Receptive and productive vocabulary profiles of high school students in Content Based and Foreign Language Instruction: A Costa Rican perspective

UNIVERSIDAD DE SALAMANCA

FACULTAD DE FILOLOGÍA

DEPARTAMENTO DE FILOLOGÍA INGLESA

TESIS DOCTORAL

Receptive and productive vocabulary profiles of high school students in Content Based and Foreign Language Instruction: A Costa Rican perspective

DAMARIS CASTRO GARCÍA

DIRECTOR

RAMIRO DURÁN MARTÍNEZ

2018

Tesis doctoral presentada por Dña. Damaris Castro García bajo la dirección del Prof. Ramiro Durán Martínez.

VºBº

Prof. Ramiro Durán Martínez

Salamanca

2018

AGRADECIMIENTOS

Primero quisiera agradecer al profesor Ramiro Durán Martínez por su apoyo constante a lo largo de esta etapa de estudios doctorales. He sido muy afortunada de haber trabajado bajo su supervisión; agradezco su guía, apoyo y motivación, así como el interés que ha mostrado hacia mi investigación y mi formación profesional. Quisiera también agradecer a todo el personal del Departamento de Filología Inglesa de la Universidad de Salamanca por su contribución en este proceso. Hago extensivo mi sincero agradecimiento a la Universidad Nacional de Costa Rica y en particular a la Escuela de Literatura y Ciencias del Lenguaje, sin su financiamiento mis estudios doctorales no hubieran sido posibles.

Quiero expresar mi más profundo agradecimiento a Fran, por su apoyo infinito y su paciencia durante este proceso. Sin su ayuda no hubiera podido terminar este proyecto exitosamente. Espero poder seguir aprendiendo de él y poder así algún día imitar su carácter, integridad, y humildad. Muy especialmente quiero dedicar este trabajo a papi y a mami. No hay palabras que puedan expresar el agradecimiento que siento hacia ellos; por enseñarme a ser persistente, a esforzarme, y a dar siempre lo mejor de mí. Gracias por todo lo que me han enseñado. En general quiero también agradecer a toda mi familia, por el apoyo permanente y por ayudarme a seguir mis sueños y alcanzar mis metas. Con su ejemplo, todos me han enseñado el sentido de responsabilidad, honestidad, y perseverancia que guían mis acciones cada día. Por esto y todo lo que me han dado les estaré siempre agradecida. Espero tener muchas oportunidades para recuperar juntos un poco del tiempo que este proyecto nos ha quitado.

Finalmente, quisiera expresar mi agradecimiento y admiración más profunda a Sherry Gapper. Gracias por su apoyo y guía durante toda mi vida académica. Durante los últimos casi veinte años Sherry ha sido mi profesora, mi amiga, mi compañera de trabajo, mi consejera, y sobre todo un modelo del tipo de profesional y persona que yo quisiera ser. La ética, el trabajo constante, la determinación, la humildad, y la increíble generosidad que ella pone en cada proyecto del que forma parte han sido una inspiración para mi desde hace mucho tiempo. Una vez más, todas esas cualidades han estado presentes durante el proceso de lectura y revisión de esta tesis. Hago extensivo mi agradecimiento a su familia, por su amable y desinteresada ayuda.

Gracias a todos, este trabajo es para ustedes.

TABLE OF CONTENTS

INDEX OF TABLES

INDEX OF FIGURES

LIST OF ABBREVIATIONS

AD	Attention Deficit
ADD	Attention Deficit Disorder
AL	Applied Linguistics
AWL	Academic Word List
BICS	Basic Interpersonal Communication Skills
CALP	Cognitive Academic Language Proficiency
CB	Content Based
CBI	Content Based Instruction
CBT	Content Based Teaching
CEFR	Common European Frame of Reference
CLIL	Content and Language Integrated Learning
CPF	Canadian Parents for French
CUP	Common Underlying Proficiency
DLE	Dual Language Education
DVK	Depth of Vocabulary Knowledge
EAL	English Additional Language
EFI	Early French Immersion
EFL	English as a Foreign Language
ELL	English Language Learners
ESL	English as a Second Language
ESOL	English for Speakers of Other Language
EU	European Union
FI	French Immersion
FL	Foreign Language
FLI	Foreign Language Instruction
FLS	Foreign Language School
FLT	Foreign Language Teaching
FSL	French as Second Language
HD	Hyperactivity Disorder
IELTS	International English Language Testing System
IH	Interdependence Hypothesis
L1	First Language
L2	Second Language
L3	Third Language

LFI	Late French Immersion
LFP	Lexical Frequency Profile
LOTE	Language Other Than English
MEP	*Ministerio de Educación Pública* (Costa Rican Board of Education)
MFI	Middle French Immersion
PBES	Public Bilingual Experimental School
PVLT	Productive Vocabulary Levels Test
SES	Socio Economical Status
SD	Standard Deviation
SIOP	Sheltered Instruction Observation Protocol
SL	Second Language
SLA	Second Language Acquisition
SUP	Separate Underlying Proficiency
TBLT	Task Based Language Teaching
TL	Target Language
TOEIC	Test of English for International Communication
TOEFL	Test of English as Foreign Language
TOEFL-RBC	Test of English as a Foreign Language- Reading for Basic Comprehension
TPR	Total Physical Response
UWL	University Word List
VIM	Vocabulary Item Measure
VKS	Vocabulary Knowledge Scale
VLT	Vocabulary Levels Test
VS(T)	Vocabulary Size (Test)
WRMT-PC	Woodcock Mastery Test Revised- Passage Comprehension

INTRODUCTION

OVERVIEW OF THE STUDY

Across the globe, different countries and regions are coming up with a number of ideas to deal with globalization and what it entails in terms of communication and languages. There is a growing interest in individuals attaining functional bilingualism with the purpose of dealing with the demands of an ever more globalized society where citizens mobility and job requirements require bilingual speakers. The Costa Rican context is no different in this sense. There is a clear interest from the citizens to learn English and the government is also directing its efforts and resources to doing just that. These attempts are, however, still limited in action. While there are many plans and reforms on the way, these require major changes in the current public system for them to be fully implemented. At the moment, the public sector offers foreign language teaching, where English is taught as a subject of the curriculum, with very poor to limited results in terms of L2 attainment. As a result, many parents rely on the private sector to offer their children a more adequate access to bilingual education. Henceforth, bilingual education will be understood as the teaching of content through means of a language different from the L1 of the learners. The aim of bilingual education is often associated with the development and strengthening of L2 competence, rather than bilingualism per se.

The intention of this study is to conduct comparative research amongst two groups of learners from two institutions that follow different teaching practices. One of the institutions is a semi-private high school, partly founded by the state, where English is taught following traditional Foreign Language Teaching (FLT) and in which English is seen as one more subject in the curriculum. The other school, a private institution, offers Content Based Teaching (CBT) through which some of the subjects are taught using English as the means of instruction. Aside from that, in this setting, students receive English lessons. This comparative study attempts to provide evidence regarding the actual levels of vocabulary attainment that students in both institutions reach. This will offer a general panorama of what might be taking place at the level of secondary education regarding English learning. Through the analysis of current trends of bilingualism and bilingual education as well as through the study of proposals of key players in the field of Second Language Acquisition (SLA) and exemplary models of second language teaching, we will try to identify elements that may contribute to the achievement of better levels of second language realization in the Costa Rican educational system. With the intention of providing information in a context where scientific studies are

scarcely found, this study attempts to serve as a starting point for a long-overdue research field in SLA in the Costa Rican educational setting. With this idea in mind, we will now move to present the details of this study.

PROPOSAL

The following paragraphs present a description of the main elements that conform the present study. Information is provided in terms of the participants of this study, the objectives pursued with this investigation, the methodology that is implemented and the different instruments that were administered to collect the data that informs this study.

Subjects

The study was conducted as a classroom research project and all 185 participants enrolled in five intact classes took part in the tests. Hence, the sample was not randomized. The participants in this study are 185 fifth-year secondary students who are learners of English as a second language in two high schools in the central province of Heredia, Costa Rica. We may anticipate that the total number of students may slightly vary between tests, as some students may be absent from class in the day a particular test is administered. This sample is homogeneous in terms of L1, age, social class and geographical location. Spanish is the first language for all students. None of the parents of the students are native speakers of English nor have the students lived in an English-speaking country. The average age of the students is 16.6 years old. Both groups attend a middle-class school in the same geographical area in the province of Heredia, Costa Rica, the third most populated province in the country. Both schools follow the same national language policy in terms of the curricular content as established by the National Board of Education, aside from the additional content teaching in the school that follows content based instruction. Currently, the main aim of the national language curricula is to help students develop communicative skills although in the classroom it still concentrates on a strong development of the reading skill. This is made evident in the objectives of the national curriculum that demands more complex development of the communicative competence for the reading skill while

these appear to be more basic for writing and even more so for listening and speaking (see *Ministerio de Educación*, 2016, 2017).

One of the schools, henceforth referred to as Content Based School (CBS), implements Content Based Teaching (CBT) as part of its methodology. In the sample, 55 students attend this private, bilingual high school. There are 21 boys and 34 girls in this sample. The average age for students is 16.3 years old. In this school, students receive 6 weekly hours of English language class where English is studied as a subject. Additionally, they attend 4 weekly hours of a content course where the subject matter is taught in English. Regarding the total number of hours of instruction, the CB students have received approximately 1,368 hours of instruction at the secondary level. These students received 361 hours of instruction per year during 7th, 8th and 9th level plus 285 hours of instruction at 10th level. Part of that total number of hours corresponds to 3 hours of content-based instruction per week. At this school, the subject taught in English varies from one level to the next. These include Ecology, Social Studies and Biology in 7th, 8th, and 9th, grade respectively. In 10th year they take an English taught preparatory course for the TOEIC test. The data for this study were collected during the first month of the 11th grade, their last year of secondary instruction.

This high school offers optional, non-permanent theater and reading English clubs. Learners also have the possibility of participating in student exchanges with high schools from the U.S.A. and England. This institution hosts students from the U.S.A and Germany for year-long stays. Every June, the English Festival is held. In preparation for this activity, students spend a month making videos, magazines, journals, and comic books. A jury selects the best pieces and assigns a prize to the winners.

The second school will be labeled Foreign Language School (FLS) and, as suggested by its name implements Foreign Language Teaching (FLT) where English is studied as part of mainstream education, as an additional subject. This group of 130 students represents the rest of the participants in the study. The average age for students in this school is 16.5 years old. There are 64 boys and 66 girls. These students receive compulsory English lessons, 6 hours of English as a Foreign Language (EFL) classes per week. These students will have received a total number of 1140 hours of instruction by the time the tests are administered. These hours are distributed as 304 hours of English instruction during 7th, 8th and 9th level, plus 228 hours of English instruction during 10th grade. Like in the other school, the data for this study were collected during the first month of the 11th grade level.

According to the English coordinator of this school, students enjoy a number of activities that allow them to be in contact with native speakers and advanced second-language speakers of English. For example, students in the last three years of high school serve as volunteers translating for medical doctors who work for a project called *Butterfly Clinic*, a project that works in marginal areas of the central cities and for which students translate and interpret information for doctors and patients. Pupils can participate of student exchange programs attending high schools in the U.S.A. as well as hosting students from the U.S.A., France, Germany, Switzerland, and Austria. This school also offers support in the form of additional tutoring lessons for incoming students who are lagging behind in English classes. Another course is offered to 10th and 11th graders in preparation for proficiency tests such as TOEIC, TOEFL, and IELTS.

Objectives

The idea of this study is based on the premise that language learning through content results in greater benefits for students. We expect the positive effects of language instruction to be made evident in the vocabulary levels that students attain. With the intention of corroborating this premise and determining the tangible effects of different types of language instruction, this investigation will pursue the following objectives:

1. To determine the receptive vocabulary levels of students attending a school that implements FLT.
2. To determine the receptive vocabulary levels of students attending a school that implements CBT.
3. To determine the productive vocabulary levels of students attending a school that implements FLT.
4. To determine the productive vocabulary levels of students attending a school that implements CBT.
5. To define whether there are significant differences in how productive and receptive vocabulary develop in different teaching language conditions.
6. To establish the effects of receptive vocabulary on the reading ability of the students in the sample
7. To establish the effects of productive vocabulary on the writing ability of the students in the sample.

8. To describe the effects that the motivation of the students exerts in second language learning.
9. To define whether gender characteristics affect the acquisition of vocabulary in this population.

Methodology

The students will take a battery of tests to measure the vocabulary levels that they have attained. To determine the students' receptive vocabulary level, they will complete the Vocabulary Levels Test (VLT) (Nation, 1983; Schmitt, Schmitt, and Clapham, 2001). This test will measure vocabulary levels through a controlled task. Also at the receptive level, the students will complete a reading test, Cambridge ESOL Skills for Life, Entry 3 (2011), where the receptive vocabulary will be tested in context. Moreover, to determine productive vocabulary levels in a controlled test, students will take the Productive Vocabulary Levels Test (PVLT), (Laufer and Nation, 1999). At the productive level but in a free writing task, the students will write a composition where they can choose among three topics to present their ideas. Along the language tests, the students will complete a personal survey as well as a motivation survey.

The students will take the different tests and complete the different instruments in different sessions. While two of the short tests or surveys (VLT, PVLT, personal and motivational survey) may sometimes be completed in one session, the reading and writing tasks will be completed in single sessions dedicated to each one of those tasks. This is done with the idea that tiredness or lack of interest will not affect the results. The different groups in these schools will take all of the tests along a period of two weeks. All of the tests will be administered during the scheduled time for the English lessons. The students will not receive credit for taking these tests, regardless of their performance on them. Students will be made aware that they are taking the tests to contribute to research in SLA, although the idea of vocabulary measures will not be discussed in depth. Students will complete a consent form and they will be guaranteed anonymity throughout the process.

Following a descriptive approach, this study will inspect the data that has been collected through a quantitative analysis of the results obtained in the different tests taken by the sample population. To calculate descriptive values and compare the different measures of receptive and productive vocabulary, the SPSS 20

Software tool will be used to perform descriptive and inferential statistics. The media and standard deviation (SD), as well as Mann-Whitney U, Anova and T-student tests, among others, will be used to determine the type of statistical differences in the results.

Instruments

As mentioned above, students in this study will complete a battery of tests to measure the vocabulary levels that they have reached as a result of English language instruction in secondary education. They will complete two additional questionnaires, one of which will provide personal and socio-economical information about the subjects, and the other will gather information about the motivation that students have in connection to second language acquisition. Each of these instruments is described below and can also be found in the appendices at the end of this document.

Vocabulary Levels Test (VLT)

Paul Nation designed the Vocabulary Levels Test (VLT) in 1983 and updated it in 1990. It was originally conceived as a diagnostic test that would help determine the vocabulary size of students and would subsequently guide instructors and students regarding the teaching and learning activities that they needed to put into practice to achieve specific vocabulary levels. The test, however, has been widely used as a tool to establish measurements of the vocabulary size of students in many different studies with the intention of extending this information to describe the language proficiency of students (see Laufer, 1998). In the present study, an updated version of the test is used, in particular the VLT 2000 band, Version 2 of Schmitt et al. (2001). This test has been used in many different studies to determine the vocabulary profile of students (see Jiménez Catalán and Terrazas Gallego, 2005; Fernández Fontecha and Canga Alonso, 2014; Agustín Llach, 2015; among others). In this test the participants must complete a total of 30 items. For each of these items, they match a word to its corresponding definition. There are 30 extra words for which no definition is provided. The students, therefore, must discriminate between these distractors and the concepts that match the definitions to give the correct answers.

