by establishing that reading comprehension is a state rather than a kit of skills or processes. Smith (2004) understands comprehension as the relations we make according to the world around us including the intentions and expectations we have related to a particular phenomenon. According to Smith's perspective (2004), reading comprehension is the way we relate the old with the new, how we make sense of the new information and how we accommodate that new piece in our previous knowledge. Smith (2004) affirms that prediction is 'the core of reading' (p. 25) and that this promotes our comprehension when retrieving meaning out of a text. Here, prediction is understood as the prior knowledge we already possess about the world that

helps us disregard among a group of options to choose the correct one so predicting before and during a text is not done carelessly but rather under the premise of previous knowledge.

Rupp, Ferne, and Choi (2006) agree with the notion of reading as a very complex and yet essential process to be successful in any future purposeful reading situation. They believe that any successful reading takes on the fluid, accurate and efficient application of bottom-up processes (p. 444). This view is supported by Carver's model (1977) and Stanovich's concept (1986) on reading comprehension where both highlight the importance of phonological awareness and word recognition as primary to future successful reading comprehension.

Carver's (1977) Rauding theory or model implies that any 'typical' reading (p. 13) done apart from studying is stable in terms of comprehension. The occurring language is understood as it is presented without difficulty. The Matthew effect in reading is Stanovich's conception (1986) where it is described that those who normally read a lot and will keep on doing so will, therefore, eventually develop bulky vocabulary and broader knowledge. The Matthew effect or "rich-get-richer" effect (p. 381) has a counterpart to this statement where poor readers will read less and as a consequence lower their vocabulary and the growth of reading ability or "the poor-get-poorer" effect (p.382).

Rupp et al. (2006) and Stanovich (1986) as well refer to Perfetti´s theory as a third conception: Perfetti's 'verbal efficiency theory' (1985) adds that reading is not just a mere mechanics of decoding words but how its deciphering leads to an understanding of a text and its future coherent construction. This theory suggests that the reader needs to pay attention to comprehend rather than just to recognize the words. Rupp et al (2006) suggest that this level of comprehension of a text is what can postulate an individual to be considered a proficient reader. A proficient reader can, then, carry out three key processes

or levels for a successful reading task: lexical access that encloses the identification of words to their phonological representation; propositional encoding that involves the recognition of words and their meanings within a context and finally text modeling which incorporates the propositions (meanings) into a coherent and logical mental illustration of the text.

Twenty-five years later, Perfetti (2010) publishes an enhanced theory that takes on its predecessor – verbal efficiency theory-. This updated theory is named the *Golden Triangle of Reading Skill*, in Perfetti's words this new additional concept differs from the former one in the sense that knowledge was not given the prime importance back then. Perfetti (2010) argues the verbal efficiency theory emphasized in the general process of reading that entailed decoding, phonological processes, retrieval or memory but little attention was paid to knowledge. Perfetti (1985) stated in this verbal theory that a skilled reader carried out efficiently the operating mechanisms mentioned above but efficiency was not considered in terms of measurement to make it plausible to differentiate whether it was present or not, and to what level.

On the other hand, the triangle or 'Decoding, Vocabulary and Comprehension' theory (DVC) includes all the notions contained in the *verbal efficiency theory* but a new element is added in a wider scope, knowledge. Here, knowledge is in regard to lexical representations plus practice that inevitably leads to a more refine and efficient reading process. The DVC triangle has the purpose to illustrate the interdependence of knowledge about words, meanings and comprehension. To sum up this framework, Perfetti (2010) outlines it by noting that once an individual has mastered decoding, the reading lies upon a series of complex honed skills through effective reading experiences. These experiences consequently produce comprehension and along this comprehension the knowledge of

words and meanings strengthens so that these provide the necessary practice for the enhancement of the reading skill.

The above notions have commonalities as they all go by scaffolds where they all start out by mentioning decoding, meaning and then comprehension. Luke (1992) draws attention not only to these previous components but enrich them by categorizing them into four roles that an efficient reader gets involved in, by citing his previous work in Luke and Freebody (1990), Luke (1992) recalls this reading model developed along with Freebody where both attempt to break the reading process down into four aspects or roles. These four roles are accordingly: code-breaker, text-participant, text-user and text-analyst. Both authors argue that effective readers first go by breaking a code meaning that a reader pays attention to letter-sound correspondences, complete words, sentences, paragraphs, semantic information, punctuation and word meaning.

As a text-participant the reader is involved in what that text is expressing within the context of the same reader's background knowledge about that topic. When turning into a text-user, the reader grasps and uses the text in order to achieve a particular social purpose. And the text-analyst practices refer to how readers critique and analyze the primary and unsaid assumptions or beliefs in a text, - to read between lines some may say, and the way a text attempts to position a person as an active reader. Luke and Freebody (1990) not only go beyond the notion of the already discussed interaction between text and reader but also enhance and empower the reader as the conductor of this interaction by assigning active roles along the whole process of reading.

There has been previously mentioned the term efficiency when it comes to evaluate the reading ability or process, to what, Pressley (1998) agrees, and suggests another term to this idea of efficient readers. Pressley (1998) talks about skilled reading as a way to become

competent readers, by adding that in order to increase reading decoding and comprehension, children need to learn two things: word identification strategies and comprehension strategies. These strategies are interdependent, with comprehension being closely linked with a reader's quick and efficient identification of the words in a text. It is obvious that it there is a rapid word identification this will definitely facilitate and sustain understanding, which in turn, will lead to speed and fluency in processing any written text.

The bottom line of this discussion about reading as a process is that is undeniable to the majority of researchers that the heart of reading is meaning; since meaning is what we search for as we read (the purpose) and it is also part of what we use to reach that specific purpose. In making meaning from a text, readers combine their knowledge of the world, the topic of the text and their knowledge of the language (Solé, 2000). Pressley (1998) emphasizes that higher-order processes are essential for achieving this purpose or objective; for example, comprehension or making inferences. These higher-order processes help develop the ideas in the text through the channel of low order processes, such as decoding, in the same text. Retrieving meaning from a text depends on highly efficient low order processing to what Pressley (1998) adds: "good readers are interpretive, evaluative and reactive to the text" (p. 62).

2.2 Models of Reading Comprehension

It has been discussed along previous pages the different notions about the reading process and how they have struggled to define it and also outlined the different array of micro processes within the macro and complex one. The different theories and concepts share similarities although certain minimal distinctions but what it is undeniable is that they all offer a very wide picture of what reading is including all its intricate technicalities.

Following the above idea regarding reading technicalities, it is also very important to mention in this paper the different notions that intent to explain how reading comprehension happens. In other words, the definition and the references to reading are rich but how exactly does the brain process a text or how does a reader approach a text and figures the meaning out of it? This opens the discussion about processes and mechanisms to understand a text but most specifically the Bottom-up, the Top-down and the interactive models. These will be reviewed briefly in the following paragraphs.

2.2.1 Bottom – Up, Top – Down and Interactive Models

Researchers agree that reading is all about constructing meaning, but the question that arises after this principle is how exactly this construction of meaning happens. It has been reviewed that the beginning point to approach any text is by understanding the code (language) first and then the recognition of words and the meanings attached to these to take on the meaning within the context and then the interpretation of what that text wanted to 'say'. But some scholars have attempted to explain how comprehension takes place by creating models of understanding for this process. (Alderson, 2000; Anderson & Pearson, 1984; Gabe & Stoller, 2002; Goodman, 1967; Pressley, 1998; Solé, 2000) There are three major frameworks or models that present different views in an attempt to explain how reading comprehension generates or takes place.

2.2.1.1 Bottom- Up Model

The first one is the hierarchical ascending model: bottom-up. This model begins from the start; this means that it claims that reading comprehension is reached by following the basic and logical sequence in any text. First the words, then the sentences, and later the

paragraphs as a means to guide the individual to comprehend the text; therefore it relies strongly on the decoding of the language. For the bottom- up approach the individual understands a text only if the decoding skills are well honed, it is text centered but fails to explain how typos can be skipped and still do not affect understanding or how it is not necessary to decode accurately every word and yet make sense of what is being read. Also, it cannot explain how inferences are made along the text before moving on to the next word or phrase (Solé, 2000).

It is highly mechanical in its nature and works under a mental translation of one piece at a time mode (Gabe & Stoller, 2002). This idea of reading is very simplistic (Pressley, 1998) by just reducing it to text decoding as a way to retrieve meaning. The Bottom-up model depends upon decoding skills to achieve meaning and even though decoding is a corner stone to reach comprehension; the fact that it reduces the process to just a mere translation of the 'code' makes it seem too petty and undermining of the complexities reading comprehension entails and that have already been presented here.

2.2.1.2 Top- Down Model

On the other hand, the Top-down approach is hierarchically descending and is the counterpart of the Bottom-up framework. This model highlights, according to Solé (2000), the importance of previous knowledge and high cognitive levels to anticipate the text content beforehand. Solé (2000) adds that this model bases its mechanisms in the construction of hypotheses and anticipatory stances whose veracity is tested once the text is processed. Grabe and Stoller (2002) explain this model as the one based on the reader´s expectations and goals widely supported by inferences and background knowledge. There is evidence that supports that the same text will be interpreted differently among readers.

The meaning intended in the text by the author will differ from one reader to another in its interpretation, one of the reasons is the reader's prior knowledge but at the end both will grasp the essence of the same text (Anderson & Pearson, 1984).

Goodman (1967) underlines reading, in this model, as a `guessing game´ where the reader guesses or predicts meaning solely based on minimal textual information by activating to the maximum the already existing knowledge. In regard to knowledge, Smith (2004) aggregates that this *prior* knowledge plus "non visual information" (p. 13) are mental resources at hand that facilitate visual information (print text) to be interpreted. To Smith (2004), the term *cognitive* has to do with knowledge and structure since these involve organization, thus a reader already possesses, what he calls, an organization of knowledge at the moment of venturing in a reading experience.

Despite the heavy weight placed on the cognitive aspect in this framework; the truth is that the reader does not allow this previous or prior knowledge processing to get out of hand at the moment of reading (Pressley, 1998) which means the reader controls and gauges the reading process. The Top-down view stresses the potential exploit and interaction of all processes of both higher- and lower order but as its counterpart, it inclines towards a rather general recognition of words to favor comprehension disregarding the acute decoding skills which are essential for a more efficient and accurate reading comprehension.

Both opposing views somehow strive to demonstrate how meaning is constructed through a text. But because of this contradiction, they both fail to explain certain features that the other model supports, and these gaps in their theory makes it complicated to lean towards one or the other. They both outline valuable data in regard to reading comprehension but leave out salient features of the process to defend just a particular

aspect. The bottom-up model places the reader as a passive decoder while the top-down view considers the reader as an active encoder (Alderson, 2000). Finally, after reviewing both models, the most sensible thing to consider is to meddle between both and find an intermediate attitude since neither one of them really offers an adequate characterization of the reading process as such (Alderson, 2000) owing to the fact they both stand at one end of the spectrum of reading comprehension.

2.2.1.3 Interactive Model

The debate between which model satisfies reading comprehension at its best has led to a simple and yet obvious idea: to combine the best of two worlds in favor of a more complete and fulfilling perspective regarding comprehension. Alderson (2000) mentions an interactive model as one where the different reading components interact regardless of their nature either 'higher up' or 'lower down' (p. 18) and Grabe (1991) deepens on this notion by adding that processing happens to be parallel rather than serial, this concept or parallelism enriches the idea of a transversal nature in reading processing. Both, Grabe and Stoller (2002) and Solé (2000) continue in the same vein and talk more about the option of an interactive model which is the combination of the bottom-up and the top-down approaches, but not a hybrid (Grabe & Stoller, 2002). In this model the reader carries out the bottom-up features such as decoding but at the same time top-down aspects are taking place like those related to predictions and expectations.

There is a simultaneous interaction (Solé, 2000) between the two models where the reader uses background knowledge to activate interpretation and construct new meaning from that text. Solé (2000) recaps her position in support of an interactive model of reading comprehension by noting that the reader is an active agent in reference to the text and that

the reading process is constantly producing and probing hypotheses in an effort to construct comprehension from the text. Urquhart (1987) suggests flexibility in terms of text interpretation because this depends on the reader and the purpose as there are many goals and interpretations that can yield from the same text.

This cognitive reading assertion was later expanded by Iser (1980) when he claimed that in the reading process, readers are the focal point because it is the reader and his prior knowledge that work together with the text. In a brief thought, the interactive model for reading refers to the dynamic relationship between bottom-up and top-down, between decoding and interpretation, and between text and reader; a balance that Alderson (2000) adds may vary with the text, purpose and reader. Meanings eventually trigger from this interaction as there is a dynamic relationship between reader and text, in which reading becomes a creative process rather than just a sheer decoding stance. The reader is therefore assigned a proactive leading role.

To this point, so far, there have been developed concepts such as what reading is and the different notions to explain such a complex cognitive process. Also, three models have been presented as a means to explicate how reading comprehension generates and the different arguments to support such models. Reading comprehension is a compound of different features that range from decoding to the construction of meaning as the ultimate goal. Nevertheless, there are still some underlying processes within reading that need to be explored and presented to better understand how they activate and generate comprehension.

2.2.2 Reading: single indivisible process or multi-componential process

Along with the discussion of the three processing models mentioned above, there has long been considerable disagreement about whether reading comprehension should be

viewed as a single process or a multi component process with separate element skills. For example, Rost (1993 in Carver, 1992 and in Weir & Porter, 1994) states that reading comprehension is a unitary process because the subskills of reading comprehension are so closely fused with each other in the process of reading that it becomes virtually impossible to distinguish among reading subskills, and measure them separately. Rost (1993 in Weir & Porter, 1994) also concludes that the variation in reading could be attributed to a single aspect: "general reading comprehension."

Even though the exact nature of reading comprehension remains divisive, many researchers in the field of language testing support the idea of viewing reading comprehension in terms of separate components. For example a number of researchers support the idea that reading comprehension is multi-divisible and can be separated into two or more components. These researchers claim that several factors can be identified and separated which influence readers' ability to successfully read and comprehend a text, such as vocabulary (Carver, 1992; Urquhart & Weir, 1998) and inferences (Enright, Grabe, Koda, Mosenthal, Mulcahy-Ernt & Schedl, 2000).

Weir and Porter (1994) argue that there are different skill components in reading just as the convention of having skills of listening, reading, writing, and speaking does exist. A number of researchers (Alderson, 1990; Grabe, 1991; Koda, 2005; Weir, Hughes & Porter, 1990) agree that reading is composed of multi-componential skills but differ on the number and scope. Grabe (1991, 2009) and Grabe and Stoller (2002) accept and comment on important aspects such as automaticity in identifying words along with syntactic knowledge, content and background knowledge as well as skimming, and scanning. Their standpoint in relation to a multi- divisible reading process can be directly quoted from Grabe (1991, p. 382) and it reads: "a 'reading components' perspective is appropriate in an

appropriate research direction to the extent that such an approach leads to important insights into the reading process, it…is indeed a useful approach".

Moreover, Alderson (2000) considers that implementing a unitary approach in the testing of reading may not fully characterize the reading comprehension construct and as a consequence it could fall short in appropriately assessing all the related reading skills. Likewise, Koda (2005) points out that reading is a complex construct influenced by a range of variables and involves a number of vast component operations. In the process of reading, every single operation depends on a broad selection of competencies. And Koda (2005) finishes by stating that reading is of a multifaceted nature, a complex mental process that cannot be observed directly. After careful considerations stemming from the previous notions in terms of reading processing, a divisible or multi-faceted knowing view appears to be more plausible for assessing reading comprehension, and it will be the view adopted in this study.

It has been established, after so many pieces of research on the matter previously discussed here, that reading; although complex to explain and for the sake of a consensus, is in a general term, the experience of constructing meaning after a written or print text. Models or metaphorical models of reading comprehension, as Grabe and Stoller (2002) call them; have been created as an attempt to create a common understanding about how this process takes place whether in an ascending or descending path or in a perfect amalgamation of the first two.

Parallel to these models, some scholars have strived to explain the process of reading comprehension as either a single entity or as a puzzle composed of many pieces. This has opened discussions over the soundness of such arguments by calling out on cognitive and mental frameworks to validate their views. Therefore and for the sake of

gaining greater understanding of what reading comprehension is and which mechanisms or tools uses to develop, the following paragraphs review the related literature on the notions in regard to the cognitive procedures behind the reading comprehension process.

2.2.3 Lower and Higher Level Processes in Reading

The concept of cognitive is defined by Smith (2004) by the internal structure of knowledge organized inside our brain that permits a model of the world, a personal theory of our understanding of the world. This means that these mental resources (knowledge and structure) that compose this personal theory are the basis to categorize different experiences while making sense of the surroundings. Smith considers this system of categories as the gear foundation of survival and of learning to sort and process experiences according to differences or similarities (p.16).

If taken into consideration what Smith (2004) says about cognitive mechanisms, Perfetti's (2010) triangle or Decoding, Vocabulary and Comprehension (DVC) reinforces the idea of structure and categorization since knowledge is the complementing aspect to a more refined and efficient reading experience. The DVC triangle illustrates the interdependence of knowledge about words and comprehension. Perfetti (2010) states that mastering decoding has reading skills to be honed through varied effective reading experiences. These experiences bring about enhancement of reading comprehension. Thus, cognitive processes require categorization of knowledge to succeed in learning or disregarding any experience. If these above considerations are taken into account, then, it is more than appropriate to mention these cognitive reading processes known by general convention as Lower and Higher order processes. Grabe and Stoller (2002) describe these components but clarify that this distinction has nothing to do with the idea of easy or

difficult or that one is better that the other. To what, Pressley (1998) states that skilled reading is the result of the coordination of both processes.

2.2.3.1 Lower- level Processes

Grabe and Stoller (2002) and Grabe (2009) explain the reading process as the coordination of multiple levels of sub-component processes. Lower-level processes are predominantly linguistic processes, which include recognition of words and retrieval of meanings (i.e. lexical access), analysis of the syntactic structure of the sentences (i.e. syntactic parsing), and the formation of elementary meaning units (i.e. semantic proposition encoding). Lexical access is the immediate recognition of words. Reading is 'a word by word affair' (Pressley, 1998, p.44) whenever the reader attempts to learn what is in the text.

Perfetti (2007) supports that reading comprehension depends on efficient word reading and discusses that Lexical Quality Theory (LQH) is relevant and extensive. LQH implies that efficient decoding of words and meanings grows in relation to literacy and language experience therefore the continuous engagement with reading helps accumulate knowledge about form and meanings in pro of an efficient comprehension. These words and meanings (knowledge) are recorded and stored in our long-term memory to emerge later on when needed (Kintsch, 1988; Kintsch & Mangalath, 2010). Westwood (2001), correspondingly, suggests the need of a systematic word-recognition instruction as a way to promote reading skills in poor readers. Lexical access is automatic once the reader accesses to the meaning and to be proficient requires extensive practice.

Syntactic parsing is the ability to recognize and discriminate words in context so the reader can maneuver words with multiple meanings and determine their position in the text. Urquhart and Weir (1998) defend parsing as crucial to build up meaning by means of the

regulations in the ordering of English. Parsing takes up on intensive exposure to also become automatic along with the lexical access. An extensive vocabulary that is immediately recognized by the eye is a fundamental component of automaticity in reading (Westwood, 2001).

A third basic process is the semantic proposition that is activated automatically, once the first two have developed, and is the act of combining words and meanings into basic clauses and sentences to make sense of the text in question. It works as a network where the three processes connect with one another, developing automaticity of lower level processes will allow adequate resources available for higher level text-processing (Grabe & Stoller, 2002).

The decoding ability is critical to mastering reading comprehension. Fluency in word recognition enables readers to concentrate on the meaning of a text. Without these so-called lower-level reading skills, the higher cognitive skills cannot function (Carrell, 1984). From the previous arguments it can be underline that fluid and automated word recognition abilities appear to be the direct cause for having efficient reading skills and become good readers and not the result of a greater reliance on contextual information (Pressley, 1998). On the other side of the coin, poor readers base their reading entirely in the use of the context to decode, which although useful is limited in terms of the capacity to apply the interactive processes of text comprehension (higher-level) (Westwood, 2001).

For example, readers who struggle to decode individual words frequently have little processing capacity for comprehending a text since they are caught up in the mechanics of word by word print recognition. Nevertheless, to fully comprehend a text requires more than just word recognition skills, phonic knowledge, phonological awareness, phoneme awareness and sight vocabulary. Readers can also use context to

identify words. Semantics and syntax knowledge help readers make educated guesses about the unknown words (Smith, 2004; Stanovich, 1980). There are other processes that take part in this whole process called reading that help the reader comprehend afar the obvious print. There are the higher-lever cognitive processes.

2.2.3.2 High- level Processes

The higher-level processes go beyond the literal; they extend to interpretation, analysis, and synthesis of information. They are of a comprehension building nature as well as of a strategic processing character. They tend to be controlled processes opposite to lower-level which become automatic as a requirement for fluent reading (Alderson, 2000; Grabe & Stoller, 2002; Grabe, 2009). This means that readers adjust their reading practice to the type of text so they can paraphrase precisely. They also relate the new ideas to their background knowledge and offer conclusions drawn upon the text. There is also the ability to consider the author´s purpose or intentions and judge the text (Solé, 2000).They become self conscious of their reading process and adjust their reading strategies either to verify or repair accordingly (Perfetti, Landi & Oakhill, 2005). Skilled readers are active (Pressley, 1998) and base this efficiency in their prediction ability as they read (Smith, 2004)

Perfetti (1985) reviews a number of studies and suggests that whether high or low-ability readers, they are capable of deciphering words within a context much faster than when in isolation. Although it is observed that low ability readers relied more on the context to grasp the meaning than high-ability readers. However, both good and poor readers detect visual prompts the same, what makes the difference is that good readers have better and more efficient word recognition processes that allow them to revise context faster and do their sampling more quickly than poor readers. Both Perfetti (1985) and

Stanovich (1986) argue that because good readers have more efficient word recognition skills it is why they rely less on the context, which works in the opposite direction with poor readers; who have more reliance on context cues.

At the same time, Kletzien (1991, p. 82) mentions that 'comprehenders', use a variety of word recognition strategies such as phonics, sight words, context clues and structural analysis but she points out this effort takes its toll on the cognitive capacity so the comprehension of the author's thoughts is sacrificed. Stanovich (1982) examined a series of studies where the results showed that a deficient decoding ability was not only exhibited in a poor reading skill but also in the listening comprehension. Therefore, poor readers are correspondingly poor listeners. Readers with a low deficit in reading comprehension struggle with the extra demanding processes that require analysis of more complex texts. Pressley (1998) states that students have been told or have discovered that they can read for the gist of the text and can carry on with the task of reading without an entire recognition of words, except if they get totally lost, but he agrees that inefficient word recognition is at the expense of the appropriate meaning analysis of the author's intentions.

When looking at a text, readers use previous information to filter, interpret, organize, reflect upon and establish relationships (Westwood, 2001). According to Perfetti (2001), in constructing meaning readers combine what they know about the world, the topic of the text, the writing system encoding the language of the text, the letters, words, and visual elements (p. 12800). Nonetheless, from an educational perspective, the cognitive process readers use to arrive to a result or product (Alderson, 2000) is of greater interest because it is on the tracks of the process where comprehension failure originates and becomes evident in the product. These are the processes teachers hope to influence by a guided instruction. But how does comprehension take place exactly? As it has been

extensively covered, at each stage during the reading process, the reader adapts to the text to make sense of the information by going back and forth between background knowledge and prediction techniques. The success of the process depends on issues such as the information in the text, the reader's background knowledge and the reader's lower and higher level reading abilities (Alderson, 2000; Pressley, 1998; Solé, 2000; Stanovich, 1980; Villaume & Brabham, 2002).

It has been widely presented in this section that reading comprehension comprises a spectrum of micro tasks or processes that lead to comprehension as the final product. It has also been discussed that these micro processes as such can take place, cognitively speaking, either from bottom to top or from top to bottom. These two views can be blended in a more interactive mode where micro processes contained in both happen simultaneously where boundaries are thought to be flexible but still work in harmony. Reading comprehension is considered successful, efficient and skilled when the reader considers the achievement of the text in terms of adequate grasp of meaning and can move on to the next phase until completion with an interpretation on his own.

2.2.4 Skills and Strategies

It is pertinent at this point of this paper to bring into consideration a more specific aspect to reading comprehension: skills and strategies. This is important since from an educational and administrative perspective reading skills are evaluated according to the product and the strategies are the decisions to arrive to that final product. But before any other aspect, it is important to start off by describing in more depth these two terms.

Researchers like Alderson (2000) state that despite all the literature available there is not a distinguishable clear line to define skills, abilities or strategies. Having said this, he

offers a broad definition where he understands skills as the techniques required to process a text while strategies are the decisions to process that text, in other words, the skills to be deployed to process text. As for, Grabe and Stoller (2002) skills are the linguistic processing abilities, acquired gradually until they work automatically and in different combinations. They are perceived as the final learning outcome in goal oriented tasks; while strategies are considered a set of abilities that are consciously controlled by the reader. Anderson (2000), as Grabe and Stoller (2002), clarify that these distinctions do not have necessarily a clear cut boundary due to the cognitive nature of reading and not by a definitional issue.

Strategies are not just simply compiled in a 'checklist' to be picked out on a whim. In fact the limits among one another are fine, even blurry at times since strategies are flexible and spontaneous (Villaume & Brabham 2002, p. 672). Although comprehension strategies are singled out to be discussed for educational purposes what it can be stated is that they spring back and forth during the reading process. Strategies can have a long lasting effect if, first, they are acquired individually and then reinforced by practice –e.g. real life tasks- until they become automatic (Pressley, 1995; Rhoder, 2002).

Solé (2000) as well defines strategies as the procedures to regulate and plan decisions related to an objective; where the reader selects and evaluates possible changes in terms of that objective. McCormick and Waller (1987, cited in Pressley, El-Dinary, Gaskins, Schuder, Bergman, Almasi & Brown, 1992) provide a definition for strategies by saying that they are techniques to process a text and they are used as a 'personal' response to that text (p. 525).

In addition to these conceptions, it is pertaining to clarify that even if *skill* and *strategy* may overlap according to some researchers in reality they are two significantly

different concepts. Then after the above literature it can be understood that *skills* are processing techniques acquired unconsciously through practice until they become automatic and *strategies* are the moves selected intentionally to accomplish a goal or objective during reading (Paris, Wasik & Turner, 1991). Strategies can grow with time to be automatic and in turn become a skill; the difference is simple, to evoke the procedure is a strategy and when is applied automatically a skill (Alexander & Jetton, 2000).

Grabe and Stoller (2002) mention that abilities are what a reader 'can' do during the reading process. They recap by stating that strategies can be better understood as abilities when the application and use of strategies are open to questioning and reflection. Kletzien and Dreher (2004) define reading ability by referring to students as 'able' to orchestrate strategies according to the text or task and are able to discern which strategy is useful and when to use it (p.56). Perfetti (1985) considers reading ability as the teaming up of comprehension and the linguistic component because the differences in the ability stem, in first place, from the linguistic aspect. Perfetti (1985) will later add knowledge to this definition of reading ability with the DVC triangle (Perfetti, 2010). Carrell (1984) corresponds by claiming that reading ability is the efficient connection between the text and our prior knowledge.

Skills, just as strategies, are also tested to offer somehow a tangible vision of reading comprehension ability. The results of a reading comprehension test differ in function of the reader's comprehension skills and strategies (Kletzien, 1991). Normally, the results are estimated in terms of the efficiency of these skills such as word recognition (decoding), vocabulary (syntax and parsing), poor phonic and phonological knowledge, contextual clues; the management of reading roles such as code-breaker, text-participant, text- user and text-analyst along with higher level skills in reading comprehension that

include interpretation, analysis and synthesis of information (Freebody & Luke, 1990; Grabe, 1991; Stanovich, 1982; Weir & Khalifa, 2008).

2.3 Assessment of Reading Ability: testing, assessment and evaluation

It is unavoidable to mention that society comprises a wide range of varied and different characteristics among its members. As part of a society, there is a common effort to have everyone access to all the diverse events this same societal group offers. This access is subjected to a set of rules and processes necessary to promote fairness in terms of measuring specific aspects, under a valid and reliable standard to have a pertinent decision made accordingly, this is what we call evaluation. Regardless of the above affirmation, what is true is that our daily life is under constant evaluation from which milk to buy at the supermarket (whole, low-fat, soy, skimmed, fat-free, organic, etc…) to the person we choose to spend our life with. We cannot escape either to evaluate or being evaluated and it has become a universal trait of our life within society (McNamara, 2000) to control and scrutinize transition from point A to point B.

Evaluation has become the means to measure a given phenomenon and decide whether it complies with the set standard or not, so it is used as a mechanism to control entry or exit, in other words, a gateway. But before anything we have to set the record straight and state what has been underlined and specified as evaluation, most specifically, in regard to language performance. According to Bachman (1990), evaluation is a methodical system to collect all kinds of bits of data to meet an ultimate goal: to make a fair and correct decision in relation to a specific event. This means that any decision must be carefully documented and be reliable and relevant to the situation being under the microscope. Bachman (1990) also states that evaluation has little to do with testing since

mere testing deals with the construction of tests and the later assignment of a numerical grade or value, no matter if this 'grade' comes from a qualitative or quantitative test, unless that 'grade' brings about a resulting consequence then evaluation is at place as the last step in the process.

In the same vein, Brown and Abeywickrama (2010) clarify the terms: assessment, measurement and evaluation in terms of testing by claiming that assessment entails an estimate or appraisal of a particular trait or attribute while measurement involves the quantification of that particular attribute, either qualitative or quantitative, whereas evaluation is the interpretation of that information to make a decision. On the other hand, Clapham (2000) comments on the generalization applied to the term assessment to confine both testing and informal assessment where the former means the construction of tests while the latter refers to informal methods to assess.

Clapham (2000) describes that this so-called imposed dichotomy is not necessarily existent, to her, both terms aim to the same goal: valid and reliable tests where the intended skills are assessed appropriately. Although these previous terms might appear confusing it is evident that they overlap to be part of an ultimate whole process: evaluation. For example, when a language achievement test is administered to determine progress so the individual can either repeat or move forward into the next level it can be understood that there was an assessment (the test or instrument), the resulting grade or value is the measurement (a scale of 1-10 or a specific rubric) and if this is taken into consideration for a decision (entry or exit the level), it can be said that evaluation was served as such.

For Cohen and Upton (2006), in regard of reading, tests have become of major importance to gain better understanding about both reading strategies and the process of validation. Alderson, Clapham and Wall (1995) and Bachman and Palmer (1996)

understand tests as the instruments to identify learners' areas of strengths and weaknesses in order to help and improve learning. (p. 42). Therefore, there is the need to have the sufficient theory and understanding related to what to assess and how to do it. As Bachman and Palmer (1996) express their position about testing in the next statement: "Unless we can demonstrate that the inferences we make on the basis of language tests are valid, we have no justification for using test scores for making decisions about individuals . . . we must demonstrate that these inferences are appropriate for the decisions we need to make" (p. 95). Bachman (2000) ascertains that the search for appropriateness of validity and fairness in testing are key elements since these define how professional not only a language tester is but also as an applied linguist.

The above authors have mentioned validity as part of the testing process or as Anderson, Bachman, Perkins and Cohen (1991) would state: "validity is the relationship between test performance and the construct, or ability, it is intended to measure" (p. 42). Validity, in words of Hughes (1992), is the inclusion of the aspects intended to be measured in a representative sample. Bachman (1990) extends on this when he says that valid tests are those that express meaningful, appropriate and useful results (p.25) in terms of measuring the abilities in question and nothing else. Validation in testing has as prime and sole purpose to exercise the fairness and the defensibility of the results from any performance assessment (McNamara, 2000).

Messick (1989 in Messick, 1996) defined validity as the evaluative judgment of the degree to which empirical evidence and theoretical rationale support the adequacy and appropriateness of interpretations and actions based on test score or other modes of assessment (p. 9). Henning (1987) offers a similar definition: "Validity in general refers to the appropriateness of a given test or any of its component parts as a measure of what it is

purported to measure" (p. 89). There are other authors like Anderson et al. (1995) that introduce a division in validity between internal and external. Internal validity is the perceived effects and content of a test whereas external validity is the resulting scores compared against the abilities that were intended to measure (for statistical matters).

As a consequence of validity to regulate the appropriateness of the abilities to be assessed there is another term that links to validity: washback. Washback is understood and explained by Hughes (1992) as the "effect of testing and learning" (p. 1) or as the classroom based issues and their impact on students' future language assessment from Brown and Abeywickrama (2010). Messick (1996) also adds that washback can refer to both positive and negative effects (beneficial or harmful) because it can either promote or inhibit learning. This concept permeated the educational and assessment scene that Alderson and Wall (1992) created a Washback Hypothesis to deduce how tests affected or influenced teaching and learning. The most desirable scenario would be to devise tests that are based directly on teaching practices where the relationship between teaching and testing is of 'partnership' (Hughes, 1992. p. 2).

There should be a dynamic correlation between testing and teaching. If testing is beneficial in terms of washback then teachers would think better of it because good teaching practices are tested adequately. However, if teaching is harmful then testing will inevitable show this gap as well and corrections will be needed, here, testing would be exerting a beneficial washback as a guide to what improvements need to be executed. Testing can be used as an indicator of language abilities and can indicate processing errors and the effects in instruction to provide data for corrective actions (Bachman, 1990). Washback is far more complex and thorny than just the mere effect of testing on teaching because of its potential to affect extends beyond individuals (students, teachers) but also

reaching educational institutions, systems and society (Bachman & Palmer, 1996). The power of tests - and washback- lies in their influence to determine life chances for individuals to access school and for teachers and the curriculum to be under scrutiny (McNamara, 2000).

Apart from validity and washback there is a third element in testing, reliability. Reliability has the aim to measure systematically the changes in a test; this means that a test can be consistent across different characteristics regardless of the testing situation (Alderson et al., 1995; Bachman & Palmer, 1996). Tests scores represent the measurement to which extent it is possible to quantify an individual's performance but if this individual scores the same value (under a benchmark of acceptance) in a different situation with the same test, then, it can be predicted that the test is highly reliable (Hughes, 1992; McNamara, 2000). A test is consistent and dependable only if the conditions are consistent across administrations and it is not ambiguous in its directions, rubrics, content and application (Brown & Abeywickrama, 2010).