For Schmitt et al. (2001), in this new version of the test, the 2000 level contains words from the first 1000 words and from the second 1000 words in a ratio close to 1:1, specifically 28 words belong to the first 1000 words, and 32 belong to the second 1000 words of the language. Research has demonstrated the VLT to be a valid test (Beglar and Hunt, 1999; Cameron, 2002; Laufer, 1998; and Schmitt et al., 2001). The test has also been found to be reliable (Read, 1988; Schmitt et al., 2001), as well as practical and easy to implement and evaluate. Students will be provided with the following sample so that they know exactly what they are supposed to do. The directions are provided in Spanish and the test itself will be explained to the students in Spanish to ensure the correct understanding of the task.

Vea el siguiente ejemplo:

1. business
2. clock _____ part of the house
3. horse _____ animal with four legs
4. pencil _____ something used for writing
5. shoe
6. wall

Usted debe responder de la siguiente forma:

1. business
2. clock __6__ part of the house
3. horse __3__ animal with four legs
4. pencil __4__ something used for writing
5. shoe
6. wall

The total number of words is calculated following Nation's (1990, p. 78) formula, "Vocabulary size = N correct answers multiplied by the N words in dictionary (the relevant word list) divided by N items in the test." Schmitt et al. (2001) speak of a criterion of mastery of 26 out of 30 possible items per level as necessary to determine that a level is actually mastered by the students. The complete test can be found in Appendix 1.

Reading Test, ESOL Skills for Life (Entry 3)

To determine the vocabulary knowledge of the students as it is used in context, students will take the test *Cambridge ESOL Entry 3 Certificate in ESOL Skills for Life* (2011). This test is a subset of a battery of tests used worldwide to determine language skills in students of English as a Second or other Language and it requires 1:15 minutes for students to complete it. It contains five texts that range between 176 and 440 words in total. The students must answer a total of 32 questions; in 22 of these questions students have to check (√) the correct answer and in the remaining 10 questions they have to provide one-word answers in 9 of these questions and a one-sentence answer in one of them. The sample test is found in Appendix 2.

Productive Vocabulary Levels Test

A preliminary version of this test was first presented by Laufer and Nation (1995) as part of a larger study that showed a connection between vocabulary size and writing ability and it was described as the active version of the VLT. The test used in the present study is an updated version presented by Laufer and Nation (1999); in the present study, in particular we use the 2000 word level test. It consists of 18 items where students must complete one word that appears incomplete in the context of a sentence. The first letters of the word are provided to direct students to the right choice. Guided by the context and the part of the word that is provided, students complete the sentence with the word that better fits the context. The number of letters given to complete the required word depends on how many of these are necessary to disambiguate the correct word and to discard possible alternatives that belong to other word levels. Laufer and Nation (1999) argue that this test measures vocabulary knowledge in a controlled productive activity whereby students are required to use certain words,

> when compelled to do so by a teacher or researcher, whether in an unconstrained context such as a sentence-writing task, or in a constrained context such as a fill-in task where a sentence context is provided and the missing target word has to be supplied. (p. 37)

The latter context better describes the format of the test used in the present study. In the calculation of the scores, the authors suggest ignoring minor spelling and grammatical mistakes. They argue that 15 or 16 out of the 18 items should be correct to consider that students readily handle the vocabulary level, but they clarify that this depends on what researchers or teachers judge adequate for their purposes. In this version of the test, Laufer and Nation (1999) established that the test is reliable, valid and highly practical. The version of the PVLT that was used in the present study can be found in Appendix 3.

Writing task

Students will also complete a free writing task in which they can choose between three topics to present their ideas. They will be given general directions in Spanish and the topics will be presented in English. The first option appeals to students describing Costa Rica and the places and things it has to offer. The other two options in this task were taken from Laufer and Nation (1995). In these two cases, comments are added to encourage students to present their ideas. Although students will not be asked to produce a specific number of words, they will be encouraged to write as much as possible. Students will be given one hour to perform the task. The test sample is found in Appendix 4.

The students can choose from the following topics:

a. What could a foreigner find and enjoy if he visited Costa Rica? Discuss the things and places that Costa Rica has to offer to visitors.
b. “Should a government be allowed to limit the number of children a family can have?” Discuss this idea considering basic human rights and the danger of population explosion. (Laufer and Nation, 1995)
c. “A person cannot be poor and happy, because money is always needed to gain something that is important to that person.” Argue for and against that idea. (Laufer and Nation, 1995)

Personal and background questionnaire

In this instrument, students give general information about their language background, parents’ jobs, and general family information. They will also present

personal information in terms of school type, age, sex, use of English, Internet accessibility, self-evaluation of English level and other personal information. This information in combination with different portions of the test results are used throughout the analysis section in the form of variables to present specific results in this study. The questionnaire is found in Appendix 5.

Motivation survey

Finally, students will complete a survey that explores their motivation in regard to learning English. This survey is presented in the form of a Likert-Scale with five pre-coded responses where 3 is the neutral point, 5 represents students who agree strongly with the given statement and 1 represents those who disagree strongly with the statement. This survey was adapted from Gardner (2004) who created it to use it in a study to determine motivation toward learning of English as a second language. In the present study the survey is given in Spanish. It gathers information on five aspects with a total of 28 items. Eight items track motivation intensity and eight items are intended to identify the students' desire to learn English; four items attempt to determine the integrative orientation of students; four items look into the instrumental orientation of students and four more items evaluate the students' attitude toward learning English. Students will be given 40 minutes to complete the survey. See the motivation survey in Appendix 6.

STRUCTURE OF THE STUDY

This study is divided into two main parts. The first part discusses fundamental theories on key aspects for bilingualism and second language acquisition. First, it discusses bilingualism, its founding concepts, characteristics and advantages. Second, bilingual education is discussed as a phenomenon whose intervening factors, definition, approaches, and types of programs exert an enormous effect on bilingualism. To support the concept of bilingual education and describe successful types of programs, three sections are dedicated to strong forms of bilingual education: Immersion Programs in Canada, Dual Language Education and CLIL. Concerning bilingual education, the case of Costa Rica is presented with the intention of depicting the current situation of bilingual education in this country; this sets the context for this study. Theory about vocabulary is added and its critical

role in the acquisition of second languages is delineated. Finally, reference to current theory about gender and motivation in connection to second language acquisition is made with the intention of illustrating their role in the process.

The second part of this thesis concentrates on the analysis of the data collected in this study. The results for each one of the tests are examined in detail. First, a complete analysis of the results of the VLT is discussed. Second, an analysis of the results of the reading comprehension test is presented. This analysis is extended to establish the relation between the size of the vocabulary of students based on the results of the VLT and the scores obtained in the reading comprehension test. Then, the results of the PVLT and those of the writing task are analyzed independently. In addition, further analysis is conducted to establish the relation existing between the size of students' productive vocabulary (according to the PVLT) and its possible effect on the writing ability of the students in the sample. For each one of the tests, the role of gender is analyzed at inter and intra school levels. The analysis section concludes by analyzing the role of motivation in each school. For each test, the results of the present study are compared with earlier studies in this same field. Finally, a set of conclusions is drawn from the results and the theory presented throughout this study.

PART I: THEORETICAL REVIEW

CHAPTER 1

BILINGUALISM AND BILINGUAL EDUCATION

In this chapter, we will begin the theoretical analysis of the fundamental concepts that have set the basis for research on bilingualism and bilingual education in the context of second language acquisition. The first part of the chapter concentrates on key theoretical concepts and presents these concepts from the perspective of authors such as J. Cummins, S. Krashen or M. Swain, who have played an essential role in shaping and guiding research in the area for many decades. The first section then continues with a review of specific characteristics of bilingualism; it attempts to construct a definition of the term while acknowledging the difficulty of this challenge, and it concludes by presenting some of the advantages that have been associated with bilingualism.

In the second part of this chapter we deal with the notion of bilingual education. We discuss the role of some of the intervening factors which play a role in the process and which are not necessarily found explicitly in the second language classroom. Attention is given to the challenge of defining bilingual education and the different forms it can take through different approaches and types of instruction in the field. Finally, the two types of bilingual instruction compared in this study (Content Based Instruction and Foreign Language Instruction) are described. This second part of the chapter sets the basis for the discussion of three forms of bilingual education that are presented in Chapter 2.

1.1. BILINGUALISM

Since ancient times, bilingualism has been one of the products of contact between different languages and it has enabled societies to interact and go through processes of change and growth. Like in today's society, diglossia, pidginization and language loss, as well as language change and death were common results of language contact in the past. However, bilingualism and second language acquisition are the most commonly studied phenomena coming from language contact. Bilingualism may be due to personal choice, immigration, or socio-political reasons. Regardless of the cause, written evidence is proof of the existence of ancient bilingual texts exemplifying the social, cultural, and political effects of said language contact. Adams & Swain (2010) offer a review of many ancient languages and of how their contact led to changes in language and language use amongst the peoples in the areas; a few examples follow.

Based on written texts, Adams & Swain (2010) describe several cases dating back to conquest times, by the mid-first century BC, when Celtic, for example, ceased to represent the script form for Gallo, Greek, and Iberian languages, because of contact with Latin, a higher prestige language. After contact with Latin took place, Latin replaced Celtic in these contexts. Contact between Greek and Aramaic, Greek and Phrygian or Greek and Latin during the second and first centuries BC also resulted in bilingual texts. Syriac, a new variety of Aramaic appeared, as well, as a result of the dominance of Greek over Aramaic after the first and second centuries AD. For these authors, even though it is not possible to determine the extent of bilingualism back in that day, there is evidence that in Egypt and Greece, middle-ranking workers were usually balanced bilinguals, an important requirement that allowed them to perform their duties accordingly. Furthermore, according to Adams & Swain (2010), from peoples who held on to more than one language in the Mediterranean, to merchants and businesspeople who enjoyed the benefits of bilingualism, to Cicero and other members of the Roman nobility not only being bilingual but also biliterate in Greek, bilingualism was a remarkable feature already present in ancient times.

In modern times, bilingualism, as a field of study, appears to have traditionally served a three-fold purpose. According to Wei (2008), three main perspectives have guided research on bilingualism. The first area of research follows a *linguistic* orientation in which researchers are interested in determining the interaction of languages in a multilingual mind and how elements like *code-switching* come into play. For Wei (2008), this area of study deals with how speakers acquire *linguistic knowledge* and how they make use of this knowledge. The second area of research involves a *psycholinguistic* perspective that studies the mechanisms involved in reception and production of language; as much in reference to accessibility and use of vocabulary, as to the knowledge, organization, and use of linguistic information. The third area of research deals with *sociolinguistic* studies, where bilingualism and multilingualism are analyzed as a social phenomenon about how people use languages and choose to interact with different individuals in various situations to the extent of demonstrating their identity to others.

The present study could be placed within a combination of the first and second approaches, although some information could be drawn in reference to the third. On one hand, it will concentrate on analyzing the linguistic knowledge of students on the basis of vocabulary knowledge. As stated in Nation (2013), knowing vocabulary entails knowing a wide array of linguistic aspects that make

up a word. These include components that carry grammatical information as well as information about the form and meaning of the word. The second approach relies on the assumption that students are acquiring linguistic knowledge so that they can use this knowledge to communicate. On this front, this study analyses the use of linguistic forms through the recognition and production of language from a receptive perspective (reading) and how students use this vocabulary productively, in writing, for example. As this study will concentrate on vocabulary acquisition, it could shed light on how acquisition of knowledge takes place, the base of the linguistic approach, and how receptive and productive vocabulary serve as a way of accessing language knowledge to communicate effectively, which is linked to the psycholinguistic knowledge approach.

On the other hand, although to a smaller extent, the investigation could also shed light on bilingualism as a sociolinguistic practice in the Costa Rican context, the third of the approaches. The results of this analysis could provide some insights that could hint at future decision-making in terms of possible language learning policies in the country. The present study could inform on how students are motivated to learn English and on whether this motivation is contributing to improvement in language levels. The habitual presence of English native speakers as well as the influence of the U.S. here might have inadvertently served to motivate students to learn the language. Along these lines, the knowledge and use of vocabulary by the population in this study, as compared to vocabulary measures in other countries, could contribute some information on English learning as a social phenomenon in Costa Rica.

1.1.1. Fundamental theoretical concepts

The field of study of bilingualism has historically undergone many changes. The negative connotations associated to bilingualism, which prevailed up until the 1960s gave way to a wealth of studies that geared toward different, more positive views of bilingualism. Studies that highlighted the benefits of bilingualism and which moved away from constant comparisons of bilingual speakers to monolingual counterparts became more visible around the 1970s and have been growing steadily since then. The field of bilingualism has developed hand in hand with the field of Second Language Acquisition. In the following sections, we will discuss key concepts derived from these two areas of study.

Many authors have contributed key concepts that have fueled academic discussions and have also laid the groundwork for different research avenues. In the following subdivisions, we will discuss some of the concepts that have been conceived by renowned authors and which remain strong today.

1.1.1.1. Wallace E. Lambert

Lambert (1977, 1981) is credited in the literature with an essential differentiation between what he called *additive* and *subtractive* bilingualism. This distinction of terms has served as a pillar for several important theories put forward later and which will be discussed below. For Lambert (1981, p. 12), *additive* bilingualism is the one that takes place when "the two languages involved have social value and respect […] adding a second, socially relevant language to one's repertory of skills." Baker (2011, p. 71-72), along these same lines, describes it as a situation "where the addition of a second language and culture is unlikely to replace or displace the first language and culture […] The 'value-added' benefits may not only be linguistic and cultural, but social and economic as well." In sum, this type of bilingualism would bring greater benefits to the speakers and would yield greater linguistic, social and possibly positive financial gains without harming the already existing language or culture. *Subtractive* bilingualism, on the other hand, is defined by Lambert as the process in which:

> […] the learning of the second language necessarily portend the slow replacement of the first or "home," language [...] because of national educational policies and social pressures of various sorts, [speakers] feel forced to put aside or subtract out their ethnic languages for a more necessary and prestigious national language. (1981, p. 12)

Baker (2011) insists that the latter form of bilingualism has negative effects on the person's identity while conveying possible cultural and ethnic identity loss as well as marginalization. Along this same vein, García (2009) sees the progressive loss of features of the first language (L1) and the imminent turn to monolingualism as negative consequences of subtractive bilingualism. From the previous differentiation of types of bilingualism, it is clear that anyone interested in the field of bilingualism and second language acquisition (SLA) would pursue and support

an additive version of bilingualism. The idea behind becoming bilingual is that of strengthening those individuals in possession of these languages, may this be due to family, education, social or political reasons.

1.1.1.2. James Cummins

Turning to some key theories for which additive bilingualism is essential, James Cummins is probably one of the most recognized authors in the literature on the topic. His name is constantly found in papers and books across the field of study. The concepts presented below represent some of the cornerstones for bilingualism studies, as well. Cummins (1979, 1980, 2008) established a distinction between two types of proficiency that can be associated with cognitive challenges that bilinguals face. He determined that *Basic Interpersonal Communication Skills* (BICS) are different from students' *Cognitive Academic Language Proficiency* (CALP). BICS represent the communication skills that students put into practice every day and which are aided by contextual clues. These skills are not cognitively demanding and become evident in their daily interpersonal interactions. CALP, on the other hand, represents a higher level of linguistic proficiency necessary for more demanding topics on specific fields of study in the classroom context. This type of proficiency cannot rely on context because it usually involves interpretation of non-concrete concepts connected to literacy knowledge in the classroom setting.

Baker (2014, p. 233) defines CALP as "[t]he level of language required to understand academically demanding subject matter in a classroom. Such language is often abstract, without contextual supports such as gestures and the viewing of objects." From this definition, we can abstract the importance of distinguishing between these two types of skills. Given that they serve very different purposes in the classroom setting and that they exert a different impact on their function, their distinction plays a key role for both language instructors and students alike. Cummins (2008) acknowledges the value of this differentiation in relation to policy-making and educational practices, as well. He points to the relevance that this distinction has for educators in defining the type of proficiency that second language learners exhibit at given points of their second language learning process. Based on this, we can deduce that a clear recognition of the differences between these two concepts will aid instructors in determining realistic objectives and on designing activities suitable for students' capacities at any given point of the learning process.