There is an evident and tense relationship between validity and reliability but it is the tester's job to balance these two out to provide an adequate assessment of any language ability. Nevertheless what it is unquestionable is that a test cannot be valid without reliability or the opposite. Messick (1992), in his progressive matrix, claims that validation is a fusion of meanings, values, interpretation and test use, hence; validation involves both science and the ethics of assessment. At the end of the testing process, what matters the most is that the devised test offers an accurate and fair reference of the learner's ability (Alderson et al., 1995). Despite the breath of investigation material and definitions regarding validity and reliability there is a blurring line no matter how aware one is about these terms. It is obvious that in real life it is impossible to discard inconsistencies but the

pertinent application of both can help minimize potential sources of discrepancies since reliability and validity are both critical qualities to construct tests and interpret test scores (Bachman & Palmer, 1996).

2.4 Taxonomies to Assess Reading Ability

It has been reviewed, along previous sections, aspects that range from reading as general concept and its notions to the models that strive to explain how this phenomenon takes place. It has also been explained how having indicators of the reading ability has permeated testing and evaluation within an educational context in an effort to have a tangible measurement of an individual´s reading ability under valid and reliable conditions. This last section is devoted to present how reading is tested within exams and how a reading item measures a specific reading skill or strategy.

All those of us who have been immersed in an educational environment have experienced the fear and stress of doing a test, a test that will measure our performance under a given rubric or standard. Alderson (2000) implies that reading is seen either as a process or a product (skills tested) but regardless of the approach one favors, the product seems to be much easier to investigate than the process; not without its own particular pitfalls. Assessment has been typically related to exams and tests that someone prepares for other individuals to take and answer. Whatever the presentation of the test is; what it is evident is that it presents a certain pattern: questions. Questions to be answered, most of the time and despite the nature of the question; test questions call for a specific behavior or answer. In testing, assessing reading through questioning is a widespread practice. It is test constructors task to always be sensible to "the level of meaning they believe readers ought

to ‘get out of’…when assessing ‘how well’ they have understood the text in question (Alderson, 2000, p. 9).

Reading tests are structured under a specific format and contain questions that can be presented before, during or after a reading section and their position can demand different comprehension aspects and serve different functions. Questions before a text can function as guides to consider certain elements or to trigger predictions. Questions within the text monitor and alert about relevant aspects to consider during the reading. Questions after the reading normally look for inferences or evaluations about the text. Reading researchers have attempted to strip down reading comprehension in different skills or abilities to identify diverse levels of reading understanding by asking questions after varied reading passages (Alderson & Lukmani, 1989; Kintsch & Van Dijk, 1978; Van Dijk & Kintsch, 1993). Various attempts to understand how reading can be scrutinized in its smallest pieces have yielded taxonomies, lists or hierarchies of skills.

As previously commented, if reading comprehension is defined in terms of the outcomes or results that yield as a product (in testing); so then, it needs to be compared to a checklist of established goals or standards, each designed to calculate the final results. Such objectives or goals are conceived by reading specialists to provide a standard as an effort to regulate the examinations of reading comprehension abilities. Many models of taxonomies of the reading process have been structured as a means to clarify how skills and strategies develop and how they can help in the construction of a test. Different scholars have created particular lists or taxonomies as an effort to outline and identify reading abilities and skills.

For example, Brown and Abeywickrama (2010) talk about seven micro skills and seven macro skills in reading comprehension. This list is conceived in relation to the multiplicity of types of texts where the genre determines the deployment of specific reading

skills. These micro skills range from deciphering graphemes, recognizing plain words and meanings and their connection in the overall sentence to rhetorical conventions, inferences, metaphors and interpretation. It can be observed that the micro skills are those related to a bottom up model while the latter ones, macro skills, to a top down approach.

Meanwhile Carver (1977, 1992) defends the idea of only five processes or abilities regarding reading comprehension. Carver (1977) ascertains that the reading ability starts by decoding words first and then move upwards to meanings and propositions. These three abilities are what Carver (1977, 1992) calls rauding or 'typical' reading (p. 13) which is stable in terms of comprehension. The occurring language is understood as it is presented without difficulty. Rauding has words become meanings to turn into complete thoughts - reading comprehension. Along rauding there are also two more abilities: scanning and skimming. To Carver (1977), rauding was the concentration of a unitary process -in terms of abilities- that worked towards reading comprehension.

As for Grabe and Stoller (2002) and Grabe (2009) they all agree on around twenty two major reading abilities or strategies. These abilities include the ability to summarize, guess vocabulary in context and make inferences and predictions among others. This long list prescribes abilities in an upward trend where the purpose behind the reading is at the bottom and critiquing and reflecting on the content at the end of the reading is at the top on the list. On the other hand, Hughes (1992) refers to nine aspects divided in micro skills and macro skills. Micro skills are identified as the ability to recognize referents, vocabulary, meanings, semantic propositions, and vocabulary in context while macro skills are scanning, skimming, prediction and interpretation. To Hughes (1992), micro skills should develop as a means to improve macro skills but in a test they must be balanced since micro skills can obscure the measurement of macro skills.

Mumby (1978) describes, in his taxonomy, about two hundred and sixty skills grouped in fifty four groups that represented the abilities regarding reading comprehension only. Among those groups, there are the ones that describe abilities to decode and deduce meaning, abilities to connect meanings and retrieve an overall idea, scanning, skimming, and the abilities to comprehend, interpret and summarize. Mumby's taxonomy (1978), in his words was not 'exhaustive' (p. 117), but had the only purpose to help guide the selection of the abilities according to their role in determined tasks. Although some scholars agree that this taxonomy does not outline processes but products (Alderson, 2000).

Another example is Barrett's taxonomy (1968 in Clapham, 1996). Taxonomy created and anchored in Bloom's principles (Pearson, 2009). This taxonomy is organized in five categories: literal comprehension, reorganization, inference, evaluation, and appreciation and as in Bloom Taxonomy (1956), literal comprehension is placed at the lowest level to escalade to upper ones being appreciation the ultimate level. Barrett's taxonomy has proven to be a very useful tool, especially for teachers, because its design represents a guide for teachers to create and identify those questions that can measure the reading comprehension abilities they need in a test (Myers & Brent-Harris, 2004). Thus, Clymer (1968 in Pearson, 2009) based his redesigning of the Barrett taxonomy on Bloom taxonomy to release his own blueprint of reading comprehension. Clymer's motive was driven by his idea that the selection of the text, the questions plus the reader's background dictated the difficulty of the task and which type of comprehension to demand. Clapham (1996) refers to Barrett taxonomy -when evaluating IELTS reading exams- as appropriate because it displays most of the skills presumably tested in reading.

Davis (1944) in his revolutionary study describes nine testable skills after collecting several hundred skills which overlapped in the literature and grouped them in nine

categories. Davis' list (1944) starts with knowledge of word meanings, the ability to recognize vocabulary in context, make inferences, and determine author's intention as to the ability of paraphrasing, summarizing and retrieving main ideas. This list is unique in matching the psychometric concept to each category and also due to the analytic tool employed to acknowledge its reliability (Pearson, 2006). About fifteen years later, Davis (1968) presents the resemblance between Bloom Taxonomy (1956) with his list by cross referring them, the former shows many of the same skills under different names of the latter. After these studies, Davis (1944, 1968) concludes that reading was not a unitary process indeed and it consisted of two major components, word knowledge (literal) and reasoning (inferential) about reading. Literal comprehension refers to information explicitly stated while inferential understanding is the implicit information but requires some level of thinking and reasoning (Hosp & Suchey, 2014).These two factors are powerful enough as well as reliable to guide in the construction of tests and reading curriculum (Pearson, 2006).

These lists or taxonomies can apparently be seductive and validated due to the theory behind them and because of the power they appeal to have to detect reader's problems. Taxonomies devise the identification of single skills and characterize them in categories to be tested accordingly; and also represent a very powerful framework for test construction (Alderson, 2000). Clapham (1996) coincides with Alderson (2000) in the attractive aspect of the different taxonomies available and the issue of disagreement among one another when referring to skills and abilities. But Clapham (1996) also adds that this disagreement has to do with the little training for detecting skills and because of the wide scope of different tests and their particular requirements.

The previously discussed taxonomies, among many others beyond the scope of this paper, have been created and influenced after Bloom's "Taxonomy of Educational

Objectives in the Cognitive Domain" (Bloom, Englehart, Furst, Hill & Krathwohl, 1956). The considerations to estimate a taxonomy as pertinent or not depends on the context and the purpose. Over the years, there have been claims that have produced a breath of research to examine the validity of Bloom's taxonomy (1956). Seddon (1978) reviewed different investigations aimed to verify the properties – educational and psychological- of Bloom's taxonomy as an attempt to appraise overall findings.

Seddon (1978) concludes that the psychological properties are classified properly and that no taxonomy is perfect in providing a perfect universal understanding of a particular aspect. Bloom's categories appeared to be constructed in the right order up to the fourth except for the last two - synthesis and evaluation – based on statistical tests and correlation matrices. Nevertheless just statistical results did not make up for a clear conclusion since the qualifications in its effects were not stressed. For this reason the key is to find that taxonomy which is the best, although imperfect, means of communication (p. 320) and Seddon (1978) suggested further experiments in comparing Bloom's taxonomy with other taxonomies to reach a more stable agreement regarding the former one.

In spite of some of the issues that this taxonomy has raised regarding its validity, as previously discussed, and for reasons attached to this paper's motivation; Bloom Taxonomy is the most suitable and it will be described in detail because of its relevance for this study. (See appendix A). This research paper required a taxonomy whose parameters would allow the appropriate identification of reading abilities with the adequate reading comprehension items or questions that measured specific reading behaviors.

Bloom et al., (1956) along other collaborators theorized about the components of educational achievement and it has confirmed to be very influential along the decades in the construction of instructional material, curricula, programs and tests. Bloom (1956)

discusses this taxonomy as a construct to avoid judgment and promote neutrality in education (p. 6). Bloom et al. (1956) conceived three domains in this taxonomy to construct a system that would be inclusive but dynamic no matter the educational orientation.

The three domains envisioned in this taxonomy are the affective, the psychomotor and the cognitive one. Each of the three domains is divided under six categories: knowledge, comprehension, application, analyze, synthesis and evaluation where knowledge is at the bottom of the process and evaluation at the top. These are meant as processes that scaffold to build upon the next level.

The cognitive domain is the concern of this paper because it is here where it lies most of the theory used by the existing taxonomies to measure reading abilities. Bloom (1956) specifies the cognitive domain as the area to consider the future outcomes or behaviors from instruction but is not in any way applicable to the teaching methods or their evaluation as such.

The following is a condensed outline of the first two levels within the cognitive domain in Bloom Taxonomy. These two levels are of major importance for this study. (See appendix A for further detail) The first level is Knowledge which is divided in three subcategories: specifics, ways and means of dealing with specifics, and universals and abstractions in the field.

Knowledge of specifics has to do with the recalling and recognition of the most basic elements: words or symbols. This means that symbols can be translated, sort of speaking, into words or vice versa. This category encloses the familiarity and ability to recognize words, meanings and definitions. Also the recognition of facts connected to those words and the ability to elaborate briefly definitions of various terms and vocabulary related to science, arts, society, etc. It is the first ability to develop in order to go further up

in the next stage.

Knowledge of ways and means of dealing with specifics goes a step further the simplest recognition of words since here there exist the ability to connect those terms into a chronological sequence. This means that facts or meanings can be translated into organized patterns. It represents a more abstract level because there is a link between the specifics and their established sequence in a field. It is the most basic understanding of the process rather than the product.

Knowledge of the universals and abstractions in the field is the third and last subcategory in the Knowledge level is related to the organization of ideas, schemes or patterns in terms of specifics and how these interconnect with the field they relate to. It is the ability to summarize the universals of a particular field in a 'frugal' but logical way. It is the referent to having a highly relative insight of a field. This level shows a high level of abstraction within this level since it requires a more complex use of the knowledge although it is somehow incipient.

The second level is Comprehension which is also divided in three subcategories: translation, interpretation and extrapolation. Translation is the ability to 'translate' or communicate into other terms, language, or into other forms of communication. This subcategory is the transition into a more complex thinking where the possession of relevant knowledge is a requisite to move upward towards a grasp of more abstract ideas. It is the ability to transform an abstract idea into a much easier term as a means to facilitate thinking. It is the conversion of abstract symbols or terms into briefer and more explicit everyday terminology.

Interpretation is the second category that expresses the ability to reorganize ideas into a new order or configuration. It is the ability to summarize, generalize and interpret

essentials and differentiate irrelevant aspects to carry out conclusions. Interpretation also deals with the grasping of the major ideas or topics and the understanding of their interrelationships. This ability is also expressed in going beyond mere repetition and rephrasing by distinguish limits to what extent conclusions can be drawn.

Extrapolation, the third subcategory, encloses the ability to judge with respect to estimations or predictions based on a sample. It is the capacity to make inferences in regard to the possible consequences or implications described in a given communication. It is expressed by the capability to determine relations between a sample and a universe or vice versa. Extrapolation extends beyond the given data or findings to be expressed as inferences within certain degree of probability but not certainty.

In Bloom's words (1956): "what we are classifying is the intended behaviors of students—the ways in which individuals are to act, think or feel as the result of participating in some unit of instruction …only behaviors related to mental acts of thinking…" (p. 12). Therefore, Bloom's taxonomy succeeds in describing the skills involved in various types of reading questions and as a result this taxonomy's classifications are applied when defining and measuring reading comprehension abilities. This taxonomy offers a wide range of elements to identify what questions intend to measure, the first two levels in particular, and a major control over what needs to be assessed or tested. Therefore, Bloom Taxonomy will be used as the guide for this paper to evaluate y determine which abilities are contained in both TOEFL and CELE departmental exams.

2.5 Summary of Literature Review

According to the questions leading this paper, the literature review has been dedicated to four sections. The first one was entirely devoted to the nature of reading and how it has been noted that reading is a complex process that differentiates us from the animal kingdom. It is crucial for the adequate functioning within a society and involves the understanding of the tiniest bits of it starting with sounds to later level up to words and sentences. It was also revised that reading goes beyond the lexical aspect because as we read we also construct meaning. The second section addresses how different models have strived to explain how reading as a process takes place. Different perspectives have been born to explicate this cognitive process such as those that place reading at the bottom where decoding or parsing lead to comprehension. Or the opposite that claims reading begins at the top with aspects such as analysis and interpretation.

There is a third perspective that places reading comprehension as a flexible and active process where both processes bottom and top run parallel to complete the reading momentum. The third part attends to aspects having to do with how the assessment of reading, educationally speaking, has permeated and how areas such evaluation and testing have outlined aspects in reference to the validation and reliability of reading comprehension tests. Evaluation has created indicators and standards as an effort to trust tests as instruments to retrieve how such a cognitive process as reading comprehension occurs. This last stance leads to the fourth section which revolves around categories or taxonomies regarding reading that have been created and devised to describe more accurately how individuals read. These taxonomies function under categories whose intention is to withdraw specific skills or abilities known to be for reading comprehension in a test.

2.6 Conclusion

As indicated above, it is more than evident the wealth of literature emphasizing on reading comprehension and on all its related processes as a mystery that needed to be revealed. There are many researchers that appear to have discovered, after extensive investigations, the particularities of the reading process per se but despite its visibility in many facets it is still a cognitive process that happens inside our brain (Smith, 2004). It is pretty clear that a great deal more needs to be researched about how individuals read and how they get to certain conclusions after a text. But efforts have been made in different areas as an attempt to understand this intricate process and far beyond that: how to evaluate in a concrete way an intangible process. In regard to evaluation matters there have been numerous efforts to create exams and tests which in turn intend to assess appropriately reading comprehension abilities. These exams strive to be valid and reliable in terms to stay true to their layout and their educational purpose. There has also been an extensive catalogue of various taxonomies in respect with reading comprehension abilities. This talks about a considerable interest not only in revealing and 'unveiling' the particularities of reading as a cognitive process but also in capturing it in a more tangible, sometimes ill conceived, manner as it is with tests and examinations. Despite the extensive breadth of theories and studies whose aim is to expand our notions and knowledge about reading, it is undeniable that we still know very little. Reading as a cognitive process presents itself inside our brain but whether we understand it or not the truth is that we do certain mental procedures that have no theory to back them up and yet recognize some of its specifics. Regardless of the position one stands for, the broad wealth of reading literature confirms that it is intriguing and still quite unexplored. But as Clymer and Robinson (1961, p.139) once said: 'Outer space is not the only frontier. Reading qualifies, too.'

CHAPTER 3

METHODOLOGY

3.0 Introduction

This chapter addresses key aspects linked to the research method: the general theoretical framework, the outline of the present study, the outline of the instrument and the detailed description of data collection and analysis procedure.

As indicated earlier in chapter one, this study addresses four research questions:

- ✓ Are the reading abilities in CELE departmental exams of English (level 4) assessed in accordance with the reading abilities in TOEFL?
- ✓ Which reading abilities in CELE departmental exams of English (level 4) are being assessed?
- ✓ Which reading abilities in TOEFL are being assessed?
- ✓ Which abilities assessed in TOEFL are missing in the CELE departmental exams of English (level 4), if any?

This study looks at the reading sections and reading abilities assessed in two different tests: the CELE departmental English exams and the TOEFL (Paper-Based). Both are designed to assess individuals' reading ability. There were three stages in the project. The first was the gathering of data and the adaptation of Bloom's taxonomy into an efficient instrument. This instrument, in turn, allowed the pertinent identification and subsequent classification of the reading abilities assessed in both exams (CELE and TOEFL). The second stage had to do with the reading of the passages and items in the

CELE and TOEFL exams to identify and classify the resulting emerging abilities in both exams with the help of the instrument. This meant the organization, analysis and categorization of the occurring abilities in both exams. The third stage was the comparison of the different abilities assessed in both exams to have solid data pertaining to the abilities both exams assessed or disregarded.

3.1 Orientating Framework

Reading abilities are of a cognitive nature, already discussed in the literature review, that goes beyond quantification, For this reason, the following is an outline of the traditions in qualitative research contributing to this study and afterwards its possible merger to construct firm and trustworthy data through a well-thought out sequenced and, if possible, simple procedure for this study.

This research was conceived due to the interest to unfold the abilities, in a scientific procedural way, in one type of exam (CELE) to be compared with those existing in a different type of test (TOEFL). This research was triggered from that curiosity. Mahrer (1988) defines a researcher as a curious investigator whose drive is to find gaps in the field and address the unknown. This proactive idea comes to reinforce the posture that any researcher should consider a holistic approach in the name of that curiosity where a variety of different disciplines may lead to a more solid and reliable data. Gough (2002) refers to research by saying that it is a quite hard exercise especially when a researcher must make decisions on whether to take into account or not but in the end research is, after all, an attempt to minimize ignorance or to generate truth.

According to Wagner's theory (1993), a researcher needs all sorts of materials for either answering questions or asking them. Wagner categorizes two materials that are interrelated: *blank spots* or the actual knowledge that can help formulate questions but no answers, and *blind spots* or the lack of knowledge that generates important and fresh questions relevant to the phenomenon as a way to get an adequate picture of it. Wagner's view (1993) promotes the construction of knowledge based on a structured plan with various methods of research with successive stages where one possible solution is related to the next step of the program.

Parting from the above ideas, grounded theory was considered pertinent to guide this study due to its capability to report research about a phenomenon as a way to generate or discover a theory beyond mere description (Creswell, 2007). By generating theory, Glaser and Strauss (2012) mean that the hypotheses or concepts do not come from data only but are methodically handled in relation to the data along the development of the investigation. Although Charmaz (2005) argues that researchers normally refer to grounded theory to mean the specific mode of data analysis; grounded theory methods are, essentially, 'a set of flexible analytic guidelines that enable researchers to focus their data collection and to build inductive middle-range theories through successive levels of data analysis and conceptual development' (p. 507).

Strauss and Corbin (1998) understand theory as the resulting set of well-developed categories whose interrelations form a theoretical explanation of a process, interaction or action. The resulting theory is strongly anchored in the data so this shields it to last despite possible reformulations or modifications (Glaser & Strauss, 2012). Egan (2002) reinforces this conception by adding that grounded theory has the faculty of prediction; moreover, the theoretical product is founded in the context where the phenomenon originated. Creswell

(2007) on the same vein claims that grounded theory criteria regarding the process of research and the foundation of the phenomenon in the data are the best benchmarks to assess the quality of the study.

In respect with benchmarks, Landis (1993 in Creswell, 2007, p. 217) developed –in a doctoral dissertation- a set of standards as a means to verify whether or not a study met the grounded theory criteria. Some of these standards are: (a) the study of a process, action, or interaction as the key element in the theory; (b) a coding process that works from the data to a larger theoretical model; (c) the presentation of the theoretical model in a figure or diagram; (d) a story line or proposition that connects categories in the theoretical model and that presents further questions to be answered and finally (e) a reflexivity or self-disclosure by the researcher about his or her stance in the study. If one wishes to evaluate a grounded theory study it is necessary to look for the process and the relationship among the concepts. Therefore grounded theory can be considered qualitative in its design because the purpose is to generate theory to explain a given practice or provide a framework that can be researched further on (Creswell, 2007).

The interest leading this study is also guided from an action research stance because the key attribute of this type of research is to better understand a particular aspect of professional practice with the only purpose to bring improvement (Richards, 2003). Action research can be seen as an intervention in terms of producing recommendations and actions towards the aspect under scrutiny. In this respect the researcher can be perceived as an emancipator or an interventionist whose main goal is to improve or bring change to a working practice (Carr & Kemmis, 1986). Action research seeks to study a certain situation, consciously, to either change it or improve it and whether there is action or not, whether change happens or the situation remains the same, this does not invalidate or

undermine the value of action research (Berg, 2001).

Kemmis & McTaggart (2008) reinforce this type of research when they state that concentrating on concrete and specific practices makes reflection upon them available. This reflection leads to the reconstruction of past conditions that can potentially be modified for present and future scenarios. This study aims to make a guided reflection upon a working practice, reading comprehension abilities in tests, to collect the appropriate evidence in favor of a change over certain aspects regarding that practice. The leading principle of this type of research is to link investigation to teaching to improve effectiveness of the teaching practice (Dörnyei, 2007).

This study can also fit in the case study category since it focuses in the specificity of a particular phenomenon: to detect and determine the reading abilities in a sample of six departmental English exams of just one specific level (level 4) within a two-year window from a particular educational institution (CELE). Any case study involves detailed examination and pondering of a situation, a site, a group or an individual and despite the debate whether a case study falls into the category of method or paradigm (Richards, 2003), what it remains true is its focus on the exploration of a particular entity or set of units where sampling is purposeful by nature. The purpose is to gain an insight of specifics about the entity or unit to get a better understanding of it.

According to Creswell (2007) a case study is delimited by time and place ruling out the idea of a generalization and it is chosen to be illustrated because it is unique, unusual, accessible or ordinary and within boundaries but whatever the case is this selection requires careful attention. Richards (2003) adds that although this research tradition can mistakenly be thought as a mere detailed description as sufficient to present results, the truth is that it entails a rigorous formulation of themes to be fully interpreted and asserted. Besides,

gathering enough information and deciding on the limits in terms of time, occurrence and maturity presents a challenge as well. The case study approach can include quantifiable data, for example, the percentages within a test resulting from the frequency of some salient characteristics (abilities) that can be compared and analytical descriptions where those same specifics can be described.

3.2 Research Design

In order to gain insight into the reading abilities of the exams under evaluation, the reading data collection in this context consisted of seven reading sections (six CELE exams and one TOEFL test) - questions included, this meant a thorough analysis of every single item to identify the resultant abilities. This approach of a qualitative character enabled the identification of the reading abilities to be investigated; nonetheless the frequency and incidence at which these occurred required a more quantitative approach.

Therefore, the treatment of the data was done under a mixed-method approach. Creswell (2003) conceives this approach as a mix where knowledge claims, based on a pragmatic posture, can be best represented by the gathering of both numerical data and text information. This notion of mixing methods is neither rare nor recent. Campbell and Fiske (1959) were, perhaps, the pioneers of this concept when endorsing what they called "multitrait - multimethod matrix" (p. 81) as a legitimate position to examine data from different perspectives. According to this matrix, as a means to guarantee validity, one phenomenon can be examined with two or more methods as long as they render correlations about the same trait at the end, regardless of the approach. Nevertheless,

Chenail (2000) advises a careful administration of various techniques at the same time in order to have a complex process become simple.

A mixed-method, then, embodied the adequate fusion to enable data to be analyzed from its "dynamic reality" –qualitative- but also from its "stable reality" –quantitative- (Cook & Reichardt, 1979 p. 10). Denzin (1978) supported multiple methods, preferably qualitative, as a way to reduce plausible weaknesses of single methods by counterbalancing one another as the key principle to reach methodological triangulation and validity. Chenail (2011) advises that whatever approach one chooses whether from a designer – case study, grounded theory, etc. - or of an eclectic nature the assorted procedures must synchronize so the data flow is coherent and smooth. Undeniably the main appeal to use a mixed method research lies in the use of the best of both paradigms because by increasing the strengths the weaknesses are reduced (Dörnyei, 2007) to benefit, at the end, the understanding of the phenomenon under study. The structured hybrid method for this study helped in obtaining a solid and valid indication of the reading abilities and their occurrence within the different reading sections.

3.3 Data for Analysis

The exams used for this study were a sample of six CELE departmental English tests of level four corresponding to two academic years, 2012 and 2013. The second exam was TOEFL whose sample reading tests were taken from a preparation book of commercial distribution – TOEFL™ Preparation Course (Mahnke & Duffy, 1996). When the exams had been gathered and the corresponding reading samples selected the data were then categorized by year and period (spring, summer, and fall) and then the reading samples by their format (multiple choice, matching, true or false and open questions). Therefore

the next step was to consider the sections and items administered in all the exams to get an overall view of their format, number of items and types of exercises including the percentage they accounted for. The following table shows an overview of the samples considered. After this was done, the next step was to determine the ability being tested in each item within each reading sample.

Table 1: Summary of the Data collected to analyze Reading Abilities.

	CELE Spring 2012	CELE Summer 2012	CELE Fall 2012	CELE Spring 2013	CELE Summer 2013	CELE Fall 2013	TOELF Standard PBT
Items per section	10	10	10	10	5	10	50
Number of passages	2	1	1	2	1	2	5

3.4 Instrument: Bloom Taxonomy

As mentioned before, a detailed instrument was used to analyze the collected data in the present study and it is outlined in more detailed below. The instrument used (Bloom's Taxonomy) for this research was accurately useful in terms of facilitating the identification and insertion of the significant and predominant categories of those abilities present in both types of exams (CELE and TOEFL) as a means to yield commonalities or divergent features.

Bloom's taxonomy (1956) was considered for this study since it offers a very acute list and description of the different abilities to be demonstrated in seven levels under the cognitive aspect. This taxonomy attempts to facilitate the design of curricular developments and of evaluation devices. Therefore this taxonomy encloses those most significant behaviors considered within an educational environment but its primary goal is to help

teachers construct and design curricula, specify objectives and prepare evaluation devices accordingly. This taxonomy reflects the effort to make the exchange of information easier and much more fluid and provides a wide spectrum of abilities and specifics to suggest an indication of an individual's general skills in reading comprehension. It was conceived to be a valuable source to help, in a constructive manner, solve problems in the creation of programs and tests.

Once the taxonomy had been selected the next step was to condense the different pertinent abilities for this study under a format that would allow an easy and efficient identification of the abilities contained in the reading sections to be analyzed. The instrument took the shape of a table with a detailed description of the different abilities enunciated in the first two levels of Bloom's taxonomy: Knowledge and Comprehension.

Mahrer (1988) relied on category systems – usually custom designed- as dependable instruments to provide a closer and better look at data. Category systems – limited to their own class of data- are careful and rigorous enough to keep the researcher honest (p. 698). These particular first two levels (Knowledge and Comprehension) were considered as the ones to help conducting the analysis because in terms of reading comprehension they both represent the first two steps to show ability. Miles, Huberman and Saldaña (2014) advocate clustering as a qualitative grouping tactic that can facilitate conceptualization of a phenomenon.

The level of Knowledge is the baseline to recognize and identify words, meanings and definitions. The Comprehension level deals with aspects link to interpretation and the ability to go beyond the literal. When both levels work hierarchically – that is in harmony- they can demonstrate the use of a specific abstraction under a particular condition. In reading, this is translated in the abilities deployed depending on the requirements of a given

reading.

The third level in Bloom's taxonomy –Application- was not considered because in respect to reading comprehension it involves the ability to write papers or essays to apply the theory to offer solutions to problems. Evidently CELE exams and TOEFL do not require these tasks therefore the instrument engaged just in the specifics of the first two levels. Mahrer and Boulet (1999) agree that a preselected list can help in finding those most valuable and interesting aspects in a study.

A possible drawback of a selected list is the omission of latent data not considered previously or that the list prescribes features before they are actually studied. Although Mahrer and Boulet (1999) are categorical when they say that this can be overcome only if the researcher goes beyond the list and judges the data in terms of their specifics- context, content and characteristics- for pertinent modifications.

After the levels had been chosen the next step was to organize them in an easy-to-use table. The resulting table, presented in Appendix B for a wider and more complete view, was made up of two sections. Each section was designed to provide a descriptive guide to identify the level and the ability for each of the exams items. As it can be seen in the table (Appendix B), level one (Knowledge) was designed to provide an informed classification of different subcategories of abilities and the identified behaviors described regarding knowledge of specifics, universals and abstraction. Level two (Comprehension) was devised to identify the exams' abilities assessed in terms of translation, interpretation and extrapolation.

This type of instrument was a valuable method to determine the abilities assessed in the seven exams sampled. The two levels involve abilities as follow: level one; knowledge of specifics, knowledge of ways and means of dealing with specifics, knowledge of the

universals abstraction in the field. While level two describes abilities such as translation, interpretation and extrapolation.

3.5 Data Analysis

During this stage, the task was to do the analysis of each reading section according to the list described in the instrument. Each exam was described in detail in terms of the format, the number of passages and items, the types of exercises and the abilities assessed. Afterwards each table was transferred into condensed tables for better visualization with a detailed display of the abilities, items and percentages.

This data analysis considered an interpretative perspective. An interpretative orientation can translate human action – depicted in a text- into analysis to disclose meanings and actions (Berg, 2001). The procedure consisted of reading every item (question or statement) in each reading passage to then analyze them in terms of the abilities they were assessing. Then, the ability was identified and tallied in the appropriate level and category.

Berg (2001) considers that the resultant emerging categories should display the relevant meanings as accurate as possible within the messages to dissolve superficial or arbitrary ones. The type of analysis (content) procedure deems all the materials or data to be analyzed first in terms of the criteria selected to be later considered or disregarded. This prevents the researcher from only analyzing materials that support a given hypothesis (Holsti, 1968).

This procedure was done for all the exams yielding a total of seven tables, one per exam analyzed. The analysis carried out, in this study, with the items of the different reading sections could fall within the category of content analysis. According to Holsti

(1968) content analysis is conceived as a versatile method of research that serves as the basis of inferences to untangle the content of any communication. Osgood (1959 in Holsti, 1968) defined content analysis as instrumental when the inferences yielded are valid despite and beyond the control of the source or of contingency when the structure is entirely subjective.

In other words, the former refers to the making of inferences that rely on spoken or written language of a message; here the process is of a psychological nature. While the latter, on the other hand, means that inferences deal with the effects on listeners or readers upon a message this of an emotional nature. To sum up, content analysis considers the analysis of texts and questions them in terms of pre existing categories and emergent themes to either generate or test a theory (Cohen, Manion & Morrison, 2007). It is systematic, replicable, observable and rule governed (p. 476). Content analysis is of a dual purpose because it involves coding, categorizing and comparing but they serve for the purpose of making meaning out of those aspects to conclude with a theory.

At the end, a total of fourteen reading passages were examined where one hundred and five items were read, analyzed, categorized and tallied for their further management in the comparison process. During this stage, the management of data reached saturation. Glaser and Strauss (2012) stress that data gathering should continue until a researcher has achieved theoretical saturation; in other words, when the common features keep on replicating previous themes or subthemes. Nevertheless after compiling the information from both types of exams, it was also necessary to process the qualitative data results (abilities) in percentages (numbers) as a way to show the occurrences of the ability per category and level per selection examined.

3.5.1 Constant Comparative Method

After data had been already analyzed in terms of the levels and abilities of both types of exams (CELE and TOEFL) the next step was to carry out a comparison of the abilities in the reading sections between both types of exams to yield the appropriate results from the analysis.

The last phase of this research consisted of a comparative analysis between the two exams in respect with abilities. Therefore the next step in the data analysis was to use the comparative method as discussed by Glaser and Strauss (2012) and Strauss and Corbin (1998). Once the items of data had been allocated under the same themes or categories they are compared and contrasted. Comparing data helps in establishing analytic distinctions and by sequencing comparisons at every level makes differences and similarities easier to understand and manage (Charmaz, 2006). Glaser and Strauss (2012 p. 102) prescribe that the 'use of the constant comparative method of joint coding and analysis is to generate theory…by using explicit coding and analytical procedures'. The objective of carrying out a theoretical comparison is to perceive the evident properties objectively within the data rather than just considering names or classifications (Strauss & Corbin, 1998).

In the present study, once the data for all seven reading sections were ordered and entered into a table, the themes were condensed as a means to make the summary evident and concise. Then the next step consisted of describing the similarities and differences between both types of exams. Glaser and Strauss (2012) refer to this phase as the constant comparative method where the goal was to clarify evident themes that had emerged and what they meant parallel to the identification of sub-themes among themes. The comparative method dictates that by revising the data the researcher confirms the valid,

the negative and the discrepancies to keep a score of the cases (Cohen et al., 2007). This helps in the generation of alternative theory as well for the ruling out of discrepancies.

The process of comparing was relatively simple by the fact that there was just a handful –yet saturated- of entries (abilities) in the table for each of the exams; however the pouring of these comparisons into words was far from being friendly. The aspects to be considered for this comparison are as follows: the levels in respect with the instrument clusters, the abilities and their occurrence- numerical frequency-, the type of exercises and items, this last in reference to the reading section length and at last the format.

3.6 Summary

In brief, this chapter presented an overview of the research groundwork, theoretical design, and the data collection used in the present study. Data collection was gathered and a detailed description on the reading samples involved in the study was given. In addition to a summary of the creation of a formal instrument used to detect and analyze which reading abilities existed under Bloom´s taxonomy and how they could be translated into the different reading sections of the exams under study. This acute instrument -purposefully outlined for this project- contained a polished list of the different abilities mentioned in Bloom's taxonomy and was used for the analysis of data. The data supplied provided with plenty of evidence to move to the comparison procedure and final analysis of the resulting data.

A qualitative approach was considered and used since a qualitative description can be useful in determining the reading abilities assessed in a test and to what extent they are actually tested. Nevertheless a quantitative approach was also used as well as a means to account for the percentages and occurrences of the different abilities in the samples

considered. In the next chapter general descriptions are given of the data analysis for the exams under analysis.