Another chief concept introduced by Cummins (1979, 1984, 1986, 2000, 2005, 2009) is known as the *Interdependence Hypothesis*. Cummins (1984, p. 29) delimits this concept as follows: "[t]o the extent that instruction in L_x is effective in promoting proficiency in L_x, transfer of this proficiency to L_y will occur provided there is adequate exposure to L_y (either in school or environment) and adequate motivation to learn L_y." According to Cummins (1979), the L2 level that children can attain is intricately related to the L1 competence that they already had when they were fully exposed to their L2. This hypothesis is essential for the bilingual educational setting as well. It assigns a great deal of significance to the L1 in the learning process of an L2, in the sense that both languages can benefit greatly from the learning that takes place through either one of them. Cummins (1986) insists that as long as academic development is effectively taking place, transfer of concepts to the other language would follow. This is possible in an environment that allows the correct amount of exposure to and motivation through the language.

Cummins (2000, p. 70) describes the importance of "attribute-based aspects of proficiency" (individual traits) and "input-based aspects of proficiency" (from the surrounding environment) in the construction of appropriate levels of second language proficiency. He insists that there is a proven relationship between aspects of L1 and L2 proficiency because both are part of the same overall cognitive attributes of the learner. Cummins (2005) adds that the benefits of the Interdependence Hypothesis are not limited to bilingual education programs alone but can also be associated with reading abilities even in languages that are not similar in terms of writing systems. The benefits of bilingualism are linked to languages themselves and to conceptual terms as well. He describes five different types of transfer that can take place across languages in a bilingual learner: transfer of conceptual elements, transfer of metacognitive and metalinguistic strategies, transfer of pragmatic aspects of language use, transfer of specific linguistic elements, and transfer of phonological awareness (Cummins, 2007). The Interdependence Hypothesis is another key aspect to consider in a second language setting. Teachers should find a way of reaping benefits from each of the learner's languages so that both would contribute to a greater enrichment of one another. This hypothesis holds a strong connection with the two terms discussed below.

Cummins (1980, 1986, 2001, 2007) also proposed a *Common Underlying Proficiency (CUP)* model while discarding the *Separate Underlying Proficiency (SUP)* model alternative. Along the lines of the Interdependency Hypothesis, the CUP model points to an interdependence of the cognitive and academic types of proficiency that are subsumed under the L1 and L2 literacy abilities of students.

This idea is typically illustrated by Cummins through an image of a two-tipped iceberg (Figure 1).

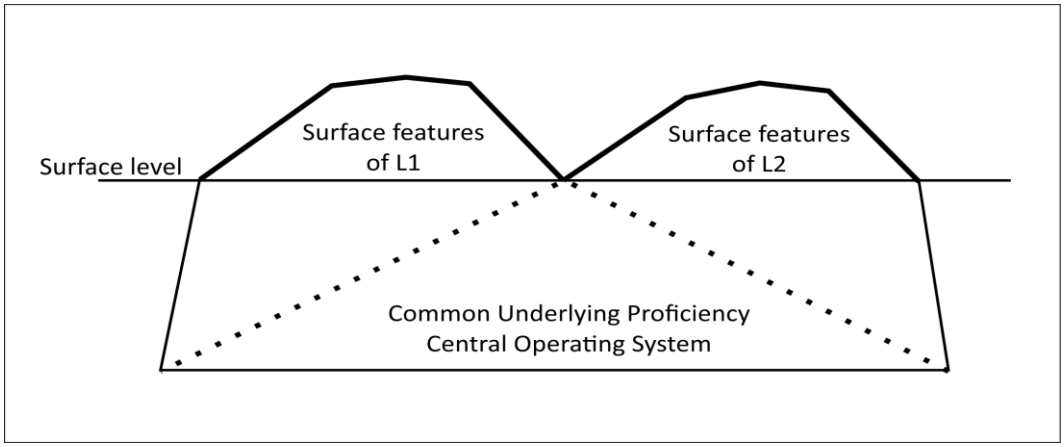


Figure 1. Cummins' Common Underlying Proficiency: The Dual-Iceberg Analogy (Adapted from Cummins, 1979, 1984 and Baker 2011).

Whereas each of the tips of the iceberg are visibly independent, representing autonomous abilities respectively for each one of the languages (i.e. morphology, phonology or syntax), the iceberg sits on a single foundation block embodying a single, underlying language proficiency (what Baker 2014, p. 234 calls "one underlying, central thinking system"). For Cummins (2001, p. 131) the CUP model stands for how the "experience with either language can, theoretically, promote the development of the proficiency underlying both languages, given adequate motivation and exposure to both, either in school or wider environment." Furthermore, Cummins (2005, p. 4) insists, "[…] when applied to bilingual education contexts, the Common Underlying Hypothesis refers to the cognitive/academic knowledge and abilities that underlie academic performance in both languages." He further contends that it is through the Common Underlying Proficiency that students can transfer academic, cognitive and literacy information, which demonstrates their level of language proficiency, from one language to the other language (Cummins, 2007).

The idea of the CUP has been made clear in the work of several authors. García (2009) speaks of the advantages shown by students who develop literacy in one language, at the time that they strengthen their literacy knowledge in another one. García believes that this is possible given that

> […] knowledge of linguistic practices as well as knowledge of the world, transfer across languages. Also, what is learned in one language does not have to be relearned in another, since conceptual knowledge transfers, and it is just linguistic labels that might have to be taught. (2009, p. 69)

She points to the importance of meaningful instruction to allow the appropriate development of a cognitive base that at the same time furthers development of language and literacy practices. Finally, the distinction between what we observe in the linguistic manifestations of our students can be differentiated from what is not observable. In this sense, Cummins and Swain (1986, p. 82) argue that "surface features of L1 and L2 are those that have become relatively automatized or less cognitively demanding whereas the underlying proficiency is that involved in cognitively demanding communicative tasks." With this, we can form a clearer image of what the CUP stands for. It is the base of knowledge in charge of sorting out cognitively demanding tasks, which feeds and becomes only slightly visible through the language manifestations of students in each one of their languages.

One last key concept has to do with learners' levels of linguistic attainment. Cummins and Swain (1986, 2014) argue in favor of the *Threshold Hypothesis***,** which regardless of what the authors label as an impossible delimitation in absolute terms (due mainly to each individual's level variation), is described as follows: "there may be threshold levels of linguistic competence which a bilingual child must attain both in order to avoid cognitive disadvantages and allow the potentially beneficial aspects of becoming bilingual to influence his cognitive functioning" (1986, p. 18). This hypothesis is commonly illustrated through a three-story house, as seen in Figure 2. Each floor represents a different threshold level with specific characteristics. This house also has one ladder for each language on either side of the house. This ladder illustrates the possibility of advancement in each of the languages toward higher levels of development.

For Cummins (1979, p. 229), "[t]he threshold hypothesis assumes that those aspects of bilingualism which might positively influence cognitive growth are unlikely to come into effect until the child has attained a certain minimum or threshold level of competence in a second language." In the representation of the three-story house, this could only be possible if the students' language proficiency has reached the upper level. Baker (2014, p. 56-58) describes the students' language characteristics in each "floor" as follows. The bottom floor houses *limited bilinguals*. When compared with their school age group "children have low levels

of competence in both languages with likely negative cognitive effects." These children are not likely to reap the cognitive benefits of bilingualism and may have trouble processing information in the classroom. Cummins (1979, p. 230) argues that, for these children, this limited proficiency in their L2 is thus likely "to impair the quality of their interaction with their educational environment through that language" which would, in turn, yield poor academic results.

For Baker (2014, p. 58), the second floor houses *less balanced bilinguals* who have "age-appropriate competence in one of their languages but not in both." For these students, there is no evidence of positive or negative results in connection with bilingualism. In addition, their thinking is similar to that of monolingual students. Finally, the third floor houses *balanced bilinguals* whose competence in L1 and L2 is age appropriate. Baker argues that at this level students have

> […] two well-developed languages [… and] they can cope with curriculum material in either of their languages. It is at this level that the positive cognitive advantages of bilingualism may appear. When a child has age-appropriate ability in both their languages, they may have thinking advantages over monolinguals. (2014, p. 58)

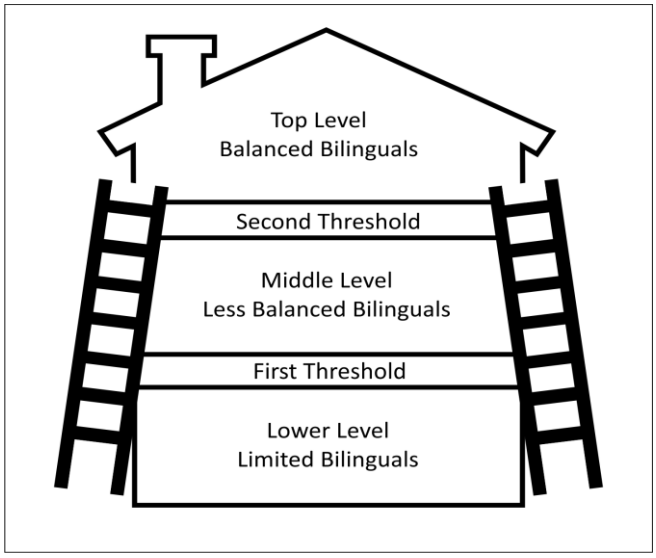


Figure 2. The Threshold Hypothesis (Adapted from Baker, 2011)

In this same vein, García (2009, p. 70) argues "high bilingual proficiency in two languages is associated with more positive cognitive effects […] development of bilingualism (whether additive, recursive or dynamic) where the additional language is acquired at no cost, can lead to improved linguistic outcomes and academic success." Knowing the implications that different language levels have on cognition, the Threshold Hypothesis (Cummins & Swain, 1986) is also called to have a forefront role in language learning settings. It should be the instructors' and curriculum planners' mission to ensure that students are provided with all the tools required to reach the "top level floor" of the language acquisition house. Only this level will guarantee the benefits that students search for when learning a second language.

1.1.1.3. Stephen Krashen

Stephen Krashen's (1982) theory of second language acquisition has also influenced approaches and methodology since it appeared. This theory originally consisted of five different hypotheses: The Acquisition/Learning Hypothesis, The Monitor Hypothesis, The Input Hypothesis, the Natural Order Hypothesis, and the Affective Filter Hypothesis. We will concentrate here on the Input Hypothesis (IH); however, the principles underlying each of these hypotheses are intricately present across the different hypotheses and thus some major traits could be identified in the IH. This hypothesis insists that providing second language learners with enough comprehensible input is the key condition to second language acquisition. According to Krashen (1982) the IH is made up of four parts, namely: IH relates to acquisition, not learning; acquisition takes place by understanding language (aided by extra-linguistic information) that contains structures that are a bit beyond the current understanding of the learner (i + 1); when input is understood, communication takes place and i + 1 is provided automatically; and finally, the ability to produce language emerges without need of instruction. Furthermore, Krashen (1982) contends that input needs to have specific features for it to be beneficial to learners. For input to be optimal, it must comply with the requirements of being: comprehensible, interesting or relevant, not grammatically sequenced, and sufficient in quantity. If the input has these features, Krashen's IH predicts that acquisition will take place.

Krashen (1982, 1989) argues that acquisition takes place when we concentrate on meaning, rather than on form and that this acquisition can be

achieved merely by understanding messages. Krashen (1991) describes the contribution of IH in terms of language acquisition and literacy development and contrasts the IH contribution to that of output, for example. He makes the following four claims: 1) comprehensible input increases proficiency; 2) methods with more comprehensible input have proven to be more effective when compared through research to other methods; 3) output and error correction do not have a strong presence in or outside class to be able to support a theory which is based on them (in reference to the one put forward through the Output Hypothesis (Swain, 1995, 2000a), for example); and 4) comprehensible input results in high levels of proficiency (these gains are still present in absence of output, skill building, error correction or comprehensible output).

Krashen (2003, 2013) and Krashen & Bland (2014) have found sufficient evidence regarding the benefits of a rich input environment; particularly from the perspective of (recreational) reading and the multiple profits it brings to vocabulary building, spelling and literacy in general. Second language researchers and language practitioners could easily identify the presence of this theory on approaches such as the communicative approach or CLIL. For several authors (Dalton Puffer, Nikula and Smit, 2010a; Harrop, 2012; Meyer, 2010; amongst others), Krashen's theory still permeates very current movements (such as CLIL today) in the form of its need for meaningful contexts and $i + 1$ requirements. Input holds a paramount role in second language acquisition. This key role assigned to input by Krashen's work has been and continues to be undeniably present in language settings today. The degree and quality of input that students receive will, to a strong degree, determine the level of achievement that they can attain in the language classroom.

1.1.1.4. Merril Swain

The next set of concepts comes from another major contributor to SLA and bilingualism: Merril Swain. When one tracks Swain's (clearly Vygotskyan-based) work over the last thirty years a number of contributions become evident. For the research that concerns us, we will focus on her ideas on the Output Hypothesis (Swain, 1995, 2000a); and the evolution that, according to my interpretation, one part of it undergoes to become the concept of *Languaging* (Swain, 2006, and Swain & Deters, 2007, among others). Swain (1995, p. 126) insists that output "pushes learners to process language more deeply (with more mental effort) than does

input." To begin the analysis of the output hypothesis, Swain (2000a: 99) argues that, through output, students could assume a more active role in learning by moving from the "strategic processing" focused on meaning that characterizes comprehension (input) to "complete grammatical processing needed for accurate production" that is more characteristic of output. She often insists on the idea that comprehension and production (hence acquisition as a final goal) demand two kinds of processing. These ideas clash with the previously mentioned input-based hypothesis.

There has certainly been some debate among these researchers. Regarding output, Krashen (1982) limits its contribution in second language acquisition to the fact that once learners produce language they will receive more input in return. Thus, output simply serves as a vehicle that ensures a better chance of getting more input in the form of corrective feedback or just as a response to communication *per se*. Krashen (1989, p. 454) admits that output can lead to "consciously learned competence." However, he insists that this is not as effective as the unconscious acquisition that results from comprehensible input. Keeping these contrasting ideas in mind, we will continue with the analysis of the output hypothesis and its undeniable contribution to the field of language learning.

Swain's output hypothesis attempts to reclaim importance over language production. Output appeared to have been relegated to a less outstanding position due to the emphasis given to input, derived mainly from Krashen's (1982) input hypothesis, in vogue at the time. Swain's (1995, 2000a) insists that language production supports second language acquisition in multiple ways. Swain & Lapkin (1995, p. 373) argue that what the output hypothesis stands for is the idea that "even without implicit or explicit feedback provided from an interlocutor about the learner's output, learners may still, on occasion, notice a gap in their own knowledge when they encounter a problem in trying to produce the L2."

In the output hypothesis, Swain (1995) identifies three main functions of output in second language acquisition. First, as described in the quote above, it serves a *noticing/triggering function* that allows learners to identify gaps of knowledge in their language. For Swain, this identification of linguistic limitations, in turn, results in conscious awareness on the students' part and may trigger certain cognitive processes that favor second language learning. Students' awareness may also lead them to actively seek a way to reduce this linguistic limitation using their knowledge about the language or by actively searching for a solution in the input they receive. The second function of output is that of *hypothesis testing.* For Swain (1995), students hypothesize using language forms and structures that they need for

the communication requirements they face. Learners try out forms to identify what works and what does not.

It is through this process of trial and error that students would decide whether what they know is correct or not, leading them to either maintaining or modifying their language forms and structures. The third function is the *metalinguistic function* or *reflective role.* This is the part of the hypothesis that, according to my interpretation, has received more attention by Swain throughout the development of this theory. According to Swain (1995), this function becomes clear when learners reflect on the hypotheses they have while using language to do so; either with themselves or with other peers (collaborative dialogue). For Swain, this practice is a metalinguistic function because students use language to reflect on language. This activity allows learners to have better control over their learning process and their language development while allowing them to own the knowledge they gain in the process.