CHAPTER 4

DATA ANALYSIS AND INTERPRETATION

4.0 Introduction

This chapter aims to present an overall description of the data analysis appropriate for qualitative research and pertinent to this present study as well as the steps taken heading to a detailed analysis of the reading sections and segments of seven exams (six CELE reading segments and one TOEFL test) according to the abilities they intended to assess. In this section a detailed analysis of the different reading sections and the abilities they assess is described in detail. The main focus of this study was to identify and compare the different reading abilities being assessed in two types of exams: CELE departmental exams and a TOEFL test. The technique used to identify such abilities was the use of the previously constructed instrument (based on Bloom's taxonomy), previously mentioned and described in chapter three, pertinent to keep track and tally the data and the categories of emerging abilities.

4.1 TOEFL: Reading Section

Test of English as a Foreign Language or TOEFL, as commonly known, is a standardized test designed to evaluate English language proficiency of individuals whose first language is not English. It is a multiple choice test consisting of three sections: listening comprehension, structure and written structure and reading comprehension. The reading section, the focus in this study, tests the ability to understand the meanings and uses of words in written English along with the ability to understand a variety of written materials. It consists of five reading segments or passages followed by questions – a total of

fifty- that require informational or inferential answers.

This is supported by Mahnke & Duffy (1996, p. 16) who state that reading comprehension section in the TOEFL is tested through questions that retrieve on the topic, facts, and details, inferences and implied information, author's purpose, attitude and tone in the passage as for vocabulary, this is tested in context. Over the years, TOEFL has gained the stature to be considered as a criterion for the validity of other tests to the point that tests are valid as long as they agree with TOEFL as it is meticulous to demonstrate that it measures what it is intended to be measured (Alderson et. al., 1987). Alderson et al. also note that despite the practical limitations due to its scope and purposes, TOEFL remains the best of its breed (p. 81) in terms of validity and reliability.

4.1.1 TOEFL: Reading Abilities

The TOEFL reading section sampled in this study was taken from Mahnke and Duffy's (1996) preparation book for TOEFL. The reading section considered is from the Diagnostic test that is one of the three complete TOEFL tests in the book. The selection of the diagnostic test over the other two was done randomly since the three complete tests have the same structure in terms of format and abilities therefore the choosing of either one does not affect in any way the reliability of the analysis of the data in this study.

The reading section in the Diagnostic test chosen for this study consists of fifty items distributed in five passages. Each passage is followed by either nine, ten, eleven or twelve questions depending on the length of the passage. The passage length ranges from 300 to 600 words. The reason for long texts lies in the fact that these attempt to reflect the academic experiences that students deal with and to facilitate their reading comprehension (Cohen & Upton, 2006).

Enright et al. (2000) consider that the manipulation of the text length, type and task allows for a reliable, efficient and fair evaluation of an individual's academic reading ability since the above conditions mirror similar academic conditions at college where students encounter comparable formal reading settings. The five passages were analyzed in terms of the items (questions) to obtain a list of the abilities that they were assessing and then have a general scope of what was assessed. The following is a detailed outline of these abilities.

After careful reading of the items and their categorizing certain categories emerged. For example, it can be observed that the TOEFL reading section falls within the first two levels of Bloom's taxonomy (1956) as previously described. These are, first, the Knowledge level, where 22 items retrieved information related to vocabulary, terminology and facts. While in the second level: Comprehension 28 items drew on connotative language to be translated into denotative, vocabulary in context, summaries and inferences. From a quantitative perspective TOEFL reading section alone assesses 56% of comprehension and just a 44% of knowledge which shows a major interest in the area of inferences and vocabulary in context as well as facts and terminology as the major aspects considered. For a better illustration of the items and abilities see Table 2 for reference.

Table 2. Description of TOEFL reading abilities			
	Ability	Format of recurring TOEFL items	Number of items
Level 1 Knowledge	Ability to distinguish referents for words from others.	The pronoun **"it"** in line 12 **refers** to (A) abcd (B) abcd (C) abcd (D) abcd	2
	Knowledge and mastery of vocabulary and terms of fine arts, science, accounting, geometry and quantitative thinking.	The author of the passage **describes** atherosclerosis as (A) abcd (B) abcd (C) abcd (D) abcd	4
	Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.	**According to the passage** what was the effect of the appearance of stoves in Cape Cod Cottages? (A) abcd (B) abcd (C) abcd (D) abcd	16
Level 2 Comprehension	-Ability to translate non-literal statements (metaphor, symbolism, irony, exaggeration) to ordinary English.	The **phrase** "reigns supreme" in line 7 is **closest in meaning to** (A) abcd (B) abcd (C) abcd (D) abcd	3
	-Ability to comprehend the significance of words in their context.	The **word** "dallied" in line 14 is **closest in meaning to** (A) abcd (B) abcd (C) abcd (D) abcd	14
	-Ability to grasp the thought of a work as a whole at any level of generality.	What does the **passage mainly discuss**? (A) abcd (B) abcd (C) abcd (D) abcd	5
	-Ability to deal with conclusions in terms of inferences.	We can **infer from** the passage that bees are well suited to collecting pollen because of (A) abcd (B) abcd (C) abcd (D) abcd	6
		Total of TOEFL items	50

In the above table it can be observed that the two areas that are given most of the weight are those related to recalling facts and terms and vocabulary in context. Within the level of Knowledge the ability to recognize terms and facts alone accounts for 72% therefore it is evident that this ability is prioritized over the other two recurring abilities contained in the Knowledge level -Knowledge of Specifics. This emphasis could be explained by the fact that this ability is one of the highest inside the Knowledge of Specifics and it is at the threshold of the next level.

Bloom et al. (1956) describe it as the recognition of facts connected to those words and the ability to elaborate briefly definitions of various terms and vocabulary. It is the very first ability to be developed as a means to move up to the next stage. The other two remaining abilities belong to the lowest level of Knowledge because they only refer to the recognition of words (10%) and their meanings (18%). For an overall representation of these percentages refer to Table 3.

Table 3. Reading abilities and their representation in percentages.

	Ability	Number of items that assess this ability	Percentage of the recurrence of the ability in TOEFL	Percentage of the recurrence of the ability per LEVEL
Level 1 Knowledge	Ability to distinguish referents for words from others.	2	4%	10%
	Knowledge and mastery of vocabulary and terms of fine arts, science, accounting, geometry and quantitative thinking.	4	8%	18%
	Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.	16	32%	72%
Level 2 Comprehension	-Ability to translate non-literal statements (metaphor, symbolism, irony, exaggeration) to ordinary English.	3	6%	11%
	-Ability to comprehend the significance of words in their context.	14	28%	50%
	-Ability to grasp the thought of a work as a whole at any level of generality.	5	10%	18%
	-Ability to deal with conclusions in terms of inferences.	6	12%	21%

These three abilities described in the Knowledge level and assessed in TOEFL represent lexical access, syntactic parsing and semantic proposition encoding respectively. Once the first two have developed, the semantic proposition activates automatically and the three represent the act of combining words and meanings into basic clauses and sentences to make sense of the text in question. It works as a network where the three processes connect with one another, developing automaticity of lower level processes will allow

adequate resources available for higher level text-processing (Grabe &Stoller, 2002; Grabe, 2009; Pressley, 1998; Urquhart & Weir, 1998; Westwood, 2001).

On the other hand, in the Comprehension level, the highest ability was that regarding vocabulary in context by representing 50% within this level. This illustrates that there is a particular emphasis in vocabulary but in relation to a context because it took up half of the items measuring comprehension. While the ability of exercising inferences recounted for just a 21%. Inferencing is one of the many abilities in the extrapolation category which represents the highest level in comprehension, but it is just mildly stressed in the TOEFL section. The other ability left with a low percentage had to do with being able to translate connotative language into denotative in the use of metaphors or symbolic language (18%) and the lowest percentage was for the ability to generalize and summarize which just reported a 11%.

The above abilities assessed by TOEFL belong to those considered of higher-level because they go beyond the literal; they extend to interpretation, analysis, and synthesis of information. Their nature is to construct comprehension and develop processing abilities for fluent reading (Alderson, 2000; Grabe, 2009; Grabe & Stoller, 2002). These abilities are expressed through the adjustment of the reading practice to the type of text so it can be paraphrased accurately. Comprehension abilities denote the relation of new ideas to background knowledge and offer conclusions drawn upon the text. They also demonstrate the ability to consider the author´s purpose or intentions, judge the text and verify or repair predictions or inferences (Perfetti et al., 2005; Pressley, 1998; Smith, 2004; Solé, 2000).

Along with the abilities that are being assessed it is also very important to consider the fixed format of TOEFL. The format is entirely conceived as multiple-choice, which has been preserved despite time and severe criticism because of its restrictiveness to assess

certain areas, that fact that guessing can be facilitated and that the washback effect can be minimum (Hughes, 1992). Perhaps the negative perception towards multiple-choice pertains to its institutionalization because once it has acquired this status it is perceived as superficial, inappropriate or even artificial when in reality this format is the most appropriate in certain specific situations -same cases under debate were the cloze and the dictation techniques (Bachman, 1990).

Although true at some point the above affirmations, TOEFL has strived, to a great extent, to design a test whose test items suit the test-taker population and are fair but also economical in terms of both time and money (Cohen & Upton, 2006). It is evident that a reading comprehension test does not resemble regular reading outside school but both situations require the engagement of comprehension processes and multiple-choice questions can as well draw on reading features at different levels and definitely create particular comprehension and response processes (Rupp et al., 2006).

Multiple-choice tests have proven to be practical and reliable but the idea that test items are easy to write is imprudent because even though they appear simple in the surface they make high demands on time, design and expertise (Brown & Abeywickrama, 2010; Hughes, 1992) besides they can be expensive and difficult to prepare because they have to be rigorously edited to ensure their fairness and reliability (Weir, 1995)

Also, it is important to note some of the most noticeable patterns –test items- that help in identifying the ability in TOEFL. For example, all the test items without exception are structured under a four-option mode (or distractors). This model of four options or distractors complies with the aspect of creating possible responses evenly distributed where just one answer is correct but must be genuinely correct to avoid dubious items (Alderson et al., 1995).

The reason behind this is that this is how one can identify high-ability readers from the low-ability ones in their ability to differentiate items along the test in terms of their facility or difficulty (Brown & Abeywickrama, 2010; Weir, 1995). Munby (1968 in Alderson, 2000) was an advocate of multiple-choice and used it exclusively by designing each distractor carefully in every question with the idea that each one represented a possible misconception of some aspect of the text. Alderson (2000) states that the construction of multiple-choice questions is a skilled `business´ because writing plausible incorrect choices is far from easy plus the modification when those items do not perform well.

Moreover in respect to the language used in the structuring of the items there are some key words that emerged as fixed patterns to hint in the identification of the abilities at play. When, for instance, an item contains the phrase *"according to the passage"* it is asking for retrieval of meaning from the text so is of a recognition nature since the procedure just requires referring back to the text and regain meaning of facts. There are other items that ask for a word and *"its closest meaning to"* this refers to the ability of estimating the meaning of a word within a context and discriminate over the other possibilities but if there is a phrase inquiring about *"its closest meaning"* the ability at test is the translation of a symbolic feature into explicit language (connotative language converted to denotative).

Some items are explicit in terms of their structure and when they establish words such as *"infer from"* or *"mainly discuss"* they are openly assessing the ability to make an inference or the ability to design a summary or grasp a general idea. Items that assess the ability to access to basic vocabulary or definitions are normally structured through verbs such as *"refer to"* or *"described as".* In this respect TOEFL is very transparent in what it is assessing and how it is done owing to its regulated framework for item patterns. The

judgment of reading comprehension items is beyond the scope of this study and it is not intended to be part of this paper. But it is pertinent to mention that TOEFL item examination was friendly in terms of identifying the abilities; very different from the area of item writing which is far from being straightforward or trouble free.

4.2 CELE: Reading Section

CELE departmental English exams take place after the end of a course. They are designed to assess and evaluate achievement of college students after their instruction. It consists of five sections: listening comprehension, linguistic competence, reading comprehension, written competence and oral competence. The reading section, the focal point in this study, as well as TOEFL, has the objective to assess the ability to understand meanings and words in written English along with the ability to understand various written texts. The format varies from period to period because it can at times have just one reading or sometimes two passages followed by questions that require informational or comprehension responses. The items vary as well since they can be presented in different modes such as multiple choice, true or false statements, match up or open questions. These latter aspects regarding format and exercises will be discussed in more detail further down this paper.

4.2.1 CELE: Reading Abilities

The CELE exams that were sampled for this study were retrieved from DEPEA prior request for this study and their availability was granted since they had already undergone probing, revision and been applied previously; therefore the material was free for release. The reading sections considered were compiled from six different exams. The

selection of the passages relates to the fact of having a considerable sample of gathered data regarding format, abilities and types of items. The reading sections in these six exams consist of a total of fifty five items distributed in nine reading passages. Each passage is followed by either five or ten questions depending on the number of the passages in each exam's reading section. The nine passages were analyzed in terms of the items (questions) to obtain a list of the abilities that they were assessing and then have a general scope of what was assessed. The following is a detailed outline of these abilities.

The visible feature that emerges before the reading of the items and the categorization phase was the different formats of CELE reading sections which present an irregular array of items per section, type of exercises and number of passages per exam. This layout differs from exam to exam and so do the abilities. The next segment is dedicated to the salient features in CELE reading sections in regard to the format and treatment of items and exercises. Refer to table 4 for a general visualization of the layout.

Table 4. Different CELE exam layouts

Type of Exam	Total Items in reading section	No. Total passages	No. Passages Per exercise	Type of exercise	No. items per type of exercise	Percentage of the exercise
CELE Spring 2012	10	2	1	Multiple choice	5	50%
			1	Matching	5	50%
CELE Summer 2012	10	1	1	Multiple choice	10	100%
CELE Fall 2012	10	1	1	Multiple choice	10	100%
CELE Spring 2013	10	2	1	Matching	5	50%
			1	True or false	5	50%
CELE Summer 2013	5	1	1	Multiple choice	5	100%
CELE Fall 2013	10	2	1	Multiple choice	3	30%
				Open questions	2	20%
			1	Matching	3	30%
				True or false	2	20%
TOEFL Standard PBT	50	5	5	Multiple choice	50	100%

It can be observed that the number of items included in the reading section is often ten except for the summer 2013 that has only five. The number of passages also differs

since in some exams (spring 2012, 2013 and fall 2013) there are two while in others only one passage is considered (summer 2012, 2013 and fall 2012). As for the types of exercises, it can also be seen that various types of exercises range from multiple choice to open questions. Some passages are tested via two different types of exercises as shown in the spring 2012, 2013 and fall 2013. While other exams based their testing completely on multiple choice like the ones in the summer and fall 2012 and summer 2013. This shows multiple-choice as a common type of exercise to test reading comprehension and despite the variety of the types of exercises and items its inclusion is evident in all of the CELE exams regardless of the period when they were designed.

The above table suggests that CELE exams layout does not consider a specific standard in regard of the reading section, as TOEFL does, in terms of what types of exercises should be included since exams appear to be miscellaneous in number of passages, exercises and items. This assortment of exercises provide with an ample collection of options in the visual aspect because it appears to be dynamic in contrast to just one type of exercise like the multiple choice mode in TOEFL. As diverse as this format may look it is pertinent to mention the apparent absence of regulations because while one exam has only five items to test reading others need ten.

Complete reading sections containing only one passage or others that have two, and reading passages that have two different types of exercises while others just solely need one type. If the purpose is to offer various and varied types of exercises to test reading it is recommended to outline specifications in terms of which types can be included and the possible combinations but also a fixed number of overall items as well as passages. Therefore even if the format seems to be cluttered it would fall into a regulated layout obeying to one of the many possible combinations but always respecting aspects such as the

number of passages and items. Included here is it also the aspect of the abilities assessed through different types of exercises but the next section will be devoted to describe this in more detail.

As for the abilities, the core of this study, it must be mentioned that due to the nature of the reading sections as previously detailed in the above section. The management of the different exams was done individually, in other words, each section was read, analyzed and categorized under the prescribed aspects belonging to the selected instrument. The emerging categories and abilities will be explained in specific according to the year and period of the exam and illustrated with the corresponding table compiled in Appendix C at the end of this paper for a better visualization.

The table will display the ability, the format of the exercise in use and an example item, when possible, similar to those tested in TOEFL where it is evident to appreciate the ability assessed. Also, it will display, just as it was presented in TOEFL, two percentages of the ability both in the section and per level. The salient abilities will be described in terms of their occurrence and percentage in each reading section. And then they will be compared to those in TOEFL to withdraw the appropriate conclusions and answer the questions leading this study.

4.2.2 CELE Exams

4.2.2.1 Spring 2012

This exam completely disregarded the Knowledge level since no item measures either vocabulary or facts. The items leaned towards the Comprehension level where the ability to summarize or generalize accounts for just 10% while the ability to paraphrase is given 40%. The major focus is reflected in the ability to interpolate (50%) which can be

expressed by going beyond the data and the ability to complete the missing information. Types of exercises that assess this last ability are called transfer tasks (Weir, 1995) where missing information is matched up such as the arrangement of jumbled texts or sentences within a text. And as an end note, the multiple choice exercise failed to cover the four-option mode by just presenting three distractors.

4.2.2.2 Summer 2012

This exam tests only one of the many abilities in the Knowledge level being the one that recalls facts with a 10% and skips through the Comprehension level. In the Comprehension level there is an assortment of various abilities assessed such as paraphrasing, connotative language and vocabulary in context all covered by just 10% each. The ability of summarizing or generalizing counts for 20% along with the ability to interpret with depth and clarity. This leaves at the end of the row the ability to make inferences and the ability to draw conclusions with only 10% each. As for the format, it is presented in multiple-choice entirely but it fails to show the four-option mode by just considering three distractors.

4.2.2.3 Fall 2012

Similar to the summer 2012 exam; this one, once again, only tests solely the ability to recall facts in the Knowledge level although here it takes up 20%. On the other hand, in the Comprehension level, the ability to identify vocabulary in context reports a 20% while the ability to make summaries or generalizations recounts for a 40%. The ability to make inferences is paired up in percentage with the ability to draw conclusions with only 10% each. The abilities in this exam were reduced since some abilities engrossed most of the

items such as the one related to generalizing. Although the multiple choice format is used in this exam, it does not succeed in including the four-option mode so characteristic of multiple choice exercises in general by just repeating the same three-distractor structure.

4.2.2.4 Spring 2013

This exam, relatively similar to the spring 2012 test, assesses two abilities only. One within the Knowledge level is the ability of recognition and recalling of facts which engrosses half of the items (50%) and are all contained in just one exercise. The exercise is limited since it simply exercises the above mentioned ability through true and false statements and this reduces the scope of abilities to be tested. The second ability described emerges after skipping most of previously mentioned abilities in the Comprehension level and lands on the ability of interpolation, same case is the spring 2012 exam, where there is a transfer task or ordering task (Alderson et al., 1995) where missing information is matched up such as the arrangement of jumbled texts or sentences within a text. This ability is assessed at 50% along the remaining five items in just one exercise which evidently shows the short scope of abilities assessed in this particular exam limited to only two.

4.2.2.5 Summer 2013

A very similar case in respect to the spring 2012 and 2013 exams is this exam in particular. This exam is a very unique case because it is the only one that contains just one passage and only five items under the multiple choice modality. The ability assessed in this reading section is the one belonging to the Knowledge level: the ability to recognize and recall facts in 100%. The scope of abilities is evidently reduced and limited, nor only to just one passage but to one ability singlehandedly. Also worth mentioning the condensed

number of only five items instead of ten and the fact that the Comprehension level is ignored completely. Nevertheless this is the only multiple choice exercise that respects the four-option mode.

4.2.2.6 Fall 2013

Another peculiar case is this exam because it presents a much cluttered reading section with two passages, each one assessed with two different types of exercises. The first passage is tested via two open questions and three multiple choice items (structured under the four-option feature). The second one is assessed by a transfer task (Alderson et al., 1995; Weir, 1995) and two true and false statements. In terms of the abilities, this exam considers both levels. In the Knowledge level the abilities measured with 10% are those regarding referents, this means the plain recognition of words (lexical level) and the ability to recognize and recall facts. The highest scoring ability in this level is the one about vocabulary with a 30%. As for the Comprehension level the two abilities that ranked the same 10% are the ones where connotative language is translated and vocabulary in context is deciphered. A recurring ability is that of interpolation which takes up the remaining 30%.

After describing each exam reading section thoroughly it is observable that the six samples differ from each other in aspects ranging from those on the subject of format (number of items and passages and type of exercises) to those relative to the assessment of reading ability. To sum up, CELE reading sections assess a total of eleven different abilities, where the ability to recognize and recall facts is almost present in all of them at different percentage. The ability to focus on vocabulary in context and the one referring to making generalizations are considered second along with the interpolation ability. Abilities

such as those to make inferences, draw conclusions or translate into denotative language are left third while other are just measured slightly like referents to vocabulary or the ability to interpret. The following table (Table 5) condenses the eleven emerging abilities tested in CELE exams and to what extent each one of the abilities is tested in every exam.

Table 5. Resulting abilities in CELE exams

	Emerging Ability	Spring 2012	Summer 2012	Fall 2012	Spring 2013	Summer 2013	Fall 2013
		Abilities and Percentages					
1	Ability to distinguish referents for words from others.						10%
2	Knowledge and mastery of vocabulary and terms of fine arts, science, accounting, geometry and quantitative thinking.						10%
3	Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.		10%	20%	50%	100%	20%
4	-Abstract into concrete or less abstract. "in your own words"	40%	10%				
5	-Ability to translate non-literal statements (metaphor, symbolism, irony, exaggeration) to ordinary English.		10%				10%
6	-Ability to comprehend the significance of words in their context.		10%	20%			10%
7	-Ability to grasp the thought of a work as a whole at any level of generality.	10%	20%	40%			
8	-Ability to comprehend and interpret with depth and clarity.		20%				
9	-Ability to deal with conclusions in terms of inferences.		10%	10%			
10	-Ability to draw and state conclusions.		10%	10%			
11	-Interpolate where there are gaps in data.	50%			50%		30%

4.3 CELE and TOEFL: Comparison of Reading Abilities

After carrying out the analysis of both types of exams separately, the last step was to compare them so see both similarities and differences between exams. The first aspect was to consider which abilities occur in both exams and then their occurrence. For example the following table (Table 6) shows the final abilities assessed in both exams.

Table 6. Resulting abilities in TOEFL and CELE exams	
TOEFL	**CELE**
Ability to distinguish referents for words from others.	Ability to distinguish referents for words from others.
Knowledge and mastery of vocabulary and terms of fine arts, science, accounting, geometry and quantitative thinking.	Knowledge and mastery of vocabulary and terms of fine arts, science, accounting, geometry and quantitative thinking.
Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.	Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.
	-Abstract into concrete or less abstract. "in your own words"
-Ability to translate non-literal statements (metaphor, symbolism, irony, exaggeration) to ordinary English.	-Ability to translate non-literal statements (metaphor, symbolism, irony, exaggeration) to ordinary English.
-Ability to comprehend the significance of words in their context.	-Ability to comprehend the significance of words in their context.
-Ability to grasp the thought of a work as a whole at any level of generality.	-Ability to grasp the thought of a work as a whole at any level of generality.
	-Ability to comprehend and interpret with depth and clarity.
-Ability to deal with conclusions in terms of inferences.	-Ability to deal with conclusions in terms of inferences.
	-Ability to draw and state conclusions.
	-Interpolate where there are gaps in data.

4.3.1 Similarities

A salient feature that emerges is that both exams apparently assess the same abilities despite the extra four abilities CELE added. The realm of abilities CELE measures is the same in regard to that of TOEFL which is a very important aspect because this means CELE exams consider the same abilities as TOEFL, but the extent to how much and often these abilities are measured is of major importance for this study as they are detailed further on in this paper. As for the format, it is worth mentioning that although CELE uses other types of exercises which restrain a full multiple choice mode there is a great incidence in the use of these presented in all the CELE exams examined. The similarities vanish to this point and the next consideration is to reflect on the differences.

4.3.2 Differences

It is obvious that from a numerical perspective the abilities assessed in TOEFL are only seven while CELE assesses eleven. Drawing on the idea that both exams assess the

same abilities, it is very important to see in more detail this apparent correspondence because it is here where the differences are considerably evident. As shown in Appendix D, at the end of this paper, the occurrences of the abilities differ in percentages; therefore the number of incidences diverges in terms of how appropriately they are assessed and qualified.

For instance, the ability with the highest percentage in TOEFL was the one related to recognize and recall facts with a 32% while CELE assesses the same ability at different percentages which range from 10% to 100%. This means that some CELE exams may completely assess just this ability while others just consider it at a low range. Moreover, the ability regarding vocabulary in context is assessed in TOEFL with a 28% but in CELE this one goes from 10% to 20% (just present in only half of the exams analyzed) placing it way below the required level by TOEFL.

Abilities such as detecting referents or vocabulary in general are considered in TOEFL by 4% and 8% respectively while just one CELE exam, from the six examined; considered these both in 10% each. TOEFL assesses the ability to translate connotative language to denotative by a 6% while the ability to make inferences is of 12%, in CELE exams these two abilities are measured with a 10% each which is not so far off from the requirements of TOEFL but they are taken into consideration by just two exams. The last ability that matched in both exams was the one related to producing a general idea or a summary about the topic which TOEFL assesses with a 10% but CELE measures from 10% to 40% and it is present in just half of the cases once again.

Furthermore, it is significant to mention in this regard that CELE exams assess other abilities not considered by TOEFL such as the interpolation ability which is given a high rate of a 30% to 50%, the ability to draw conclusions (10%), paraphrasing (10% - 40%) and

the ability to comprehend and interpret (20%). These rates suggest the idea that certain abilities are prioritized over other without an apparent framework since they also fluctuate in their occurrence and assessment and this also suggests a considerable attention to certain abilities while undermining those relevant in TOEFL.

In addition, there is not an evident framework in reference to the format, in other words, despite CELE considers, up to a point, the same abilities as TOEFL; exams differ in the number of passages, the number of items and the types of exercises which in turn affect the measurement of the abilities according to those outlined in TOEFL. The apparent absence of regulations is based on the fact that there are exams with only five items to test reading while others have ten.

It is observable the evident existence of reading sections combinations with either one or two passages; or passages with either one or two dissimilar types of exercises. Furthermore, exercises that measure just one ability or exercises whose incidence is not clearly ascertained because these exercises vary from test to test. Therefore, CELE exams are presented as miscellaneous with plenty of passages, exercises and items which look attractive from a visual aspect in contrast to a single type of exercise (multiple-choice) as in TOEFL. Unfortunately, even if the format may appear as unique and inventive; it happens to be cluttered, incomplete and irregular and in turn poor in terms of the assessment of abilities in comparison to TOEFL.

CHAPTER 5

CONCLUSIONS

BUAP students, CELE English students in particular, are struggling to reach a specific TOEFL score to graduate and exit from the university. They require the appropriate reading skills and abilities to grow to be skillful fluent readers. Most of these students are detected until they have undergone a previous TOEFL examination and had failed it. Many factors can be blamed for this failure such as the timing, the duration or the format of the test which in turn may limit the effectiveness of students' performance. Nevertheless, one of the core problems at display is the lack of the pertinent reading abilities –the object of investigation in this paper- to undergo such examination satisfactorily. This failure, unfortunately, indicates underachievement and it is both debilitating and nerve-wrecking.

5.1 General Conclusions

The main research question addressed if CELE departmental exams assessed the same reading abilities as TOEFL. The second and third questions were about which particular reading abilities both exams tested; while the fourth was concerned with possible missing reading abilities, if any, in either of both exams. In relation to these questions the following conclusions were drawn from this study.

From the analysis in all the aspects concerning CELE exams, it is pertinent to shed light on the idea behind these exams. They appear to have some foundation about what to assess in terms of the abilities that TOEFL measures despite the differences in format. Evidently, CELE English exams format does not comply with the pertaining one from

TOEFL; although CELE gives the impression to have an established layout. If this format is to be described it should comprise a regulated framework where many possible combinations can take place but always considering aspects such as a strict number of overall items, the number of passages and items; and types of exercises. Additionally, if there are to be various and diverse types of exercises the adequate specifications in reference to the combinations are to be included.

Also, CELE exams swing in extremes because while they can completely disregard or undermine certain abilities to the point that they are assessed poorly or not at all; others are over prioritized -one or two are the main core of the reading section in some cases. The aspect of abilities is also of key importance since they have to correspond to TOEFL in occurrence and percentages despite the format CELE considers. But the format itself is indeed a key aspect of TOEFL along with the fixed abilities it assesses; consequently CELE format would have to be adapted as well to adequately match up.

The previous information shows at a wide scope that the abilities CELE considers for assessment correspond to those of TOEFL but they are assessed neither in percentage -same rates per ability- nor in occurrences much less in the format of every exam. The abilities are present in all CELE exams undeniably, but they are randomly and dispersedly tested. This means that some CELE exams fluctuate from measuring a handful of the abilities to simply testing just one but not the entire set of salient abilities all at once (as it is established in TOEFL).

CELE exams are unsuccessful in measuring the same reading comprehension abilities as TOEFL and this is translated in salient abilities not being assessed accordingly or appropriately. Obviously, these findings, explain somehow, and just to some extent, CELE English students failure at TOEFL examinations. In the long term, CELE English

students will only confirm this deficiency and the lack of the necessary breadth of reading abilities.

5.2 Contribution and Implications

As underlined earlier, there is a significant problem with the achievement of both reading abilities and consequently a passing TOEFL score of CELE English students. CELE exams emphasize a more communicative approach with a variety of exercises in every exam as a sign of dynamism and creativity. Unfortunately, the most available certification to graduate, TOEFL, does not follow a communicative approach. This is reflected in an imbalance where CELE students are falling further behind as they move up along levels with minimum training from CELE English exams in the required TOEFL reading abilities.

CELE students usually venture to present the TOEFL test between level 4 or 5 with no evident improvement in reading and just a limited number of them can actually display some success. Clearly, for CELE students, TOEFL readings are difficult and exhausting and far beyond their level of mastery. In order to successfully obtain the required TOEFL score to graduate, CELE students need to work twice as hard and in most of the cases they become frustrated and resentful to reading and thus to English. Addressing the needs of these students in particular is undoubtedly a challenge.

However, the following are recommendations to be considered. There is the urgency to adapt CELE exams to these needs as a means to support those struggling students. This means the imminent change in the CELE exam construction in addition to the planning and monitoring of both exams and students. This change contends that there is the need to redo

procedures and policies in favor of an adaptation. For example, establish a regulated layout for the reading section such as four passages because this would resemble passages in a TOEFL test but this should be considered in terms of test construction, time and grading. Otherwise, it is advisable that they always contain at least two reading passages with length specifications and ten multiple-choice items each. The items should definitely follow the same structure as the ones in TOEFL somehow and have the same percentage and occurrences of abilities -within a scale- to match those of TOEFL.

Any TOEFL manual could be handy, for the meantime, to start with for the item writing task, as a means to familiarize with both the type of items and what they are testing. It is obvious and mandatory the need, in the future, of a structured training (by an expert) in reading abilities for teachers since it is in the classroom where they can teach these and then structure the exams with the rightful foundations. This training can be extensive to other language teachers as an action to reduce low levels in reading in their own particular language certifications.

Another consideration is that, if the modifications to the reading section take place it will take up both human and material resources. Test construction is demanding and time consuming, therefore; if the reading section is done according to TOEFL the different passages can be recycled in the future instead of exhausting resources every term in producing more. This can be applied to the other test sections as well. The task of constructing CELE departmental exams would turn friendlier and less constrained with the perk to be more reliable and valid.

5.3 Limitations

Limitations are, like most of the time, related to the human factor, in other words, the administration behind the programs and then the teachers involved in test construction. Teachers who participate in the test making task are academically trained for teaching but testing and evaluation are not necessarily the specialization of most. Consequently, teachers are extremely committed and well meant when participating in preparing exams, but are at some point blindfolded when it comes to test construction - I include myself in this last consideration-.

Then, the limitation to be pondered upon is the application of this paper's findings in regard to the format of the test section in question. The consideration of a fixed format for the reading section may not be welcome because it, in some way, goes against the prevailing internal test approach. Nonetheless, it is the format of the current and most demanded certification students need to deal with after their CELE instruction. Moreover, this paper just focused on reading abilities and did not consider other aspects such as listening, speaking or writing. Consequently, if the reading section format is modified the other sections will misbalance the departmental exam requiring either a similar change within a TOEFL format or further research for their particular improvement.

5.4 Further Research

The purpose guiding this thesis and the resulting analyses and findings comprised in this investigation provide a tangible proof to support the application of the advised changes and modifications in the CELE departmental English exam -reading section. Nevertheless, some issues remain to be considered in the future with respect to the assessment of reading

abilities and the corresponding item writing. For instance, a future pending issue for follow up would be the application of the reading format and monitor how students improve after two or three periods (an academic year). Moreover, the construction of in-class reading exercises similar to TOEFL to both reinforce the reading abilities and train students with the format.

This thesis just considered the reading abilities assessed in two types of exams which bring up the issue of the item construct to identify any given ability. This paper just retrieved the ability from the already written items but the question that this raises is how to identify, or even better, how to write the adequate items to measure a specific ability and imprint reliability and validity to reading tests. The latter consideration - item writing- is perhaps of major interest after this paper and further research could address the writing of reading items in terms of the abilities under test. A possible consideration could be the construction of a guide or manual to facilitate item writing for CELE departmental exams. Finally, this could, as a consequence, influence innovative transformation in how teachers carry their teaching practice, assessment and future testing in benefit of their students- especially the struggling readers.

REFERENCES

Alderson, J. C., & Lukmani, Y. (1989). Testing Reading Comprehension Skills (Part 1). *Reading in a Foreign Language, 6 (2), 425-438.*

Alderson, C. J. (1990). Testing Reading Comprehension Skills (Part 1). *Reading in a Foreign Language, 6 (2), 425-438.*

Alderson, J. C., & Wall, D. (1992). Does Washback Exist?. *Applied Linguistics, 14(2), 115-129.*

Alderson, J.C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation.* Cambridge University Press.

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Alderson, J. & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching, 35 (2), 79-113.*

Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol. III* (p. 285-310). Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, R.C., & Pearson, P.D. (1984). A schema- theoretic view of basic processes in reading comprehension. In P.D. Pearson (Ed.), *Handbook of reading research* (255-291). New York: Longman

Anderson, N.J., Bachman, L., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validity of a reading comprehension test: triangulation of data sources. *Language Testing, 8, 41-66.*

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, England: Oxford University Press.