Swain (1995) insists that this focus on output does not mean that input should be left aside. On the contrary, she insists, "output facilitates second language learning in ways that are different from, or enhance, those of input" (371). Swain (2000a) also contends that there is a need for both input and output focus in language learning processes. For her, *collaborative dialogue* serves a key role in language learning since it "facilitates the appropriation of both strategic processes and linguistic knowledge. These are insights that a focus on input or output alone misses." Hence, both input and output have a very active role in collaborative dialogue and that gives them a fundamental role as a problem-solving tool and knowledge-building base (Swain, 2000a).

Swain (2000a) assigns a great deal of value to *collaborative dialogue*. Swain (2000a, p. 97) defines it as "dialogue that constructs linguistic knowledge. It is what allows performance to outstrip competence. It is where language use and language learning can co-occur. It is language use mediating language learning. It is cognitive activity and it is social activity." Students engage in collaborative dialogue while they try to solve linguistic problems and while doing so they build knowledge that serves as a device to explore their cognitive capacities in more depth. This idea of collaborative dialogue will help us transition to the next concept put forward by Swain: *languaging*.

Swain (2000a) emphasizes the need to find a term to substitute the word "output," which she finds restrictive in defining the nature of the *activity* she has been describing through it. In that article, she opts to use concepts such as

"'speaking,' 'writing,' 'utterance' […] and collaborative dialogue," while, as she expresses, she comes up with a concept to better describe the phenomenon as including the "socially-constructed cognitive tool" that it encompasses (103). Finally, Swain (2006) argues that she attempts to foreground the concept *Languaging*, using Vygotsky's insights to support the notion. For her, this concept "puts the focus in second language learning on the importance of *producing* language […]" (p. 95, emphasis in original). She admittedly refers to the concept having been used with different meanings and on different research fields in the past, and she proceeds to present her conception of languaging for the SLA field. The following is a description of what Swain (2006) defines as *languaging*:

> For me, it conveyed an action – a dynamic, never-ending process of using language to make meaning […] I am using it to refer to *producing* language, and, in particular, to producing language in an attempt to understand – to problem-solve – to make meaning […] languaging about language is one of the ways we learn a second language to an advanced level […] languaging is a process which creates a visible or audible product about which one can language further […] Languaging, as I am using the term, refers to the process of making meaning and shaping knowledge and experience through language. It is part of what constitutes learning. Languaging about language is one of the ways we learn language. This means that the languaging (the dialogue or private speech) about language that learners engage in takes on new significance. In it, we can observe learners operating on linguistic data and coming to an understanding of previously less well understood material. In languaging, we see learning taking place. (p. 96-98)

We can link the term *languaging* to the concept of the *metalinguistic function* included in the Output Hypothesis (Swain, 1995, 2000a) described above. As can be observed in the previous paragraphs, both of these concepts serve the purpose of solving linguistic problems while making meaning and creating knowledge to support language learning, all the while using the language (through explanation, questioning or conceptualization of language forms for oneself or others) to learn more about that language. According to Swain & Watanabe (2013), languaging is present in collaborative dialogue and as such, serves as a language learning tool. They insist that it serves as a mediator of thinking and it, *per se*, mediates in the process of learning a second language.

Languaging deserves greater attention given the impact that it brings to a language learning setting. Providing students with opportunities to speak or write

with peers while they "language" may result in greater progress in the learning process. On this topic, Lapkin, Swain and Psyllakis (2010, p. 478) argue, "[i]f an individual is not given an opportunity to language, then the power to create meaning, to plan, to attend, to organize, and so on dissipate." Knouzi, Swain, Lapkin and Brooks (2010, p. 24) add: "through the use of language to intentionally organize and control their mental processes during the performance of cognitively complex tasks, [learners] discover meanings and connections that were unknown to them prior to the act of languaging."

The concept of *languaging* has been further explored in connection to scaffolding (Swain & Suzuki, 2008), self-scaffolding (Knouzi et al., 2010), and in association to Zones of Proximal Development (Lapkin et al., 2010). Swain (2013, p. 205) further emphasizes the role of emotion in connection to languaging in the second language classroom and insists "[t]houghts and emotions come into expression through languaging (collaborative dialogue and private speech)." She argues for the importance of teachers giving enough opportunities for students to language in class. This languaging practice could enable teachers to access the information that surfaces in these exchanges and they can, in turn, provide students with the help they may need to advance in the learning process. We can see here the value of allowing students to have a more active role in language classes. When students actively engage in their learning process, the number of benefits appears to be endless.

Two well-known theories that reflect on SLA have been briefly discussed above: Krashen's (1982) Input Hypothesis (Comprehension Hypothesis) and Swain's (1995, 2000a, 2006) Output Hypothesis (Languaging). Although these were conceived and have developed independently (and they were certainly presented separately here), they were described with the intention of offering two perspectives of the same issue. Hopefully, readers can accept that these notions should be seen and used as a complement of one another. They are not to be viewed as two ideas that describe different phenomena, but as two outlooks that describe a single phenomenon: second language acquisition.

1.1.1.5. Ofelia García

More recently, a new term has become part of bilingualism literature: *translanguaging*. Originally used by Cen Williams in the 1990s to describe a

pedagogical practice where students received input in one language and produced language forms in another language (Baker, 2014), the term has taken on a new meaning. Baker (2014) credits Ofelia García for widening, improving, and moving this practice from the classroom planning context to the normal everyday-life use of bilingual communication practices. As García (2009) puts it:

> […] translanguagings are *multiple discursive practices* in which bilinguals engage in order to *make sense of their bilingual worlds*. Translanguaging, therefore, goes beyond what has been termed code-switching […] although it includes it, as well as other kinds of bilingual language use and bilingual contact. (p. 45) (emphasis in original)

For García (2009), this approach to bilingualism relies on common bilingual practices more than on language alone, as has traditionally been the case. She insists that translanguaging not only serves the function of facilitating communication but also helps bilinguals reach a greater understanding of the world they live in. For García, translanguaging can be found in three particular settings. It can take place at the family level when different members of the family have different language levels and turn to another language to ensure equal access to communication for all members. In reference to this first setting, Baker (2014, p. 207) comments on one advantage of translanguaging in the household: "what a child has learned through one language in school can be expanded, extended and intensified through discussion with the parent in the other language" at home; thus, allowing major benefits at the cognitive level. For García (2009), a second possible setting for translanguaging can take place in a bilingual community. It could be used in this context to enable residents to deal with bilingual information in signs or public services. Finally, a third setting can take place at the level of communities that include language minority groups. In this sense, it would permit access to information for parents or community members with limited or no proficiency in the official language.

While Baker (2014, p. 206-207, bold in original) insists that translanguaging should be limited to individuals for whom both languages are either well-developed or developing strongly, he lists four possible benefits of translanguaging in the classroom context. The first is that it "may promote a deeper and **fuller understanding** of the subject matter." Second, it "may help children **develop oral communication and literacy in their weaker language**." This would result in

"**fuller bilingualism and biliteracy**" in both languages. Third, it "can facilitate **home-school cooperation,**" in cases, for example, in which the parent becomes more involved in school activities such as helping the kids to do homework. Fourth, "the integration of, for example, fluent English speakers and English learners" can be done smoothly through translanguaging. Although in the literature we find that there is a strong sense of connection between translanguaging and second language learning in contexts of minority groups, this practice can be just as much part of second language learning and teaching in the EFL classroom context.

We believe that using the L1 to facilitate comprehension of the L2 can definitely serve as a useful tool in a foreign language setting. Participants in the process must ensure that there is an appropriate balance in the use of the L1, as a tool to strengthen L2 development, given the appropriate reasons. This idea finds support within theories such as Cummins' (1979, 2000, 2005) Interdependence Hypothesis and its idea in connection with L1's contribution to L2 strengthening. Given that transfer of concepts can take place between L1 and L2, translanguaging can serve as a platform for learners to extend their knowledge. Through translanguaging, the Common Underlying Proficiency (Cummins, 1979, 1980) of learners is put to work through the use of shared features of languages to allow communication between speakers in different settings, thus strengthening learners' knowledge through language use in these novel contexts.

Furthermore, García and Wei (2014) explain the concept of trans-languaging as follows:

> […] for us translanguaging is an approach to the use of language, bilingualism and the education of bilinguals that considers the language practices of bilinguals not as two autonomous language systems as has been traditionally the case, but as one linguistic repertoire with features that have been societally constructed as belonging to two separate languages. (p. 2)

From this excerpt we confirm the importance of translanguaging as a phenomenon in which both languages are seen as a single unit that is constantly making use of the overall knowledge the learners have in communicating with others. Its scope extends to all the spheres in which a bilingual person participates. That being the case, according to García and Wei (2014), translanguaging affects and exerts its presence in the language education domain, and especially considering the advantageous use that it could produce in a language classroom, the understanding

and proper implementation of this concept in second language settings becomes paramount. The idea of translanguaging as defined in the last quote, also agrees with the concept of CUP (Cummins, 1980, 1986, 2007). That has to do with the notion that languages share one single underlying proficiency system that serves speakers as a base of knowledge where they can transfer academic and cognitive information across languages. We can conclude then that there are enough theoretical foundations to support the idea of translanguaging and its important role for bilingual speakers.

1.1.2. Bilingualism underpinnings

Although bilingualism has been an intrinsic part of humankind ever since trading and general interaction amongst social groups first took place, it is still today a fascinating and intriguing topic for specialist and non-specialist in the field. Researchers try to find answers to the different phenomena intervening in bilingual and multilingual contacts with the intention, among other things, of discovering patterns and specific characteristics of the human language faculty. Non-professionals on the field, on the other hand, marvel at the child, teenager or adult who is capable of using different languages from their linguistic repertoire to deal with everyday life situations, seemingly without much effort. Interestingly, despite being an element that is an intricate, old part of our society, there are factors that make bilingualism a current topic of study.

In today's society, globalization, and its demand for fast-paced communication, leads to continuous contact with people of different language groups. As a result of these demands, bilingualism has become more of a conscious endeavor in countries where monolingualism used to be the norm. Given that society is increasingly more oriented toward business and services, more and more people are required to be bilingual to satisfy the needs of highly ranked corporations and transnational companies. These facts lend importance to the topic and to the interest of researchers in finding better ways to achieve promising results in the attempt of becoming bilingual. It is obvious that parents are striving to find educational opportunities that deliver functional bilingualism. But, what exactly is bilingualism?

Above all, we must recall that bilingualism can take on many different meanings depending on who uses the term and for what purposes the term is used.

Wei (2005), for example, provides a list of 37 possible definitions for varieties of "bilinguals." As a result of this extensive variation, several questions arise: What defines bilingualism? Can we reach a straightforward definition of the concept? What entails to be a bilingual speaker? Specialists in the field of bilingualism argue that it is necessary to consider different elements when dealing with the concept of bilingualism. The following paragraphs present a description of concepts that are key to understanding and shaping the concept of bilingualism. Some of these fundamental notions will, in turn, be revisited later in this research in relation to the present study and the Costa Rican context in particular.

Baker (2011), although admittedly recognizing important connections between the following two distinctions, differentiates "individual bilingualism" (the individual possession of at least two languages) from "societal bilingualism" (the group possession of more than one language). We must admit that this distinction is key to differentiate a society which in itself is bilingual, from one that seeks bilingualism through means such as formal instruction in which not all individuals are involved. In addition, Baker (2011) argues, that when we think of the concept of bilingualism, we should bear in mind that there is a clear difference between *language ability*, usually referred to as *degree*, and *language use* in reference to the *function* given to a language. Both of these are also distinguishing features across societies.

Keeping the previous dichotomies in mind, Baker (2011) proposes eight different dimensions along which the concept of bilingualism should be analyzed. The first dimension concerns *ability*. The capability of bilinguals varies across each one of their language abilities (reading, writing, listening, speaking) and the subdivision that each one of these abilities has. Bilinguals could be more active in some language abilities while they could be more or less passive in others. The second dimension is *use*; that is, *where* we learned the language and *how* we use it may vary from one person to the other. For example, an English professor may speak English in the classroom and in job-related activities and not use this language so much at home; while a person married to an English native speaker may use English once she gets home, but not in her workplace. As Baker (2011, p. 16) states it, "[b]ilinguals typically use their two languages with different people, in different contexts, and for different purposes." The third dimension in Baker's list is *balance of two languages*. Although the balance in use of the two languages can change along our lifespan, usually one language tends to be more dominant than the other. As a consequence of this, bilinguals who are equally competent in both languages are not common.

Regarding *age,* the fourth dimension, the author distinguishes between *simultaneous* bilingualism, to describe children who learn two languages as their first language; and *consecutive* or *sequential* bilingualism, to describe children who learn a first language before starting to learn a second language. Fifth, the *development* dimension refers to *incipient* bilingualism. That would be the case if one of the languages is poorly developed while the other is strongly developed. *Ascendant* bilingualism takes place if the second language is developing somewhat steadily; *recessive* bilingualism describes a language that is starting to fade out, and *language attrition* describes the permanent loss of a language. Baker's sixth dimension concerns *culture* and it distinguishes between *monoculturalism*, the capacity of being proficient in only one culture; as opposed to *biculturalism,* which entails cultural proficiency of various languages' cultures as well as their different components.

In reference to *contexts*, the seventh dimension, Baker describes how some bilinguals are immersed in the community where their different languages are spoken while for others the second or other language is absent from their surrounding context. Along the lines of the previous discussion on Lambert's (1977, 1981) theory, another distinction for context can be drawn between *subtractive* bilingualism, in which one language is sacrificed in favor of another; and *additive* bilingualism, when the learning of another language is encouraged without necessarily sacrificing the first language. The final dimension describes *elective bilingualism* as the individual's possibility of choosing to learn another language in addition to their existing language. A distinction can be made in this dimension between *elective* and *circumstantial bilinguals.* For the latter, learning a language comes as a response to demands of the environment. In this case, the new language will help them perform better in a given situation in their daily lives. This learning is usually chosen to satisfy needs that are faced in the context where the individual is immersed. In that sense, there is less of a choice than in the former, as the new language must function accordingly in the given context.

Grosjean (2010, 2013) discusses what he considers basic concepts about bilingualism and concurs with Baker in several of them. For Grosjean (2013), the following concepts can lead us to the foundations and delimit the scope of the concept. The first two components are *language fluency* and *language use*. In terms of the former, he argues that there are common features regarding fluency in bilingualism. Namely, speakers are not necessarily equally fluent in all their languages; they might have an accent in at least one of their languages (which also identifies them as speakers of more than one language), and they might have

learned their second or other language after childhood. He insists that including fluency as a determining factor in the definition of bilingualism can turn out to be problematic given the implications that the above aspects have in speech production. The inclusion of these aspects in a definition may render many speakers unfit for the designation *bilingual*, whilst exclusion of the aspects results in a more inclusive description. Here is where the second component, *language use*, interconnects with the previous one. The rate of use of each of the languages will influence the previously mentioned factors of fluency or even age of learning. Given that not all languages are needed (and consequently used) equally, it is reasonable to identify differences in terms of fluency (amongst other things) amid languages in a single individual.

A parallel analysis is suggested in terms of language skills. Not all language skills are used to the same degree; thus, fluency may characterize some language skills more than others. To define the *language use* feature accurately, Grosjean (2013) gives close consideration to the *language history of bilinguals*. According to Grosjean, there are some contemplations that can help us build an image of a bilingual based on that person's language history. First, it is necessary to specify the age and the way in which the language was learned. Second, it is important to determine the level of proficiency attained, and to know to what extent the speaker used the language. Finally, it is essential to know whether the strength of use of one language surpassed at some point a language that was stronger earlier in the history of the speaker. The latter deserves further consideration because it is possible that, throughout their lives, speakers may have moved, changed jobs, or simply reconfigured the use of languages. This, of course, can affect the strength of language use and determine dominance or weakness of languages.