Bachman, L. F. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing, 17 (1), 1-42.*

Berg, B. L. (2001). *Qualitative Research Methods for the Social Sciences.* Allyn and Bacon.

Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives. Handbook 1: Cognitive domain. New York: Longmans, Green.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment, principles and classroom practices* (2nd Ed.). White Plains, NY: Pearson Education, Inc.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, *56*(2), 81-105.

Carr, W., & Kemmis, S. (1986). *Becoming Critical: Education Knowledge and Action Research*. Routledge.

Carrell, P. L. (1984). Schema Theory and ESL Reading: Classroom Implications and Applications. *The Modern Language Journal, 68, (4), 332-343.*

Carver, R.P. (1977). Toward a theory of reading comprehension and rauding. *Reading Research Quarterly, (1), 8-63.*

Carver, R.P. (1992). What do standardized tests of reading comprehension measure in terms of efficiency, accuracy, and rate? *Reading Research Quarterly, 27 (4), 346-359.*

Charmaz, K. (1983). The grounded theory method: An explication and interpretation. In R. M. Emerson (Ed.). *Contemporary field research: A collection of readings*. Boston: Little, Brown and Company

Charmaz, K. (2005). Grounded theory in the 21st century: Applications for advancing social justice studies. In N. K. Denzin and Y. S. Lincoln (eds.). *The Sage Handbook of Qualitative Research (*3rd) (p. 507-532). Thousand Oaks, CA: Sage.

Charmaz, K. (2006). *Constructing grounded theory.* London: Sage.

Chenail, J. R. (2000). Navigating the seven C´s: Curiosity, Confirmation, Comparison, Changing, Collaborating, Critiquing, and Combinations. *The Qualitative Report Journal 4 (3/4).* Retrieved September 26, 2005, from the World Wide Web: http://www.nova.edu/sss/QR/QR4-3/sevencs.htlm

Chenail, J. R. (2011) Ten Steps for Conceptualizing and Conducting Qualitative Research Studies in a Pragmatically Curious Manner. *The Qualitative 16 (6), 1713-1730.*

Clapham, C. (1996).*The development of IELTS a study of the effect of background knowledge on reading comprehension*. Cambridge University Press.

Clapham, C. (2000). Assessment and Testing. *Annual Review of Applied Linguistics, 20, 147-16.*

Clymer, T., & Robinson, H. M. (1961). Reading. *Review of Educational Research, 31 (2), 130-144.*

Cohen, D., & Upton, T. A. (2006). *Strategies in Responding to New TOEFL® Reading Tasks.*TOEFL Monograph No. MS-33. Princeton, NJ: ETS.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (6th). Routledge.

Cook, T., & Reichardt, C. (Eds.) (1979). *Qualitative and quantitative methods in evaluation research*. Beverly Hills, CA: Sage.

Creswell, J. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications.

Creswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd).Thousand Oaks, CA: Sage.

Davis, F.B. (1944). Fundamental factors of comprehension of reading. *Psychometrika*,9, 185–197.

Davis, F.B. (1968). Research in comprehension in reading. *Reading Research Quarterly*,3, 499–545.

Denzin, N. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods.* New York: McGraw-Hill.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Egan, T. M. (2002). Grounded theory research and theory building. *Advances in developing human resources, 4*(3), 277-295.

Enright, M. K., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 reading framework: A working paper*. TOEFL Monograph No. MS-17. Princeton, NJ: ETS.

Freebody, P., & Luke, A. (1990). Literacies programs: Debates and demands in cultural context. *Prospect: An Australian Journal of TESOL, 5(3), 7-16.*

Glaser, B. G., & Strauss, A. L. (2012). *The Discovery of Grounded Theory: Strategies for Qualitative Research* (7th). Chicago, Aldine Publishing Company.

Goodman, K. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist, 6,* 126-135.

Gough, N. (2002). Black spots, blind spots, and methodological questions in postgraduate research. (DUSA Research conference at Deakin University). Retrieved September 26, 2005, from the World Wide Web: http://www.deakin.edu.au

Grabe, W. (1991). Current development in second language reading research. *TESOL Quaterly, 25 (3), 375-406.*

Grabe, W., & Stoller F. L. (2002). *Teaching and Researching Reading*. London: Pearson Education Longman.

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.

Henning, G. (1987). *A guide to language testing: development - evaluation - research.* Rowley, Massachusetts: Newbury House.

Holsti, O. R (1968). Content analysis. In G. Lindzey & E. Aaronson (Eds.). *The Handbook of Social Psychology.* Reading, MA: Addison-Wesley.

Hosp, J. L., & Suchey, N. (2014). Reading Assessment: Reading Fluency, Reading Fluently, and Comprehension-Commentary on the Special Topic. *School Psychology Review*, 43, (1), 59-68.

Hughes, A. (1992). *Testing for Language Teachers.* Cambridge University Press.

Iser, W. (1984). *The Act of Reading: A Theory of Aesthetic Response.* Baltimore: Johns Hopkins UP.

Kemmis, S., & McTaggart, R. (2008). Participatory Action Research. Communicative Action and the Public Sphere. In Denzin, N. K., & Lincoln, Y. S. (Eds.). *Strategies of qualitative inquiry* (Vol. 2) (p. 271-330). Sage.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

Kintsch, W. (1988). The role of Knowledge in Discourse Comprehension: A construction-Integration Model. Psychological review, 95 (2), 163-182.

Kintsch, W., & Mangalath, P. (2010). The Construction of Meaning. *Topics in Cognitive Science, 3, 346-370.*

Kletzien, S. B. (1991). Strategy use by good and poor comprehenders reading expository text of differing levels. *Reading Research Quarterly, 26, (1), 67-86.*

Kletzien, S. B., & Dreher, M. (2004). *Informational Text in K-3 Classrooms. Helping Children Read and Write*. Newark, DE: International Reading Association.

Koda, K. (2005). *Insights into second language reading*: *A cross-linguistic approach*. NY: Cambridge University Press.

Langer, J. A. (1995). *Reading Assessment Redesigned: Authentic Texts and Innovative Instruments in NAEP's 1992 Survey*. Washington, D.C.: The Center.

Luke, A. (1992). Reading and Critical Literacy: Redefining the "Great Debate". *Paper of the 18th New Zealand Conference on Reading. Wellington. May 10-13, 1992.* Retrieved from ERIC database. (ED 345 211).

Mahrer, A. R. (1988). Discovery-oriented psychotherapy research: Rationale, aims, and methods. *American Psychologist, 43(9),* 694-702.

Mahnke, M.K., & Duffy, C.B. (1996). *The Heinemann TOEFL Preparation Course.* Macmillan Heinemann.

Mahrer, A. R., & Boulet, D. B. (1999). How to Do Discovery-Oriented Psychotherapy Research. *Journal of Clinical Psychology, 55(12), 1481-1493.*

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

Messick, S. (1992) Validity of Test Interpretation and Use. In M. C. Alkin, (Ed.). *Encyclopedia of Educational Research (6th) (p. 1487-1495).* New York: Macmillan.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241-256.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: a methods sourcebook*. Sage Publications.

Mumby, J. (1978). Communicative Syllabus Design: A Sociolinguistic Model for Designing the Content of Purpose-Specific Language Programmes. Cambridge University Press.

Myers, S.S., & Brent-Harris, L. (2004).Teachers' Guide Manual for Formulating Reading Comprehension Questions. USAID.

Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M.L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), Handbook of reading research: Vol. II (p.609-640). Mahwah, N J: Lawrence Erlbaum Associates.

Pearson P. D., & Hamm, D. N. (2006).The assessment of reading comprehension: key historical influences in the USA. In Sainsbury, M., Harrison, C., & Watts , A. (Eds.). *Assessing reading from theories to classrooms. (p. 76-102).* Slough, UK: National Foundation for Educational Research.

Pearson, P. D. (2009). The Roots of Reading Comprehension Instruction. In Israel, S. E., & Duffy, G. G. *Handbook of research on reading comprehension*. (p.3-31). Routledge.

Perfetti, C. A. (1985). *Reading Ability*. New York: Oxford University Press.

Perfetti, C.A. (2001) Reading Skills. In N. J. Smelser & P. B. Baltes (Eds.). International encyclopedia of the social & behavioral sciences (p. 12800-12805). Oxford: Pergamon.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook.* Chapter 13 (p. 227-247). Oxford: Blackwell.

Perfetti, C. A. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading, 11 (4), 357-383.*

Perfetti, C. A. (2010). Decoding, vocabulary, and comprehension: The golden triangle of reading skill. In M. G. McKeown & L. Kucan (Eds.), *Bringing reading researchers to life: Essays in honor of Isabel Beck* (p. 291-303). New York: Guilford.

Pressley, M., El-Dinary, P.B., Gaskins, I., Schuder, T., Bergman, J., Almasi, L., & Brown, R. (1992). Beyond direct explanation: Transactional instruction of reading comprehension strategies. *Elementary School Journal, 92,* 511-554.

Pressley, M. (1995). More about the development of self-regulation: Complex, Long-term, and thoroughly social. *Educational Psychologist, 30 (4), 207-212.*

Pressley, M. (1998). *Reading instruction that works: The case for balanced teaching.* New York: Guilford.

Rhoder, C. (2002). Mindful reading: Strategy training that facilitates transfer. *Journal of Adolescent and Adult Literacy, 45(6), 498-512.*

Richards, K. (2003). *Qualitative inquiry in TESOL*. Basingstoke: Palgrave Macmillan.

Rupp, A., Ferne, T., & Choi, H. (2006) How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective. *Language Testing, 23 (4), 441-474.*

Seddon, G. M. (1978). The properties of Bloom's taxonomy of educational objectives for the cognitive domain. *Review of Educational Research,* 303-323.

Smith, F. (2004). *Understanding reading: A psycholinguistic analysis of reading and learning to read*. Mahwah, N J; London: Lawrence Erlbaum Associates.

Snow, C. (2002): *Reading for understanding: Toward a R&D program in reading comprehension.* Rand Corporation.

Solé, I. (2000). *Estrategias de lectura.* Barcelona, ICE de la Universidad de Barcelona.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16 (1), 32-71.*

Stanovich, K. E. (1982). Individual differences in the cognitive processes of reading: 1. Word decoding. *Journal of Learning Disabilities, 15(8), 485-493.*

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21(4), 360-407.*

Stevenson, D. (1987). Test of English as a Foreign Language. In Alderson, J.C., Krahnke, K. J., & Stansfield. C.W. (Eds.). *Reviews of English Language Proficiency Tests*. Washington, DC: Teachers of English to Speakers of Other Languages.

Stoller, L. F., Anderson, N. J., Grabe, W., & Komiyama, R. (2013). Instructional Enhancements to Improve Students' Reading Abilities. English Teaching Forum, 1, 1-33.

Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research – Techniques and Procedures for Developing Grounded Theory.* (2nd). CA: Sage.

Urquhart, A. H. (1987). Comprehensions and Interpretations. *Reading in a Foreign Language, 3 (2), 387-409.*

Urquhart, A. H., & Weir, C. J. (1998). *Reading in a Second Language: Process, Product and Practice*. Harlow: Longman.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Villaume, S., & Brabham, E. (2002). Comprehension instruction: Beyond strategies. *The Reading Teacher, 55(7), 672-675.*

Wagner, J. (1993). Ignorance in educational research: or, how can you not know that? *Educational Researcher 22 (5),15-23.*

Weir, C. J., Hughes, A., & Porter, D. (1990). Reading Skills: Hierarchies, Implications Relationships and Identifiability. *Reading in a Foreign Language,7(1), 505-510.*

Weir, C. J., & Porter, D. (1994). The Multi-Divisible or Unitary Nature of Reading: The language tester between Scylla and Charybdis. *Reading in a Foreign Language, 10 (2), 1–19.*

Weir, C. J. (1995). *Understanding and developing language tests*. New York: Phoenix ELT.

Weir, C., & Khalifa, H. (2008). A cognitive processing approach towards defining reading comprehension. *Cambridge ESOL: Research Notes, 31, 1-10.*

Westwood, P. (2001). *Reading and learning difficulties: Approaches to teaching and assessment*. Camberwell, Victoria: ACER.

APPENDIX A

BLOOM TAXONOMY

(Bloom et al. 1956, p. 201)

Cognitive Domain

KNOWLEDGE

1. 00 KNOWLEDGE

Knowledge, as defined here, involves the recall of specifics and universals, the recall of methods and processes, or the recall of a pattern, structure, or setting. For measurement purposes, the recall situation involves little more than bringing to mind the appropriate material. Although some alteration of the material may be required, this is a relatively minor part of the task. The knowledge objectives emphasize most the psychological processes of remembering. The process of relating is also involved m that a knowledge test situation requires the organization and reorganization of a problem such that it will furnish the appropriate signals and cues for the information and knowledge the individual possesses. To use an analogy, if one thinks of the mind as a file, the problem in a knowledge test situation is that of finding in the problem or task the appropriate signals, cues, and clues which will most effectively bring out whatever' knowledge is filed or stored.

1.10 KNOWLEDGE OF SPECIFICS

The recall of specific and isolable bits of information. The emphasis is on symbols with concrete referents. This material, which is at a very low level of abstraction, may be thought of as the elements from which more complex and abstract forms of knowledge are built.

1.11 KNOWLEDGE OF TERMINOLOGY

Knowledge of the referents for specific symbols (verbal and non-verbal). This may include knowledge of the most generally accepted symbol referent, knowledge of the variety of symbols which may be used for a single referent, or knowledge of the referent most appropriate to a given use of a symbol.

- To define technical terms by giving their attributes, properties, or relations.
- Familiarity with a large number of words in their common range of meanings.

1.12 KNOWLEDGE OF SPECIFIC FACTS

Knowledge of dates, events, persons, places, etc. This may include very precise and specific information such as the specific date or exact magnitude of a phenomenon. It may also include approximate or relative information such as an approximate time period or the general order of magnitude of a phenomenon.

- The recall of major facts about particular cultures.
- The possession of a minimum knowledge about the organisms studied in the laboratory.

1. 20 KNOWLEDGE OF WAYS AND MEANS OF DEALING WITH SPECIFICS

Knowledge of the ways of organizing, studying, judging, and criticizing. This includes the methods of inquiry, the chronological sequences, and the standards of judgment within a field as well as the patterns of organization through which the areas of the fields themselves are determined and internally organized. This knowledge is at an intermediate level of abstraction between specific knowledge on the one hand and knowledge of universals on the other. It does not so much demand the activity of the student in using the materials as it does a more passive awareness of their nature.

1. 21 KNOWLEDGE OF CONVENTIONS

Knowledge of characteristic ways of treating and presenting ideas and phenomena. For purposes of communication and consistency, workers in a field employ usages, styles, practices, and forms which best suit their purposes and/or which appear to suit best the phenomena with which they deal. It should be recognized that although these forms and conventions are likely to be set up on arbitrary, accidental, or authoritative bases, they are retained because of the general agreement or concurrence of individuals concerned with the subject, phenomena, or problem.

- Familiarity with the forms and conventions of the major types of works, e.g., verse, plays, scientific papers, etc.
- To make pupils conscious of correct form and usage in speech and writing.

1. 22 KNOWLEDGE OF TRENDS AND SEQUENCES

Knowledge of the processes, directions, and movements of phenomena with respect to time.

- Understanding of the continuity and development of American culture as exemplified in American life.
- Knowledge of the basic trends underlying the development of public assistance programs.

1. 23 KNOWLEDGE OF CLASSIFICATIONS AND CATEGORIES
Knowledge of the classes, sets, divisions, and arrangements which are regarded as fundamental for a given subject field, purpose, argument, or problem.

- To recognize the area encompassed by various kinds of problems or materials.
- Becoming familiar with a range of types of literature.

1. 24 KNOWLEDGE OF CRITERIA
Knowledge of the criteria by which facts, principles, opinions, and conduct are tested or judged.

- Familiarity with criteria for judgment appropriate to the type of work and the purpose for which it is read.
- Knowledge of criteria for the evaluation of recreational activities.

1. 25 KNOWLEDGE OF METHODOLOGY
Knowledge of the methods of inquiry, techniques, and procedures employed in a particular subject field as well as those employed in investigating particular problems and phenomena. The emphasis here is on the individual's knowledge of the method rather than his ability to use the method.

- Knowledge of scientific methods for evaluating health concepts.
- The student shall know the methods of attack relevant to the kinds of problems of concern to the social sciences.

1. 30 KNOWLEDGE OF THE UNIVERSALS AND ABSTRACTIONS IN A FIELD
Knowledge of the major schemes and patterns by which phenomena and ideas are organized. These are the large structures, theories, and generalizations which dominate a subject field or which are quite generally used in studying phenomena or solving problems. These are at the highest levels of abstraction and complexity.

1.31 KNOWLEDGE OF PRINCIPLES AND GENERALIZATIONS
Knowledge of particular abstractions which summarize observations of phenomena. These are the abstractions which are of value in explaining, describing, predicting, or in determining the most appropriate and relevant action or direction to be taken.

- Knowledge of the important principles by which our experience with biological phenomena is summarized.
- The recall of major generalizations about particular cultures.

1. 32 KNOWLEDGE OF THEORIES AND STRUCTURES
Knowledge of the body of principles and generalizations together with their interrelations which present a clear, rounded, and systematic view of a complex phenomenon, problem, or field. These are the most abstract formulations, and they can be used to show the interrelation and organization of a great range of specifics.

- The recall of major theories about particular cultures.
- Knowledge of a relatively complete formulation of the theory of evolution.

INTELLECTUAL ABILITIES AND SKILLS
Abilities and skills refer to organized modes of operation and generalized techniques for dealing with materials and problems. The materials and problems may be of such a nature that little or no specialized and technical information is required. Such information as is required can be assumed to be part of the individual's general fund of knowledge. Other problems may require specialized and technical information at a rather high level such that specific knowledge and skill in dealing with the problem and the materials are required. The abilities and skills objectives emphasize the mental processes of organizing and reorganizing material to achieve a particular purpose. The materials may be given or remembered.

2.00 COMPREHENSION
This represents the lowest level of understanding. It refers to a type of understanding or apprehension such that the individual knows what is being communicated and can make use of the material or idea being communicated without necessarily relating it to other material or seeing its fullest implications.

2.10 TRANSLATION
Comprehension as evidenced by the care and accuracy with which the communication is paraphrased or rendered from one language or form of communication to another. Translation is judged on the basis of faithfulness and accuracy, that is, on the extent to which the material in the original communication is preserved although the form of the communication has been altered.

- The ability to understand non-literal statements (metaphor, symbolism, irony, exaggeration).
- Skill in translating mathematical verbal material into symbolic statements and vice versa.

2.20 INTERPRETATION
The explanation or summarization of a communication. Whereas translation involves an objective part-for-part rendering of a communication, interpretation involves a reordering, rearrangement, or a new view of the material.

- The ability to grasp the thought of the work as a whole at any desired level of generality.
- The ability to interpret various types of social data.

2.30 EXTRAPOLATION

The extension of trends or tendencies beyond the given data to determine implications, consequences, corollaries, effects, etc., which are in accordance with the conditions described in the original communication.

- The ability to deal with the conclusions of a work in terms of the immediate inference made from the explicit statements.
- Skill in predicting continuation of trends.

3.00 APPLICATION

The use of abstractions in particular and concrete situations. The abstractions may be in the form of general ideas, rules of procedures, or generalized methods. The abstractions may also be technical principles, ideas, and theories which must be remembered and applied.

- Application to the phenomena discussed in one paper of the scientific terms or concepts used in other papers.
- The ability to predict the probable effect of a change in a factor on a biological situation previously at equilibrium.

4.00 ANALYSIS

The breakdown of a communication into its constituent elements or parts such that the relative hierarchy of ideas is made clear and/or the relations between the ideas expressed are made explicit. Such analyses are intended to clarify the communication, to indicate how the communication is organized, and the way in which it manages to convey its effects, as well as its basis and arrangement.

4.10 ANALYSIS OF ELEMENTS

Identification of the elements included in a communication.

- The ability to recognize unstated assumptions.
- Skill in distinguishing facts from hypotheses.

4.20 ANALYSES OF RELATIONSHIPS

The connections and interactions between elements and parts of a communication.

- Ability to check the consistency of hypotheses with given information and assumptions.
- Skill in comprehending the interrelationships among the ideas in a passage.

4.30 ANALYSIS OF ORGANIZATIONAL PRINCIPLES

The organization, systematic arrangement, and structure which hold the communication together. This inc1udes the "explicit" as well as "implicit" structure. It includes the bases, necessary arrangement, and the mechanics which make the communication a unit.

- The ability to recognize form and pattern in literary or artistic works as a means of understanding their meaning.
- Ability to recognize the general techniques used in persuasive materials, such as advertising, propaganda, etc.

5.00 SYNTHESIS

The putting together of elements and parts so as to form a whole. This involves the process of working with pieces, parts, elements, etc., and arranging and combining them in such a way as to constitute a pattern or structure not clearly there before.

5.10 PRODUCTION OF A UNIQUE COMMUNICATION

The development of a communication in which the writer or speaker attempts to convey ideas, feelings, and/or experiences to others.

- Skill in writing, using an excellent organization of ideas and statements.
- Ability to tell a personal experience effectively.

5.20 PRODUCTION OF A PLAN, OR PROPOSED SET OF OPERATIONS

The development of a plan of work or the proposal of a plan of operations. The plan should satisfy requirements of the task which may be given to the student or which he may develop for himself.

- Ability to propose ways of testing hypotheses.
- Ability to plan a unit of instruction for a particular teaching situation.

5.30 DERIVATION OF A SET OF ABSTRACT RELATIONS

The development of a set of abstract relations either to classify or explain particular data or phenomena, or the deduction of propositions and relations from a set of basic propositions or symbolic representations.

- Ability to formulate appropriate hypotheses based upon an analysis of factors involved, and to modify such hypotheses in the light of new factors and considerations.
- Ability to make mathematical discoveries and generalizations.

6.00 EVALUATION
Judgments about the value of material and methods for given purposes. Quantitative and qualitative judgments about the extent to

which material and methods satisfy criteria. Use of a standard of appraisal. The criteria may be those determined by the student or those which are given to him.

6.10 JUDGMENTS IN TERMS OF INTERNAL EVIDENCE
Evaluation of the accuracy of a communication from such evidence as logical accuracy, consistency, and other internal criteria.

- Judging by internal standards, the ability to assess general probability of accuracy in reporting facts from the care given to exactness of statement, documentation, proof, etc.
- The ability to indicate logical fallacies in arguments.

6.20 JUDGMENTS IN TERMS OF EXTERNAL CRITERIA
Evaluation of material with reference to selected or remembered criteria.

- The comparison of major theories, generalizations, and facts about particular cultures.
- Judging by external standards, the ability to compare a work with the highest known standards in its field – especially with other works of recognized excellence.

APPENDIX B

LEVEL 1

KNOWLEDGE

LEVEL	CATEGORY	SUBCATEGORY
1.0 Knowledge Remember by recognition or recall.	**1.1 Knowledge of Specifics** Recall: Verbal or non Verbal symbols Low abstraction	**1.11 Knowledge of terminology** Essential terms. Literal.
		Define terms by attributes, properties or relations.
		Ability to distinguish referents for words from others.
		Familiarity with large number of words and meanings.
		Knowledge and mastery of vocabulary and terms of fine arts, science, accounting, geometry and quantitative thinking.
		1.12 Knowledge of Specific Facts Precise and Specific info, data, facts, words.
		Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.
		Acquaintance with significant names, places, events, authors, particular periods and events in the news.
	1.2 Knowledge of Ways and Means of dealing with Specifics Chronological sequence of patterns. Organization.	**1.21 Knowledge of Conventions** To treat or to present ideas or phenomena in a field.
		Familiarity with forms and conventions (verse, plays, scientific papers, maps, charts)
		Correct forms of speech and writing. Acceptable forms of language. Correct pronunciation, punctuation.
		1.22 Knowledge of the Trends and Sequences. Processes, directions, movements with respect of time.
		Understand continuity, trends, effects, developments (health, culture, evolution of man, history events, economics, politics, environment, administration, government)
		Know forces, past and present, and their world interdependence.
		1.23 Knowledge of Classifications and Categories. Classes, sets, divisions, and arrangements.
		Recognize a given area according to the problems or materials.
		Familiar with various types of literatures.
		1.24 Knowledge of Criteria Criteria to test or judge facts, principles, opinions or conduct.
		Familiarity and knowledge about criteria to judge a type of work and the purpose to read it.
		Criteria used to judge recreational activities, social sciences, a source of information, a work of art, nutrition, and economy incomes.
		1.25 Knowledge on Methodology. Methods of inquiry. Techniques. Procedures.
		To know methods to attack and evaluate problems (health, social sciences)
		To know techniques and methods to answer questions about the world (sciences)
	1.3 Knowledge of the Universals and abstractions in the Field Organization of ideas, schemes, and patterns. High abstraction and complexity.	**1.31Knowledge of Principles and Generalizations.** Particular abstractions which summarize the principle or generalization.
		Recall and understand generalizations and principles of biological phenomena, cultures, chemistry, civilizations, economy, learning, evolution, history.
		Develop knowledge and understanding of principles, functions, and elements.
		1.32 Knowledge of Theories and Structures. Clear and systematic interrelated principles and generalizations of a complex phenomenon that form a theory or structure
		Recall and recognition of major theories about cultures, philosophic bases, chemical principles, evolution, congress structure, and government.

APPENDIX B

LEVEL 2

COMPREHENSION

LEVEL	CATEGORY	SUBCATEGORY
2.0 Comprehension Objectives, behaviors, or responses which represent an understanding of the literal message. Extensions beyond communication itself.	**2.1 Translation** Transform an abstract idea into everyday terms to facilitate thinking.	**Translation from one level of abstraction to another** ***Ability to translate:***
		-Abstract into concrete or less abstract. "in your own words"
		-Lengthy communication into briefer or less abstract terms.
		-An abstraction by giving an illustration or sample.
		Translation from symbolic form to another form or vice versa
		-Ability to translate relationships expressed in symbolic form to verbal form and vice versa.
		-Ability to prepare graphics of data into visuals.
		Translation from one verbal form to another
		-Ability to translate non-literal statements (metaphor, symbolism, irony, exaggeration) to ordinary English.
		-Ability to comprehend the significance of words in their context.
		-Ability to translate (with or without a dictionary) foreign language prose or poetry into good English.
	2.2 Interpretation Recognize essentials and differentiate irrelevant aspects. Analysis. Beyond mere repetition and rephrasing.	**Interpretation.**
		-Ability to grasp the thought of a work as a whole at any level of generality.
		-Ability to comprehend and interpret with depth and clarity.
		-Ability to distinguish warranted, unwarranted or contradicted conclusions.
		-Ability to interpret and make proper qualifications.
	2.3 Extrapolation Determine the relations between a sample and a universe and vice versa. Extend the trend beyond the given data. Make inferences of probability.	**Extrapolation.**
		-Ability to deal with conclusions in terms of inferences.
		-Ability to draw and state conclusions.
		-Predict trends
		-Interpolate where there are gaps in data.
		-Ability to estimate or predict consequences.
		-Ability to sensitize to render predictions when inaccurate.
		-Ability to distinguish consequences with a high degree of probability.
		-Ability to differentiate value judgments from predictions of consequences.

APPENDIX C

CELE EXAMS (READING SECTION)

Table 1. Description of CELE Reading Abilities					
Spring 2012					
	Ability	Format and item example	Number of items	% of the recurrence of the ability by SECTION	% of the recurrence of the ability per LEVEL
Level 1 Knowledge			0	0	0
Level 2 Comprehension	-Abstract into concrete or less abstract. "in your own words"	The writer thinks that hiding a key under a doormat or flower pot : (A) abcd (B) abcd (C) abcd	4	40%	40%
	-Ability to grasp the thought of a work as a whole at any level of generality.	The best title for the text is (A) abcd (B) abcd (C) abcd	1	10%	10%
	-Interpolate where there are gaps in data.	Matching exercise, sentences have to be placed back where they belong	5	50%	50%
		Total of items	10		

APPENDIX C

CELE EXAMS (CONT'D)

Table 2. Description of CELE Reading Abilities Summer 2012					
	Ability	Format and item example	Number of items	% of the recurrence of the ability by SECTION	% of the recurrence of the ability per LEVEL
Level 1 Knowledge	Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.	**According to the author** which entities must work collectively to improve computer viruses legislation Cottages? (A) abcd (B) abcd (C) abcd	1	10%	100%
Level 2 Comprehension	-Abstract into concrete or less abstract. "in your own words"	**According to the text** what it is considered as a computer cybercrime? (A) abcd (B) abcd (C) abcd	1	10%	11%
	-Ability to translate non-literal statements (metaphor, symbolism, irony, exaggeration) to ordinary English.	What does Bruce McConell **imply** when states:"organizations must rely on their own defenses for now" (A) abcd (B) abcd (C) abcd	1	10%	11%
	-Ability to comprehend the significance of words in their context.	The paragraph , the line `thus, since federal legislation…, the word **loose** means: (A) abcd (B) abcd (C) abcd	1	10%	11%
	-Ability to grasp the thought of a work as a whole at any level of generality.	Which is the **theme** of the text (A) abcd (B) abcd (C) abcd	2	20%	22%
	-Ability to comprehend and interpret with depth and clarity.	**Why** is the creation or distribution of computer viruses and international issue? (A) abcd (B) abcd (C) abcd	2	20%	22%
	-Ability to deal with conclusions in terms of inferences.	**In which case**, distribution of computer viruses will probably **NOT** face criminal charges (A) abcd (B) abcd (C) abcd	1	10%	11%
	-Ability to draw and state conclusions.	**Based on the text**, what can it be said that any person who distributes viruses is a criminal? (A) abcd (B) abcd (C) abcd	1	10%	11%
		Total of items	10		

APPENDIX C

CELE EXAMS (CONT'D)

Table 3. Description of CELE Reading Abilities					
Fall 2012					
	Ability	Format and item example	Number of items	% of the recurrence of the ability by SECTION	% of the recurrence of the ability per LEVEL
Level 1 Knowledge	Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.	**According to the text** what sciences are drawn by Marketing (A) abcd (B) abcd (C) abcd	2	20%	100%
Level 2 Comprehension	-Ability to comprehend the significance of words in their context.	In line 2 the **word** "Holy Grail" **means** (A) abcd (B) abcd (C) abcd	2	20%	25%
	-Ability to grasp the thought of a work as a whole at any level of generality.	What is the idea behind the concept of branding? (A) abcd (B) abcd (C) abcd	4	40%	50%
	-Ability to deal with conclusions in terms of inferences.	**Based on the text** can it be stated that the medial prefrontal cortex completely defines who we are (B) abcd (C) abcd (D) abcd	1	10%	12.5%
	-Ability to draw and state conclusions.	What is the author's main intention? (A) abcd (B) abcd (C) abcd	1	10%	12.5%
		Total of items	10		

APPENDIX C

CELE EXAMS (CONT'D)

Table 4. Description of CELE Reading Abilities					
Spring 2013					
	Ability	Format and item example	Number of items	% of the recurrence of the ability by SECTION	% of the recurrence of the ability per LEVEL
Level 1 Knowledge	Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.	True and false statements	5	50%	100%
Level 2 Comprehension	-Interpolate where there are gaps in data.	Matching exercise, sentences have to be placed back where they belong	5	50%	100%
		Total of items	10		

Table 5. Description of CELE Reading Abilities					
Summer 2013					
	Ability	Format and item example	Number of items	% of the recurrence of the ability by SECTION	% of the recurrence of the ability per LEVEL
Level 1 Knowledge	Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.	Qigong is perfect for those who (A) abcd (B) abcd (C) abcd (D) abcd	5	100%	100%
Level 2 Comprehension			0	0	0
		Total of items	5		

APPENDIX C

CELE EXAMS (CONT'D)

Table 6. Description of CELE Reading Abilities					
Fall 2013					
	Ability	Format and item example	Number of items	% of the recurrence of the ability by SECTION	% of the recurrence of the ability per LEVEL
Level 1 Knowledge	Ability to distinguish referents for words from others.	The word **"their"** in line 12 **refers** to (A) abcd (B) abcd (C) abcd (D) abcd	1	10%	20%
	Knowledge and mastery of vocabulary and terms of fine arts, science, accounting, geometry and quantitative thinking.	**According to the passage** why did O'Keeffe destroy her work in 1915 (A) abcd (B) abcd (C) abcd (D) abcd	1	10%	20%
		True and false staments	2	20%	40%
	Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.	Open question	1	10%	20%
Level 2 Comprehension	-Ability to translate non-literal statements (metaphor, symbolism, irony, exaggeration) to ordinary English.	Open question	1	10%	20%
	-Ability to comprehend the significance of words in their context.	The **word** "core" in line 1 is **closest in meaning to** (A) abcd (B) abcd (C) abcd (D) abcd	1	10%	20%
	-Interpolate where there are gaps in data.	Matching exercise, sentences have to be placed back where they belong	3	30%	60%
		Total of items	10		

APPENDIX D

TOEFL and CELE

Reading Abilities

TOEFL	PBT	Spring 2012	Summer 2012	Fall 2012	Spring 2013	Summer 2013	Fall 2013	CELE
Ability to distinguish referents for words from others.	4%						10%	Ability to distinguish referents for words from others.
Knowledge and mastery of vocabulary and terms of fine arts, science, accounting, geometry and quantitative thinking.	8%						10%	Knowledge and mastery of vocabulary and terms of fine arts, science, accounting, geometry and quantitative thinking.
Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.	32%		10%	20%	50%	100%	20%	Recognition and Recall of facts about cultures, biology chemistry, and physics processes, society, health, natural resources, nutrition.
XXXXXXXXXXX		40%	10%					-Abstract into concrete or less abstract. "in your own words"
-Ability to translate non-literal statements (metaphor, symbolism, irony, exaggeration) to ordinary English.	6%		10%				10%	-Ability to translate non-literal statements (metaphor, symbolism, irony, exaggeration) to ordinary English.
-Ability to comprehend the significance of words in their context.	28%		10%	20%			10%	-Ability to comprehend the significance of words in their context.
-Ability to grasp the thought of a work as a whole at any level of generality.	10%	10%	20%	40%				-Ability to grasp the thought of a work as a whole at any level of generality.
XXXXXXXXXXX			20%					-Ability to comprehend and interpret with depth and clarity.
-Ability to deal with conclusions in terms of inferences.	12%		10%	10%				-Ability to deal with conclusions in terms of inferences.
XXXXXXXXXXX			10%	10%				-Ability to draw and state conclusions.
XXXXXXXXXXX		50%			50%		30%	-Interpolate where there are gaps in data.