The third feature in Grosjean's (2013) list of considerations regarding bilingualism is the *function of languages*. The main idea underlying this feature is that different languages are found in different realms of our lives because they are used to perform different functions. The basis for this claim comes from Grosjean's (1997) *complementarity principle*: "Bilinguals usually acquire and use their languages for different purposes, in different domains of life, with different people" (in Grosjean 2013, p. 12). Grosjean (2013) explains that covering a wider range of domains with one language, that is, using one language more, would result in greater fluency in that language. Contrariwise, less use of a language results in weakening of that language and thus weaker fluency. Any one of the languages spoken by the individual could move up or down in this dominance continuum.

Consequently, even the strongest language at one point in life could end up becoming weak or even forgotten.

Moreover, the fourth relevant concept discussed here is *language mode*, defined as "the state of activation of the bilingual's languages and language processing mechanisms at a given point in time" (Grosjean 2013, p. 15). According to this author, bilinguals are constantly presented with language choices. Then, they decide what language to *activate* or *deactivate* in any given situation. The possibilities are described as varying along a continuum between a monolingual mode and a bilingual mode and speakers may move along this continuum as needed. Imagine, for instance, the following interaction between participants A and B. When interacting in a *monolingual mode*, participant B shares only one language in common with participant A, or in another scenario B is either reading or listening to the radio in one given language (in which case no response is required). In these cases, only one language is *active*. The *bilingual mode* is *active* when the different participants in the interaction share more than one language in common and thus these languages are used interchangeably during conversations. The *bilingual mode* allows either for the inclusion of words or for a complete switch of language during the interaction. Participants in this mode may move along any part of the monolingual/bilingual continuum at any point during the conversation.

The fifth feature of bilingualism described by Grosjean (2013) has to do with *interaction with bilinguals and monolinguals* and their effect in terms of *language choice*, *code-switching*, and *borrowing*. Along the lines of the *language mode* discussed above, Grosjean (2013) describes *language choice* in relation to the base language selected by speaker A depending on whether the interaction is with monolinguals or bilinguals. With a bilingual speaker, *code-switching* and *borrowing* is more likely. On one hand, *code-switching* takes place when a complete change of base language occurs even if just in one portion of the conversation (a word, a phrase or sentence). Borrowing, on the other hand, may involve a *loanword* (Grosjean, 2013), in which content and form are borrowed; or *loanshift* (Grosjean, 2013), which may result from taking a word and extending its original meaning in the base language or creating a new meaning by adapting words from one language to a pattern in the other language. Finally, the last feature considered by Grosjean (2013) is the idea of *biculturalism*; the author insists that, although possible either way, being bilingual does not necessarily mean being bicultural, and vice-versa. *Bicultural bilinguals* portray some sort of *bicultural behavior* that allows them to move along a continuum ranging from a *monocultural mode* and a *bicultural mode*, as described above for other modes. Bilinguals

manage to adapt most of their behaviors to a wide range of situations in their everyday life and navigate along this continuum as needed.

Moving on to another author, Mackey (2005) provides an in-depth analysis of four fundamental constructs in his description of the concept of bilingualism and analyzes the concept from an individual rather than social perspective. These constructs are degree, function, alternation, and interference. The first notion is *degree*. For Mackey (2005) this basic feature of bilingualism allows us to "determine how bilingual [an individual] is" (28). For him, this purpose can be achieved through testing the speaker's skills, in terms of *comprehension and expression,* in both languages. The author points out that the degree of mastery of skills may vary, so we have to describe aptitude in terms of the *variables, skills, and levels* in each one of the languages. He proposes a detailed analysis of each skill (listening, reading, writing, and speaking); and of the level required for each skill (phonological/graphic, grammatical, lexical, semantic and stylistic levels).

The second characteristic is *function*, and refers to the uses and conditions under which the language is practiced. For Mackey (2005), these functions may be external if they are "determined by the number of areas of contact and by the variation of each in duration, frequency, and pressure" (24) or they may be internal, which "include non-communicative uses, like internal speech, and the expression of intrinsic aptitudes […]" (32). He provides a very exhaustive division of each function. In the category of external functions, when he considers all the potential sources that could nurture language, he discusses, first, *the language in the household*; be this one the one spoken by parents, employees, or by the learner if she is working as an employee in a foreign country, for example. Second, the *community language,* referred to in connection to the language spoken in the community or the ethnic group, at church, at work or during leisure activities. The *school language* refers to the language taught as a subject or used as a medium of instruction in the school setting.

Communication means language covers all the digital or electronic sources as well as print as possibilities to access a different language. Finally, the *correspondence language* is the one in use due to work or family reasons and it is the last external factor considered to enable and conserve bilingualism. All of these external forms of contact may vary in terms of *duration* (length of use in years, months or days), *frequency* (number of contact per week, day or hour) and *pressure* (political, economic, cultural, religious, and so on) (Mackey, 2005). The internal functions, on the other hand, which serve a non-communicative purpose, have a more personal use. The first one is internal *use*. Like the name says, it is the interior

use given to a language in instances such as counting, calculating, praying or cursing, dreaming, note taking or diary writing. The second internal function has to do with aptitude. This describes the intrinsic capacity that may influence our language ability through features such as sex, age, intelligence, memory, language aptitude and motivation (Mackey, 2005).

As mentioned above, the third characteristic in Mackey's (2005) list is *alternation. Alternation* depends on the influence of each language in the individual's overall behavior as well as the individual's degree of mastery of the language. The rate at which bilinguals alternate between languages is influenced by their fluency, and by external and internal factors. According to Mackey (2005), *topic, person,* and *tension* condition alternation, the latter being determined by the speakers' situation, and whether they feel more or less pressured to use a given language. Finally, the fourth factor used by Mackey (2005) to describe bilingualism is *interference,* the habit of including structures or elements of one language when producing the other. Interference varies in *amount* and *frequency*; it presents variation in terms of the *medium* (oral or written), and *discourse style* (narrative, descriptive, and so on). Moreover, it varies in terms of the role of the speaker (i.e. register); and finally, in the cases described above, it varies from *situation to situation* and from *text to text*.

The author also describes a three-step procedure to delimit the type of interference that could take place. First, one needs to define which foreign constituent was brought in to be part of the base language. Second, one must examine what the speaker does with this constituent. And finally, one should determine how well that foreign element is covering for the original element. Mackey's (2005) aptly provides a thorough description of the different types of interference we could account for in bilinguals. These include *cultural, semantic, lexical, grammatical* and *phonological* interference. The latter covers intonation, rhythm, catenation (junction, syllabation) and articulation. Each of these types of interference would demand a detailed analysis of facts to come up with an accurate description of the extent of interference in every given bilingual individual.

As can be concluded from Mackey's (2005) discussion above, a simple definition of bilingualism is not easily attained. A description of bilingualism and of a bilingual speaker in particular, along the lines suggested above, would demand meticulous attention of a number of aspects that, at the end, can only help us reach a somewhat accurate description of this phenomenon at an individual level. A simplified, more general, characterization would be necessary if the intention is to describe the social phenomenon, especially if one considers that combining all the

elements described above into a single, general definition appears to be an impossible task. The previous list of elements, however, demonstrates once again that the phenomenon we are dealing with is very complex.

Authors such as Edwards (2006), very much like Baker above, also contend that a number of elements should be considered in relation to the concept of bilingualism. Edwards (2006), despite minor labeling differences, agrees with Baker (2011) about most of the elements that he considers fundamental. First, for Edwards, one should consider what *degree* of knowledge of the language in question is required for someone to be called bilingual. Second, he insists that one should consider the different *abilities* that are part of a language as well as their corresponding aspects. Take writing, for example, and then reflect on the ability in terms of spelling, sentence structure, punctuation, and so on. We, in turn, would need to do the same for all the corresponding components in all of the other abilities of the language. Third, the author argues that one should consider the *personal factors* that are part of the process. We may anticipate that this adds a great degree of complexity to the description given that elements such as gender, attitude, intelligence, and closeness of L1 and L2 should be contemplated. Finally, the issue of which *label* of bilingualism you are referring to needs to be addressed. A varied list of categories created throughout the years attempt to delimit the concept. Concepts such as "active or passive," "simultaneous or successive," "additive or subtractive" bilingualism are distinguished here. As the author states, defining the term truly represents a challenge.

A closer inspection of the elements discussed above reveals that there is a good level of agreement among these (and other) authors as to what elements constitute the underpinnings that inform bilingualism. Although some authors break the concepts down to very specific details that suppose a meticulous analysis of each aspect, we can still identify consensus. Despite minor details and variation in the label given to the concept and/or on the number of elements analyzed, authors acknowledge a list of quite similar aspects.

All authors discussed above refer to what Baker (2011) calls *ability, use,* and *balance* of languages. For Grosjean (2013), these are covered by the terms language *fluency*, *use* and *function*; for Mackey (2005), they spread across *degree* and *function* or even *alternation* (in the case of use), while in Edward's (2006) they are found under the constructs of *degree* and *ability*. The component of *age* is mentioned directly only by Baker (2011), whereas it is intrinsic in Grosjean's (2013) *history of bilinguals* or Edward's (2006) *personal factors*. Baker's (2011) *development* is similar to Edward's (2006) idea of *label*. Baker's (2011) notion of

context shares features with Grosjean's (2013) language *function* and *mode* and with Mackey's (2005) *alternation*. Furthermore, *culture* is addressed directly by Baker (2011) and Grosjean (2013), but Mackey (2005) does not refer to it, preferring to concentrate on *interference*, a concept not addressed by any of the other scholars, as also is the case of Baker's (2011) *elective* construct.

In sum, although some notions overlap with others across the different examples; or for some other notions authors use a different name, they all serve to indicate the complexity of the phenomenon and the many elements affecting it. Taking this brief discussion into consideration, we gather that despite apparent differences there is some general consensus as to what elements are considered to be the liable foundations on which the concept of bilingualism should be built. Aware of the challenge that it represents to come up with an accurate, clear-cut definition of the concept, we now move to review literature with this purpose in mind.

1.1.3. Toward a definition of the concept

Once one starts reading the literature in the field of bilingualism, it is common to encounter assertions such as: "the majority of the world's populations is bilingual," "half or more of the population in the world is bilingual," and "there are more bilingual than monolingual speakers in the world." After brief consideration, one can conclude that these statements make a lot of sense. If we think about it, we are all familiar with the following or similar ideas. We all know bilingual speakers. We know people who use other languages in their jobs or in their households. We know that our relatives are looking for bilingual schools for their children so that they can become bilingual. But why is bilingualism so important?

The following data illustrate the complexity of the concept. In April 2018, *Ethnologue* listed a total of 7,097 living languages in the world; and in that same month, the U.S Department of State included a total of 195 states on their official list of independent states in the world. These numbers alone can help us form an idea of how much language contact is bound to take place in the world when we are required to have constant interaction between speakers from that large number of languages and the not-so-large number of official states. Languages are considered part of a nation's identity, and citizens are, in general, proud of the

language of the country they live in. It appears that as long as language differences continue to exist and as long as we maintain contact with languages that are different from ours, bilingualism will continue to identify humans. But, what do bilingualism and being bilingual entail? Can bilingualism be defined? Many definitions have been given to the concept through the years.

Early literature on bilingualism insisted on pointing at deficiencies in the intellect and social skills of bilingual speakers. One example typically mentioned in the literature will be described below. In Wei (2005, p. 17), reference is made to Bloomfield's (1927, p. 395) categorization of a Menomini-English bilingual speaker as lacking knowledge in both languages. According to Wei, Bloomfield portrayed him as "[h]e may be said to speak no language tolerably." In particular, that author criticizes his "atrocious" vocabulary, grammar and syntax making a clear point to say that this speaker had not been able to learn either one of the languages appropriately. A few years later, Bloomfield (1933) would define bilingualism as "the native-like control of two or more languages" (in Baker, 2011, p. 8). As time has passed, some definitions have become more general and less strict. Weinreich (1968) provided the following definition: "[t]he practice of alternately using two languages will be called bilingualism, and the person involved bilingual" (in Hoffmann 1991, p. 15). Weinreich here gives a definition that is geared more toward the use of the language than to the strict proficiency demanded by Bloomfield earlier.

Researchers in the field acknowledge the difficulty of providing a straightforward definition. Though Baker (2011, p. 15), for example, claims that giving a definition for "who is or is not bilingual is essentially elusive and ultimately impossible," his insightful approach can guide us to categorize or at least offer an approximation to the concept, as he posits it. He argues that by using *distinctions and dimensions* like those discussed above, we could *refine our thinking* to find a better way of defining the individuals and the phenomena under study. Grosjean (2013, p. 5), on the other hand, defines bilingualism as "the use of two or more languages (or dialects) in everyday life" and for him "[*b*]*ilinguals are those who use two or more languages (or dialects) in their everyday lives*" (Grosjean (2010, p. 4), emphasis in original)."

This definition is broad enough to cover many types of bilinguals; and, when put to the task, may provide a good scale against which to measure speakers' bilingualism. The fact that it does not refer directly to proficiency allows for a variety of degrees of fluency across diverse language abilities. However, it calls for availability of language skills that would be sufficient to carry out daily activities.

In that same vein, Mackey (2005, p. 22) defines bilingualism as "the alternate use of two or more languages by the same individual." He points to the fact that the concept has broadened, since the beginning of the twentieth century, and has thus taken on an orientation that distances itself from the former idea of identical mastery in the two languages.

Several other definitions impute more or less specificity or do so in connection with different aspects. Skutnabb-Kangas and McCarty (2008, p. 2) define bilingualism as that exercise "involving proficiency in and use of two or more languages by an individual; the term does not always imply an equally high level of proficiency in all the relevant languages." This definition aims at proficiency as well as use on the part of the speaker, although it clearly allows the possibility of different degrees of proficiency. In a sense, it reminds us of Edwards' (2006, p. 9) idea in relation to the concept of "bilingual" usually describing "non-fluent" speakers. Moreover, Myers-Scotton (2006, p. 2) contributes to the discussion by defining bilingualism as "the term for speaking two or more languages." She asserts that being bilingual does not mean that you would have achieved full command of two languages, as equal fluency is rather rare. The clear delimitation to the speaking ability does not fade out as Myers-Scotton (2006) discusses several aspects related to the topic and she later redefines the term as "the ability to use two or more languages sufficiently to carry on a limited casual conversation" (p. 44). She insists that consideration should also be given to *grammatical competence* (recognition and production of "well-formed" language) and *communicative competence* (the appropriate use of language) when dealing with the concept. This will set the bar a bit higher when it comes to the degree of language ability required to deal with "casual conversations," even if it is "limited."

Wei (2005) gives another definition for bilingualism and opens up the possibility of the term describing people who may have partial mastery of only some language abilities, but not others. For him, bilingualism characterizes people

> […] who understand a second language – in either spoken or written form or both – but do not necessarily speak or write it, a more common usage of the term refers to someone who can function in both languages in conversational interaction. (2005, p. 13)

Furthermore, Wei (2005, p. 6) offers a general definition for "bilingual" as "someone with the possession of two languages. It can, however, be taken to

include the many people in the world who have varying degrees of proficiency in, and interchangeably use, three, four or more languages." Once again, reference is made in this definition to the varying degrees of proficiency that are included.

Several authors have been mentioned in relation to their contributions to delimit the concept of bilingualism. These definitions may be more or less specific, and they may concentrate on some features more than on others with the intention of serving particular purposes. Along these lines, Wei (2005) provides a list summarizing 37 possible definitions for varieties of *bilinguals*. This is yet another sign of how complex it is to define this term accurately or how specific it may become when subjected to certain requirements or given contexts. His list ranges from definitions commonly found in the literature such as *additive*, *passive*, or *active* bilinguals, to less familiar ones such as *horizontal*, *symmetrical*, or *vertical* bilinguals.

García (2009), and later García and Wei (2014), emphasize the importance of looking at bilingualism as a dynamic phenomenon, and thus to the complexity and malleability of bilingual language practices. García and Wei (2014) insist that more than being interdependent (Cummins, 1979, 2000, 2005), the languages of bilinguals adapt to society constructs sometimes and may lead to practices that are independent of constraints. Rather than two linguistic systems as Cummins claims, García and Wei (2014) state that dynamic bilingualism is one linguistic system that integrates features of both languages. This single linguistic system, according to García and Wei, may at times adapt to the constraints of the society, while operating with absolute freedom at other times. The treatment of this notion, however, still resembles the idea of the CUP (Cummins, 1980, 1986, 2007).

As seen above, definitions range from general to specific and may take different elements into consideration. While elements such as varied levels of proficiency in speakers as well as variation within abilities in one single speaker are common, different authors provide different definitions for the concepts depending on their vision of the phenomenon. Some appear more concerned with conversational skills while others refer to all skills of the language; some discuss the importance of being able to use the language while others demand grammatical competence in this use. What is clear, is that we now have definitions that are a lot more accepting of the various levels of proficiency and of the different types of bilinguals that are part of today's interconnected societies.

1.1.4. Advantages of bilingualism

Through the years, many benefits have been associated with bilingualism. Lambert (1981) highlighted the advantages of bilingualism in connection with language, cognitive and social development. He insisted that cognitive flexibility, creativity, divergent thinking and problem-solving are examples of these bilingual advantages. Considerable progress has been made determining and describing these advantages with great precision. Ellen Bialystok has been a major contributor in the area of cognitive development in bilingual individuals. Bialystok (1988 and 2005, among many others) and Bialystok and Majumber (1998) have shown, repeatedly, the nature and functioning of these benefits from the perspective of skill components linked to *analysis of linguistic knowledge* and *control of linguistic processing*. Bialystok (1988, p. 561) argues that the former is "the skill component responsible for the structuring and explication of linguistic knowledge." She contends that depending on the use children give to language in any given task or activity, the knowledge they require demands a more explicit (analyzed) or implicit (unanalyzed) level of linguistic knowledge. Bialystok (1988, p. 561) defines control of linguistic knowledge as "the executive component responsible for directing attention to the selection and integration of information." In this case, different uses of language (and different metalinguistic tasks) demand attention to be directed to different aspects, either to focus on certain aspects of information or to dismiss distracting cues.

The following are examples of many studies in which Bialystok and colleagues have shown how bilinguals differ from monolinguals in distinct experimental environments and, above all, how bilinguals' advantages are confirmed in different settings. Bialystok (1988) indicated how bilingual children could exert more control over their processing of language earlier in age and more easily than monolingual children. She also associated the level of bilingualism and its effect on solving metalinguistic problems. Bialystok, Martin and Viswanathan (2005) demonstrated that bilinguals from different age groups outperformed monolinguals in tasks that required control of attention and inhibition of misleading information. In this study, children, adults, and late-adult bilinguals performed better than monolinguals in those age groups. Only in the case of university undergraduate bilinguals was there no clear difference. The authors argue that bilingualism does not offer extra benefits at this age because participants at this level already enjoy an efficiency peak.

In a different study, Bialystok, Mahumder and Martin (2003) found that the previously identified advantages bilinguals enjoyed in tasks requiring inhibition of misleading information could be extended to tasks that do not require inhibition of attention as well. Bialystok et al. (2005, p. 110) concluded that for children, middle-aged adults, and older adults, "the bilingual advantage is for greater control and benefits their performance for both types of items." In this study, no difference was found between monolingual and bilingual young adults. Bialystok et al. attribute this to the "control of efficient processing" that already characterizes members of this age group (see Martin Rhee & Bialystok (2008), where these results are confirmed and extended to new task requirements). Bialystok et al. (2005) argue that bilingualism exerts a beneficial role in executive control processing throughout the lifespan, saying that this happens because

> the need to manage two active language systems and to manipulate attention to both during language use is carried out by the same general executive functions that are responsible for managing attention to any set of systems or stimuli. The experience of exercising these attentional systems enhances their function, and the benefit can be seen whenever control of attentional processing is required. (p. 117)

Bialystok and Viswanathan (2009) offer a slightly different perspective on cognitive development by distancing themselves from subjects sharing similar backgrounds and analyzing subjects coming from different cultural and linguistic backgrounds. This new feature in the analysis helps extend the benefits of bilingualism to a larger population. In this study, bilinguals perform better than monolinguals in tasks assessing the executive control components of inhibitory control and switching, which requires higher cognitive flexibility, but not in response suppression, the third component of executive control. Bialystok & Viswanathan (2009, p. 499) deduce in this study that "the modification of components of executive function by bilinguals is different at various stages of the lifespan, possibly because of differences in accumulated experience or the stability of executive functioning." Bialystok & Barac (2012) dabbled into the context of immersion programs and confirmed that the advantages found in other settings apply in this context. They studied, once again, the development of metalinguistic awareness and executive control.

Bialystok & Barac (2012, p. 71) determined that "metalinguistic performance improved with increased knowledge of the language of testing and

executive control performance improved with increased experience in a bilingual education environment." This is an encouraging piece of evidence for second language acquisition learners and teachers as it shows how second language learners reap important benefits derived from bilingualism, just as bilinguals derived from other contexts (i.e. those raised in bilingual households) do.

Bialystok (2005) explored three areas of cognitive development to determine whether bilingual or monolingual children acquired language differently or within a different time frame. She investigated quantitative concept abilities (i.e. mathematical operations), task switching and concept formation (i.e. tasks demanding control of attention) and theory of mind (a central concept to intellectual growth) that involves "knowledge that beliefs, attitudes, and perceptions are constructed by individual minds that have a particular (literal or metaphorical) point of view" (2005, p. 423). She found that, in all three scenarios, whenever attention and effort were required to analyze and/or eventually ignore pieces of information, bilinguals always solved problems better. Once again, in all cases, bilingual children outperformed their counterparts (and showed earlier mastery than their peers) on questions that demand ability to control attention. In yet another study, Calvo & Bialystok (2014) analyzed the role of socio-economic status (SES) and bilingualism in the development of cognition and language. These authors acknowledge the frequent connection that is drawn between SES and bilingualism, and between these aspects and language and cognitive development. They deem it important to determine the connection (if any) that exists between them. Calvo & Bialystok (2014) conclude that children from working and middle-class benefit equally from bilingualism, and a higher SES benefits both monolingual and bilingual learners. Both SES and level of bilingualism exert influence on language and cognitive development, but they do so through different mechanisms and independently from one another.

Aside from the cognitive advantages discussed earlier, bilingualism is associated with several other advantages in the literature. Baker (2014, p. 1) insists: "bilingualism has educational, social, economic and cultural consequences." Baker (2014) lists six general advantages and gives one or more subdivisions for each. First, *communication advantages* are associated with a) a wider range of communication possibilities regardless of whether they are needed to communicate with extended or international family members in immigrant families or to create job opportunities; and b) access to literacy in two languages. Second, *cultural advantages* can be linked to a deeper sense of enculturation and multiculturalism, providing speakers access to experiences in the world of each of their languages.

Cultural advantages also provide the bilingual individual with a higher degree of tolerance and recognition of diversity.

Third, Baker (2014) mentions *cognitive advantages* and lists many findings in this area. Some of the bilinguals' benefits mentioned in association to cognition included intellectual superiority, more elasticity in thinking, higher sensitivity to communication, faster cognitive development, skilled selective attention, enhanced analysis abilities, greater control of attention and better metalinguistic awareness. Several of these claims have been referred to in Bialystok's analysis above. Fourth, regarding *character advantages*, Baker mentions greater self-esteem and a strong identity. Fifth, some *curriculum advantages* are linked with superiority in curriculum achievement and better abilities to learn subsequent languages. Finally, bilingualism may have *cash advantages* since it can allow for a better economic situation and for employment benefits. Baetens Beardsmore (2008) agrees with many of the benefits listed above and reports bilinguals to have better spatial perception, greater creative thinking and better operational knowledge (needed for mathematical operations).

1.1.5. Remaining concerns about bilingualism

Enough evidence has been mentioned to conclude that, when bilingualism takes place under the appropriate conditions, it results in a large number of benefits for bilingual speakers. Although we are distant today from the idea of bilingualism being detrimental for children, different studies now provide evidence of a few concerns associated with bilingualism. A closer analysis of these challenges shows that they could be the result of specific sociological, individual, or contextual conditions or simply of poor implementation of bilingual instruction. Taken together, these so-called disadvantages, go unnoticed in light of the benefits brought about by bilingualism. Thus, they are addressed here simply to acknowledge these world-renowned authors. Studies such as those of Bialystok (2005), Martin Rhee & Bialystok (2008), and Calvo and Bialystok (2014) have confirmed that bilingual participants in their studies (regardless of the age group) tend to have a more reduced vocabulary than their monolingual counterparts.

Baker (2014) lists disadvantages linked with bilingualism and bilinguals. First, he warns of potential problems students may face if the languages they use at school have not been developed to the level required for curricular demands.

Underdeveloped language skills may lead to underachievement simply because they cannot deal with the information they need to succeed. Second, the experience may demand more effort from parents as they try to satisfy youngsters' needs at various levels of language development. Third, bilingualism may demand more emotional effort from parents to provide extra monitoring of their children's languages and to offer practice opportunities for them. Fourth, some young people may experience identity problems as they may find it difficult to establish self, cultural or ethnic ties. Finally, problems may originate in the home, extended family, community or country settings, where people could challenge, criticize or display negative attitudes toward the bilinguals' choice of language. Though there are certain drawbacks associated to bilingualism, they can be dealt with if the proper measures are taken. We must recall that the advantages outnumber by far, superior outcomes than the concerns mentioned above.

1.1.6. Concluding remarks

The concept of bilingualism has evolved greatly in the last decades. More updated definitions about the concept show a lot more flexibility regarding what bilingualism entails. Different authors have worked to provide evidence regarding the key elements that are fundamental for bilingualism to exist. The theories that have been presented in this section will guide the analysis that will be taken in this study on bilingualism in the Costa Rican context. Given the widely recognized advantages associated with bilingualism, the present study is yet another opportunity to explore the concept in a context that has not received extensive attention.

1.2. BILINGUAL EDUCATION

This section serves as an introduction to the topic of bilingual education. It first describes what Baker (2011, 2012) identifies as intervening factors in bilingual education; namely, societal, cultural, economic and political perspectives and their influence over bilingual education. That description is followed by a brief reference to different definitions that appear in the literature about bilingual education. Then, some of the most historically recognized approaches and models in second language instruction are described. These include approaches such as the Grammar

Translation method, the Direct Method, the Communicative Approach or the Task Based Approach, to name but a few. After that, a concise description of diverse types of bilingual education programs, according to Baker (2011), is presented. Finally, a succinct description of Mainstream Bilingual Education and Content Based Instruction, as they will be assumed in this study, is provided.

1.2.1. Intervening factors in bilingual education

Bilingual education is quite a complex phenomenon that deals with aspects that go beyond education alone. Baker (2011) analyzes bilingual education from the point of view of intervening elements such as society, culture, economy and politics, the strongest affecting factor for Baker. In addition, another important feature that Baker highlights is that bilingual education does not necessarily limit itself to a specific period of our lives. On the contrary, when it starts during childhood, it usually extends throughout the rest of our lives. It may, he insists, concentrate more or less during preschool, further or higher education, but it may as well last a lifetime. Along the lines mentioned above, Baker (2012) discusses the idea that bilingual education involves many more aspects than just the obvious concepts of bilingualism and education that may come to mind. He contends that a multidisciplinary understanding would be required for bilingual education to be achieved in harmony. For him, elements such as economics, philosophy, history, politics, language planning and others are inevitable components of the experience.

Baker (2012) further analyzes bilingual education from several standpoints that, he maintains, have an influential effect on bilingual education, namely language planning, pedagogy, politics and economy. His point of view on these perspectives will be briefly described here. To begin, *bilingual education as language planning* represents the vision of language planners and, as such, it serves specific purposes like language maintenance or language revitalization. The core interest of language planning would have to do with reproduction and use of language. In this perspective, strategies and actions contribute to four objectives: language acquisition in the family, language learning from childhood to adulthood, minority language use for economic purposes and the social, cultural and leisure contribution of minority languages to society. All these elements combine to achieve objectives in the perspective of language planning.

The second perspective is that of *bilingual education as pedagogy*. For Baker (2012), this perspective is best understood when analyzed through the results

that are delivered by bilingual education. He enumerates seven outcomes of bilingual education. First, bilingual education allows the complete advancement of two languages; it does not seek to substitute one language with another. Second, bilingual education results in opening the students' minds to a respectful appreciation of other cultures and beliefs. Third, bilingual education generally results in biliteracy. Fourth, bilingual education offers a number of cognitive advantages, such as creativity, problem solving ability and sensitivity to communication. Fifth, students' self-esteem may be heightened, especially in cases of minority students. Sixth, bilingual education contributes to improved curriculum achievement that is reflected in raised standards in schools. Finally, bilingual education may serve a critical role in setting local, regional, or national standards of identity for the students and community involved in the process. Baker (2012) asserts that whereas bilingual education does not guarantee higher standards of education *per se*, or more effective outcomes in general, in today's society bilingual education is commonly associated with "more marketable employees." The importance of bilingualism has assigned a fundamental role to bilingual education and its pedagogy. Although, bilingual education faces issues related to the type of languages learned at school, how these are learned or whether the learning of these languages stops once students leave school, for Baker, pedagogy continues to play a key role in the core development of bilingual education.

The third perspective mentioned by Baker (2012) is that of *bilingual education as politics*. Baker argues, "[b]ilingual education can be fully understood only in relation to political ideology and political opportunism" (p. 294). This connection between bilingual education and politics has turned the issue of bilingual education into a matter that is analyzed with respect to economy and it has raised questions of whether bilingual education is efficient or effective. For Baker (2012), to conceive bilingual education as a phenomenon independent from politics would be a naive conceptualization as "[w]henever bilingual education exists, politics is close by" (p. 300). He further maintains that bilingual education usually reflects political ideologies, its changes and initiatives; thus, it cannot really be studied as an element that is disconnected from politics.

The fourth perspective associated with bilingual education in Baker (2012) is the *economic perspective of education*. This perspective studies bilingual education in connection with how much money is invested in it and what economic benefits can be extracted from it. Bilingual education is analyzed from the perspective of the effectiveness and efficiency of its costs. For Baker (2012) "bilingual education provides higher levels of achievement in fewer years of study.

Student progress is faster, and higher achievement benefits society through less unemployment and a more skilled workforce" (p. 302). With these characteristics in mind, bilingual education can serve as an effective way to advance the economy. There are diverse ways of taking advantage of bilingual education to strengthen the economy. We surely can think of how bilingual education has brought economic benefits to people around us, be this through job opportunities for companies that require English, working with foreign tourism or having the option of traveling to other countries as a way of improving living conditions.

Baker concludes that integrating all the different perspectives—language planning, pedagogy, politics, and economics—would be ideal but it rarely ever happens; on the contrary, separation of perspectives is usually the norm. Above all, he believes bilingual education lies mainly under the influence of political forces and their leaders. Baker calls language planners and people involved with bilingual education pedagogies to join forces so that both the general public and the politicians can really see the true benefits of bilingual education.

1.2.2. Defining bilingual education

As with other concepts such as *bilingualism* or *bilingual*, defining bilingual education is a complicated task. Baker (2011) insists: "[b]ilingual Education is a **simplistic label for a complex phenomenon**" (p. 207) (emphasis in original). He maintains that what makes the phenomenon so complex is that it does not deal only with education but also with intervening elements from the society, culture, politics or economy, as discussed above. As a result, coming up with a straightforward definition of bilingual education is not a simple task. Marsh (2012) notes that when we encounter the concept of bilingual education, we tend to assume that learners in these programs are or will eventually become bilingual speakers; that, unfortunately, is not always the case. The following definitions and uses of the concept bilingual education show how this term is at times given a more specific meaning and at other times a more general use.

> Bilingual education is different from traditional language education programs that teach a second or a foreign language. For the most part, these traditional second or foreign-language programs teach the language *as a subject*, whereas bilingual education programs use the language *as a medium of instruction*; that is, bilingual

> education programs teach content through an additional language other than the children's home language. More than anything else, bilingual education is a way of providing meaningful and equitable education, as well as an education that builds tolerance towards other linguistic and cultural groups. In so doing, bilingual education programs provide a general education, teach in two or more languages, develop multiple understandings about languages and cultures, and foster appreciation of human diversity. (García, 2009, p. 6)

> Bilingual education is an umbrella term encompassing all forms of immersion education in addition to other bilingual models. (Tedick, 2014, p. 157)

> […] the term *bilingual education* refers to the use of two (or more) languages of instruction at some point in a student's school career. Each language is used as a medium of instruction to teach subject matter content rather than just the language itself. (Cummins, 2014a, p. 4, emphasis in original)

> Bilingual education is an umbrella term that encompasses a range of education programs that have been designed for an even wider range of children and a host of special circumstances. Essentially, bilingual education refers to any school program in which more than one language is used in the curriculum to teach non-language academic subject matter or the language of schooling does not match the language of the home or community, but the reasons for incorporating the languages, the specific languages chosen, the structure of the program, and the relation between the school languages and the community vary widely and influence educational outcomes. (Bialystok, 2016, p. 2)

As seen in the quotations, the definition is usually applied to programs involving bilingual education while in others it extends to models dealing with language education. Nevertheless, most of the definitions include the idea of using the languages in the process for curriculum instruction. As Bialystok (2016) states, "Bilingual education is not perfect and it is not one thing" (p. 11), it is, indeed, a phenomenon that is complex and that is connected to multiple other aspects of society (see Baker's (2012) perspectives above). The different definitions above show that different contextual characteristics of bilingual education must be considered if one intends to define bilingual education adequately. As Bialystok's definition points out, there are multiple reasons behind choosing specific languages to include in a bilingual education program. In her definition too, reference is made

to the type of program and to the desired outcomes on which these elements depend within each particular program setting.

Moreover, Baker (2011) provides a conceptualization that appears to be applicable to more specific contexts. He establishes a distinction between two types of bilingual education: *transitional* and *maintenance*. The former attempts to substitute the child's L1, usually a minority language, with the majority language in the given context. The child is supposed to assimilate socially and culturally to the society as well. The latter makes an effort for the child to maintain the L1, its culture and identity while at the same time acquiring an L2. Baker argues that *developmental maintenance* seeks not only to maintain the L1, but also to develop it along the L2 to reach complete biliteracy resulting in *enriched bilingual education*. This differentiation between transitional and maintenance bilingual education describes specific contexts such as that of minority students in the U.S., for example. The ideal situation, however, regardless of the context, would be one in which *developmental maintenance* takes place. As such, developmental maintenance is one in which bilingual education supports the idea of additive bilingualism in its students.

1.2.3. Approaches in second language teaching

From a broad perspective, bilingual education tends to be guided by a method or approach that determines the general organization of instruction in the classroom setting. In the literature, sometimes there is variation in the meaning given to the terms "approach" and "method," whereas at other times, they are used interchangeably (Hummel, 2014). The latter is the case in this section. Regardless of the term used, we will conceive a method or approach as "a teaching perspective that reflects a theoretical view or model" (Anthony, 1963, cited in Hummel, 2014, p. 107). In the next section, following Hummel (2014), we will briefly delineate the most important approaches that have guided second language teaching practices throughout history. These illustrate the variety of perspectives that have been implemented as well as various foci that these have had. Hummel (2014) describes the teaching practices during ancient times when the instruction of Latin and Greek was done through texts called *hermeneumata.* The approach used could be identified through these texts. These featured vocabulary lists grouped by subjects, tables with the alphabet and syllable descriptions and basic dialogues. The latter

was a very popular technique to teach languages up to the 1800s. This approach gave way to other methods which could be more familiar to us.

According to Hummel, the *Grammar Translation Method* (sometimes referred to as the 'Classical Method') was used predominantly from the Middle Ages to the nineteenth century. This method concentrated on grammatical rules and vocabulary memorization. Pronunciation received very limited attention since the interest was to enable students to read and write in Greek and Latin at first, and then in other languages. Grammar rules and vocabulary were used to translate texts from the target language into the L1 and vice versa; thus, the name Grammar Translation became a popular name for these practices. For Hummel this method could still be traced to teaching practices requiring explicit grammar instruction, rule memorization and emphasis on reading.

Another method described by Hummel (2014) is the *Direct Method*, which confronted the ideas typical of the Grammar Translation Method. The Direct method was implemented during the end of the nineteenth century and throughout much of the twentieth, first in North America and then in many other countries. Originally designed by the German immigrant Maximilian Berlitz, this method prioritized oral production and strictly avoided first language usage. It insisted on target language use in the classroom, implicit learning of grammatical rules, and learning of basic, practical vocabulary. Hummel maintains that this method was found successful in private schools with small, motivated groups. The results, however, were not as positive in large populations of public school students.

For Hummel (2014), the *Audiolingual Method* appeared in the U.S. toward the mid-1900s. It promoted repetitive drills and pattern practices of the same phrases with slight substitutions of certain units (nouns or adjectives). This method, like the Direct Method, also emphasized the development of oral communication skills. Less attention was given to vocabulary and although grammatical rules were not explicitly taught, there was a strong emphasis on repetition of grammatical patterns. After some time, one of the main shortcomings associated to this model was the learners' failure to develop real communication skills.

Hummel (2014) adds that *Suggestopedia* appeared in the 1970s, initially promoted by the Bulgarian Georgi Lozanov. It is characterized by an affective, more humanistic tendency that became popular during the second half of the twentieth century. This approach catered to emotional and affective features. Lozanov claimed that learners could be more receptive to language exposure if they were freed from stress and anxiety. Language learning was implemented in relaxing

environments that featured soothing music and comfortable seating for learners. The efficacy of this method was seriously questioned and the lack of evidence of its success caused the method never to be widely adopted.

In addition, Hummel (2014) argues that the *Total Physical Response (TPR) Method* became popular in the United States during the 1960s. For Hummel (2014), this method supported the idea that L2 learning could follow the dynamics common in L1 learning where speakers first concentrate on listening to the language around them and only later use the language for communicating through speaking. TPR required learners to respond to the instructor's commands by performing the actions demanded by them. Learners were not under pressure of producing language and thus were liberated from the stress that is associated with speaking in an L2. According to Hummel, this approach was widely implemented, although experimental support is mostly lacking.

Hummel (2014) maintains that during the 1970s and 1980s, first in the U.S. and then elsewhere, *Communicative Language Teaching*, or the *Communicative Approach*, appeared as a strong movement in second language teaching. This approach sought to solve the problems that other methods experienced in the development of communicative competence. It focused on developing the ability to communicate meaning and express oneself in real life situations. It relied on authentic materials to aid students to use language productively and receptively in situations simulating real life, for meaningful, authentic purposes. The main limitations of this method remain; first, it lacks specific methodology and thus allows for much variation in the classroom; and second, as it favors fluency over accuracy, at times students may lack precision when they communicate. This approach continues to be actively present in second language classrooms all over the world and clearly influences the most current SLA methodology.

According to Hummel (2014), the *Task Based Instruction* or *Task Based Language Teaching (TBLT)* appeared as a response to certain limitations in the Communicative Language Teaching approach. TBLT emphasized communication in concrete situations that resulted in tangible outcomes put into practice through specific tasks. It helped learners become aware of what they could do while using the language and at the same time promoted accuracy. This approach attempted to boost the learner's active participation in the classroom through meaningful language exchanges that sought for solutions to seemingly real problems.

Finally, Hummel (2014) points to the *Postmethod Perspective,* brought forward by researchers such as Kumaravadivelu, for which teachers should not

settle for a specific method or approach. Instead, they should look for a specific teaching methodology suiting the needs in which the teaching context is immersed. As such, the teaching context plays the active role of determining how the teaching of a second language should be implemented. This perspective responds to the demands of globalization and the characteristics of particular populations. We could consider the case of immigrants in many countries and the multiplicity of first languages and cultures that interact in these contexts.

Along these lines, García and Wei (2014) mention that Jacobson put forward "the concurrent approach" in the 1990s, going against the separation of languages in schools. Although the approach assigned only specific uses to code-switching, and was never really legitimized, it did bring up the possibility of a type of language teaching that did not involve the separation of languages, as commonly seen nowadays. The language separation practice has since been consistently questioned (i.e., Cummins 2007, 2014a, 2014b). García and Wei (2014) discuss the advantages of viewing language learning as a dynamic process supported by the interactions of bilingual students in classrooms and the ample potential that a dynamic languaging has for learning: "translanguaging not only promotes a deeper understanding of content, but also develops the weaker language in relationship with the one that is more dominant" (p. 64).

The most current types of bilingual education programs (i.e., Dual Language Learning and CLIL) do conserve many features of the communicative approach. Researchers are, however, making strong claims as to the role of the L1 in these contexts. Although reaching a consensus might appear difficult, the current state of a globalized society demands special methodology to be implemented. Multiple actors participate in implementing bilingual education programs, nonetheless. Some of these types of program are discussed in the next section.

1.2.4. Types of programs in bilingual education

Regarding bilingual education, Baker and Wright (2017, p. 199) present a summary (reproduced in Table 1) of the different types of bilingual education sought after by different programs. They provide characteristics of the type of children that commonly attend these programs, how language is dealt with in the classroom and the common features for societal and language aims.

Table 1. Typology of program models for bilingual students. Taken from Baker and Wright (2017)

MONOLINGUAL FORMS OF EDUCATION				
Type of Program	**Typical Type of Child**	**Language of the Classroom**	**Societal and Educational Aim**	**Aim in Language Outcome**
MAINSTREAMING/ SUBMERSION	Language Minority	Majority Language	Assimilation	Monolin-gualism
MAINSTREAMING/ SUBMERSION with Pull-Out Majority Language Instruction Support	Language Minority	Majority Language	Assimilation	Monolin-gualism
SHELTERED (STRUCTURED) IMMERSION	Language Minority	Majority Language	Assimilation	Monolin-gualism
SEGREGATIONIST	Language Minority	Minority Language (forced/no choice)	Apartheid	Monolin-gualism
WEAK FORMS OF BILINGUAL EDUCATION				
Type of Program	**Typical Type of Child**	**Language of the Classroom**	**Societal and Educational Aim**	**Aim in Language Outcome**
TRANSITIONAL	Language Minority	Moves from Minority to Majority Language	Assimilation/ Subtractive	Relative Mono-lingualism
MAINSTREAMING with World (Foreign) Language Teaching	Language Majority	Majority Language with L2/FL lessons	Limited Enrichment	Limited Bilingualism
SEPARATIST	Language Minority	Minority Language (out of choice)	Detachment/ Autonomy	Limited Bilingualism
STRONG FORMS OF BILINGUAL EDUCATION				
Type of Program	**Typical Type of Child**	**Language of the Classroom**	**Societal and Educational Aim**	**Aim in Language Outcome**
IMMERSION	Language Majority	Bilingual with initial emphasis on L2	Pluralism and Enrichment	Bilingualism and Biliteracy
MAINTENANCE/ HERITAGE LANGUAGE	Language Minority	Bilingual with emphasis on L1	Maintenance, Pluralism and Enrichment	Bilingualism and Biliteracy
TWO WAY/ DUAL LANGUAGE	Mixed Language Minority & Majority	Minority and Majority	Maintenance, Pluralism and Enrichment	Bilingualism and Biliteracy
MAINSTREAM BILINGUAL	Language Majority	Two Majority Languages	Maintenance, & Biliteracy and Enrichment	Bilingualism

As can be drawn from this information, there are several possible outcomes of bilingual education depending on the intervening factors of each given program.

1.2.5. Mainstream education and Content Based Instruction (CBI)

The Costa Rican students in the present study belong to two different types of program. One group is part of what can be defined as Mainstream Education and the other group participates in what is described as Content-Based Teaching.

As Baker (2011) and Baker and Wright (2017) note, it is not always easy to assign a certain teaching method to a given type of program. We propose characterizing these two types of program, however, according to the following descriptions. To begin, I will refer briefly to *Mainstream Education (with Foreign Language Teaching)*, a weak form of bilingual education, according to Baker (2011) and Baker and Wright (2017), and then to Content Based Instruction. In describing Mainstream Education (with Foreign Language Teaching), Baker (2011) notes:

> it is sometimes informally called a 'drip-feed' language program. The term 'drip-feed' highlights the kind of language element in mainstream schooling. Second (foreign) language lessons of around half an hour per day may constitute the sole 'other' language diet […] the language [is] a subject in the curriculum similar to science and mathematics. This is distinct from teaching through the medium of a second language where curriculum content is the main focus rather than language learning […] **Mainstream education** rarely produces functionally bilingual children. A very limited knowledge of a foreign language tends to be the typical outcome for the mass of the language majority. (p. 218, emphasis in original)

Baker (2011) goes on to explain that this is certainly not always the case. He highlights that there are countries in Europe, Asia and Africa where students in Mainstream Education achieve fluency in languages such as English, turning this into their second or even third language. He maintains that, in such contexts, factors such as motivation, high language status, and economic or professional interests contribute to that success.

The other type of program in this study involves Content-Based Instruction, which is defined by Reynolds (2015) as "an approach to the instruction of second

languages that integrates academic content concepts and skills with language learning skills by employing materials from academic subjects to achieve on par with grade-level academic expectations and language proficiency development" (p. 284). This definition suggests a cooperative balance between language and content instruction where proficiency in the language is attained in combination with academic content. Furthermore, Dupuy (2000: 212) defines Content Based Teaching as "[...] teaching a content area in the target language wherein students acquire both not language being the focus of curriculum; and (2) instructors providing students with comprehensible input." According to Dupuy (2000), in Content Based Instruction an attempt is made to resemble first language acquisition. Through that focus on meaning rather than form, the language input is slightly above that of the students' current level. It follows, then, that this input aims to provide an opportunity for students to use the language meaningfully while advancing in the acquisition of this language. Content Based Instruction could be placed in the context of what Baker (2011) and Baker and Wright (2017) call Mainstream Bilingual, a strong type of bilingual program, in the sense that two majority languages are used in the classroom for instruction (Spanish and English, in this case). Students are taking some of the course load in English (the second language) and participants such as the institution, the parents, and certainly the students are aiming at bilingualism and biliteracy attainment through their enrollment in this program.

1.2.6. Concluding remarks

According to the information discussed in this section, bilingual education entails more than the teaching of a foreign language in an EFL classroom. It demands the use of the language with communicative purposes in this classroom setting. Positive outcomes of bilingual education are highly attainable if students use the target language to learn content information in the form of Math, Science or History lessons taught through the target language. Other forms of so-called bilingual education programs where English is taught as another subject of the curriculum have repeatedly been proven to be inefficient. We need to keep in mind that there are many other aspects that are part of bilingual education and which are not usually associated with a language classroom; such are the cases of the economy and the politics of the country where the bilingual education program is located.

These considerations are important in the case of the Costa Rican setting, the focus of the present study.

CHAPTER 2

STRONG FORMS OF BILINGUAL EDUCATION

In this chapter we will discuss three types of programs that represent what Baker and Wright (2017) define as *strong forms of bilingual education for bilingualism and biliteracy.* [1] First, we will describe Immersion Programs, concentrating on the case of Canada. Immersion programs are credited with unprecedented success in terms of achievement of language levels as well as in terms of the scope of the society that these programs have reached. They serve as an example of parental and societal involvement that has taken the form of top down as well as bottom up organization.

The second type is Dual Language Programs. Emphasis will be given to the case of dual education in the United States. While this type of programs takes many forms depending on the context, here we will identify key characteristics as well as identifying features of successful programs. The third type to be discussed is Content and Language Integrated Learning, whose main context is Europe. While Baker and Wright (2017) do not refer explicitly to CLIL in their "Typology of program models for bilingual students" (p. 199), it is discussed later as part of their section on "Bilingual education in majority languages" (p. 235). Given that this program has been implemented in Europe to large scales and that it is yielding positive results in many settings, its features and principles can be seen as worthy examples to follow in other programs and contexts.

These programs were selected because they are beacons for bilingual education in that they all seek bilingualism and biliteracy (Baker and Wright, 2017) from the perspective of additive bilingualism (Lambert, 1981). We will deal with these types of programs in detail to describe the key features that make them beneficial for students. Such strong programs have been found exemplary both in their implementation and in the type of outcomes that they produce. Thus, it is useful to identify the features that may be implemented in other settings and which can result in functional second language education. After analyzing several forms of bilingual education here, in the next chapter we will inspect the current state of bilingual education in Costa Rica.

[1] An earlier version of concepts covered in this chapter has been published in: Castro-García, D. (2017). *Three forms of bilingual education: Immersion, Dual Language, and CLIL*. Heredia: Ediciones Escuela de Literatura y Ciencias del Lenguaje. This research was carried out as part of this doctoral thesis.

2.1. IMMERSION PROGRAMS: THE CASE OF CANADA[2]

This section describes various aspects of French immersion education in Canada. First, a brief outline is provided of the socio-political background in which the programs originated. Then, information about the goals and main defining features that identify Canadian immersion programs is presented. Information is also provided regarding the Association of Canadian Parents for French, a unique organization that has played a key role in these programs. After that, there is a brief description of the structure of immersion programs. This is followed by a description of findings which are grouped into three different subsections that address common benefits associated with immersion education, challenges faced by immersion education and other findings that have derived from research in these programs.

2.1.1. Background information

It is a well-known fact that the Canadian immersion programs have enjoyed long-term, international recognition because of their success in implementing second language teaching programs. Ever since French immersion programs were established in the 1960s, Canada has served as a referent for teachers and researchers interested in bilingualism taking place in a variety of immersion programs. As Baker (2011) posits, the original parents experiment to have a group of 26 kindergarteners immersed in a French program in St. Lambert, Montreal, in 1965, gave birth to an ongoing language learning practice that has produced excellent results in a framework of immersion bilingual education. The origin of the French immersion movement in Québec, rather than in Canada itself as a country, can be better understood considering that, in Canada, each province or territory is in charge of their education, and that the figure of a federal ministry of education is not included in its jurisdiction.

The socio-political Canadian context is a fundamental driving force for the creation and implementation of immersion programs. Canada is divided into 10

[2] As part of the research carried out for this doctoral thesis, earlier versions of several parts of this section have been published in Castro-García, D. (2017). Bilingualism, CLIL, and Immersion programs: The case of Costa Rica. In G. Nieto Caballero (Ed.), *Nuevas aportaciones al estudio de la enseñanza y aprendizaje de lenguas* (pp. 123–136). Cáceres: Universidad de Extremadura.

provinces and 3 territories. French holds majority language status in Québec, a historical French settlement since the 1600s. On a similar vein, although at the level of minority language status, French stands strongly present, to a larger extent in a sizeable number of speakers in Ontario and to a lesser extent in New Brunswick. In the rest of the provinces, French is a minority language, as well. Quebecois are known for their interest in preserving and maintaining the French language in their province. Baker (2011) argues that, in Canada, bilingualism, through second language acquisition, serves a societal function that aims to lessen conflict and seeks harmony between language groups while creating a more integrated society. He also notes that speaking a second language gives all citizens access to careers, higher education, jobs, travel, communication and information. For Baker, bilingualism in Canada has become key to secure jobs in the local government or civil service, to work as a teacher or in the mass media.

In Canada, French is present in different spheres: from their multilingual parliament, which serves as a sign of prestige and value toward the language, to local stores and households. The interest of Quebecois in strengthening the language is reflected on the different activities that they carry out to promote French language use. Baker (2011) sees it reflected in activities such as voluntary language classes or complementary and supplementary schools, and in vacation classes or Sunday schools. Through these activities, communities and families seek opportunities to promote the use of the language. Baker (2011) notes that it is common for bilinguals to avoid speaking the majority language in offices and shops when they feel that the majority language threatens their minority language.

Furthermore, Baker (2011) insists that immersion education should be viewed beyond educational terms as it involves a political, social, and cultural ideology. For him, immersion education is a movement toward a different kind of society rather than just a different type of bilingualism. For Baker (2011), "[b]y promoting bilingualism in English speakers, immersion education in Canada may support French language communities, increase the opportunities for francophones outside from Québec and help promote bilingualism in the public sector […]" (p. 271). With this social context in mind, we begin the analysis of immersion programs in Canada.

Different authors hint at other elements leading to immersion programs. According to Roy and Galiev (2011), French immersion programs appeared as a response to an overt interest of Quebec citizens to exert a better economic and political control over their province. Gibson and Roy (2015), on the other hand, report that, in 1963, the Canadian government appointed the Royal Commission on

Bilingualism and Biculturalism to analyze a series of challenges regarding the unequal distribution of English and French speakers in that country. This commission recommended protecting and supporting minority communities and their languages. Work at the governmental and commission levels continued for several years. Following recommendations given by that commission, the Parliament passed the first Official Languages Act in 1969. Subsequently, the government funded and promoted French instruction for French-speaking citizens and French as Second Language instruction for the majority groups.

Along with the governmental interest in a more even distribution of speakers of French and English in the region, parents were also concerned about their children's future and the new requirements their changing society imposed on their offspring. Gibson and Roy (2015) insist that the socio-political and cultural chaos was what inspired a group of parents to search for instructional methods to provide their children with the command of French that they required. With the help of specialists from McGill University in Montreal, the active parents convinced the district to approve an experimental plan recognized today as French Immersion Programs. This initial experiment is well known for its evident success in which students reap the benefits of proficient second language acquisition while maintaining an adequate level of English. The well-researched basis of the experiment, and the strong parental, socio-political, and institutional support contributed to the advantages accrued from these programs. They, in turn, have given way to continuing research, some of which will be discussed below.

2.1.2. Goals and defining features of French Immersion education

We could argue that a general goal that could be abstracted from reading about Canadian Immersion programs is that it seeks to attain bilingualism in French while still developing appropriate levels of proficiency in English. While this may account for the bigger picture, several other details that should be considered as objectives clearly go beyond bilingualism. In this vein, Baker (2011) maintains that the original goal of French immersion programs was clearly identifiable. The program aimed at children becoming bilingual and bicultural while still enjoying the benefits offered by their curriculum. To achieve this, the so-called "experiment" had clear objectives, which the author lists as follows: First, children would have to be literate in French; that is, read, write, as well as speak in the language; second, children would attain standard curriculum levels in all areas (including the English

language); and third, children would learn to value both English-speaking and French-speaking Canadian traditions.

Bingham Wesche (2002) states that early French immersion originally predicated three basic assumptions: 1) the notion that incidental, daily exposure to a language in a natural context results in children's innate acquisition of language and that this ability is reduced as age progresses; 2) the idea that frequent, varied and prolonged exposure to the language is necessary for learners to become fluent; and 3) the view that there is a need for language exposure and engagement that ought to be made available through activities that are interesting for learners and that take place in a natural rather than a formal instructional setting. These activities were supposed to facilitate language comprehension and demand production from students. Bingham Wesche (2002) insists that these features remain as basic elements of immersion education today and that they are reflected in early school starting ages. He adds that they are seen in initial instruction through French (many times taught by native speakers) that is progressively reduced in hours to incorporate English instruction; and in the use of French to teach curriculum that incorporates varied topics of interest for students.

Furthermore, different authors have reviewed fundamental features of the programs relating to the Canadian context where immersion takes place. Note that many of the characteristics considered by the authors are shared across the different analyses. Bingham Wesche (2002) describes a number of contextual features that characterize the original immersion program of St Lambert and that have prevailed across many other immersion programs in Canada. These features, the author says, have come to be considered crucial to the success of immersion programs. They include the following: 1) Students in the program were *speakers of the majority language*. As such, French represented an additional language and English continued to be developed as they were exposed to it at home and in the community at large. 2) All the students shared a similar language proficiency in French, so they all identified with one another in that sense, sharing motivation and interest in learning. 3) As an optional program, the students joined it voluntarily, thus contributing to enhance their motivation. Students, as well as parents, had a positive attitude toward the learning of the language, which can be thought to contribute in boosting that motivation. 4) There was strong social support toward the learning of both languages. Parents, community, and society in general positively valued the language learning process. 5) The languages and communities that these languages represented had several aspects in common. The two languages in the process were typologically related, sharing cognates and writing system features. At the

community level, people shared similar cultures and beliefs, life routines and holidays. 6) Local, political control in charge of funding issues and decision-making allowed for parental involvement that, in turn, endorsed constant innovation. 7) Immersion teachers were native speakers with extensive experience curriculum-wise, who were willing to teach in French. 8) A wealth of authentic material, pedagogical and otherwise, was available for students and teachers. For Bingham Wesche (2002), all these conditions laid the foundation for success in early immersion programs.

Swain and Lapkin (2005) bring us to a more contemporary analysis of Canadian immersion features. For Swain and Lapkin (2005) the current linguistic and ethnographic diversity of Canada ought to be considered when one thinks about the features that those programs should have. Although this article dates back more than ten years, its ideas are even more valid for today's reality. For these authors, acknowledgement of the changing sociopolitical context of Canada and the constant increase in the diversity of its population inevitably require a change in the pedagogy identifying immersion programs. While attempting to pave the way toward that change in pedagogy, Swain and Lapkin reviewed the former list of eight characteristics of the Canadian immersion programs published in Swain and Johnson (1997), analyzing whether the items on this list required adjustments to fulfill the current demands of Canadian society.

1) Swain and Lapkin (2005) argue that, given the diverse linguistic background of students, French, which was originally termed the second language, should be referred to as the immersion language. This is justified by the significant presence of students whose first language is not English, as was once the case, but one of many other possible languages.

2) The second feature contemplates that the curriculum of the immersion students is the same as the curriculum of core students; this feature remains unchanged according to the authors.

3) Third, they call for evident support for the students' first languages. The authors argue that in Swain & Johnson (1997) there was already a demand for the L1 to be part of the curriculum. Swain & Lapkin (2005) continue to stress the importance of L1 presence in the curriculum. This, of course, comes with the inevitable consideration of taking into account the various linguistic backgrounds represented in schools today. As a possible solution, the authors suggest the use of dual-immersion practices to take advantage of language varieties in the classroom. Other authors have stressed the importance of L1 in immersion programs. Cummins

et al. (2005) emphasized that "[p]re-existing knowledge for English language learners is encoded in their home languages. Consequently, educators should explicitly teach in a way that fosters transfer of concepts and skills from the student's home language to English" (p. 38). Cummins et al. (2005) acknowledge the challenge that this represents in Canadian classrooms where over 20 different L1 languages can interact within one school and describe the benefits that *dual language texts* can bring to minimize this situation in such varied language environments. For Cummins et al. (2005), dual language textbooks allow for the use of students' L1. This language serves as a resource while allowing the presence of multiple cultures in the classroom. All this is possible while the students' identity is also strengthened.

4) The original fourth feature in the review presented by Swain and Lapkin (2005) remains unchanged. This feature, a defining one for the authors, calls for a practice of additive bilingualism for students in immersion programs.

5) Their fifth key feature remains the same except for the substitution of the term L2 by immersion language, following the first characteristic described above. This feature refers to the idea that the immersion language is mostly restricted to classroom use.

6) The sixth feature conserves the same idea; that is, students in immersion programs share a similar linguistic background in regard to the immersion language, in this case French. Despite the variety in the students' language background, they all share a limited knowledge of French.

7) Teachers are bilingual (French/English). The authors mention the rarity of programs having multilingual instructors (who know languages other than English and French). This is even more significant if the diverse students' background is to be accounted for by these teachers.

8) The culture of the classroom must address the L1 cultures of all the individual cultures represented in it, and not just the L1 culture, as it was originally called for in Swain & Johnson (1997). This feature is clearly attempting to cover a much wider scope of cultures than was intended at first. Once more, there is an imminent need to acknowledge not only the variety of linguistic backgrounds, but also the multiple cultural forms that come with them. In their article, Cummins et al. (2005) argue that this situation can be overcome through the implementation of *identity texts*, as mentioned above. By means of this methodological practice, "instruction communicates respect for students' languages and cultures and

encourages students to engage with literacy and invest their identities in the learning process" (p. 42).

In sum, the results of this revised version of the eight characteristics of immersion programs responds to and seeks to integrate the vast linguistic and ethnic diversity in such a way that both first languages and cultures, in all their forms, become part of today's immersion programs.

Along the same lines as those of the authors above, Baker (2011) maintains that even today there are particular elements that identify the undeniable success of Canadian immersion programs. He describes eight identifiable, key features that have traditionally defined immersion programs: 1) He notes that these immersion programs aim at bilingualism in two languages that are majority languages of recognized prestige (French and English). 2) These programs are optional, so parents enroll their children in them voluntarily. 3) Children can use their first language (English) in beginning stages and during out-of-class activities. Keeping in mind the ideas put forward by Swain and Lapkin (2005) and Cummins et al. (2005) we understand that this is far more complicated now, due to the large numbers of cultures and L1s that are part of today's Canadian classrooms. 4) Teachers are fully bilingual (even if students do not perceive that from the beginning), and this contributes to high quality language exposure and modeling. 5) Classroom work concentrates on significant, real communication, through which students need to convey daily needs. This allows students to create associations between their learning and authentic language use in everyday situations. 6) Students share a similar linguistic background when they begin, this contributes to their self-esteem and motivation. 7) The curriculum in these programs is the same as that developed by mainstream, core students; this serves as a guarantee for equality at the social level. 8) Immersion represents a societal, political, or even economic as well as educational movement. There are many participants in immersion programs; parental involvement in particular keeps the community well informed and active in the education process.

The lists developed by Bingham Wesche (2002), Swain and Lapkin (2005), and Baker (2011) regarding Canadian immersion programs coincide in certain elements they consider essential, in some cases with slight changes in the wording that authors use. These include the following:

- All the lists provide descriptions of the programs as dealing with additive bilingualism. Baker (2011) describes bilingualism in two majority

languages, relating it to additive bilingualism. He suggests avoiding the term immersion, which could carry a subtractive meaning, and suggests using the term *submersion* instead.

- All descriptions refer to the possibility of L1 use. For Baker, L1 is permitted up to the first year and a half, while Swain and Lapkin (2005) discuss offering students overt support in all home languages. Bingham Wesche (2002) talks about L1 support at home and in the community.
- All descriptions mention teachers being (competent) bilinguals, and/or ideally multilingual.
- In all descriptions, reference is made to students sharing a similar background in the immersion language. Bingham Wesche (2002) terms it instructional language, Swain and Lapkin (2005) refer to it as immersion language, and Baker continues to call it second language.

Baker (2011) and Swain and Lapkin (2005) agree on the idea of students in immersion programs developing the same curriculum as students in core programs. Bingham Wesche (2002) points out three features that are not included by either Baker or Swain and Lapkin. They refer to the shared linguistic and social features common between the languages and communities; the local political and administrative control over immersion programs and the availability of authentic materials for immersion students. On the other hand, three features set apart Swain and Lapkin's (2005) and Baker's (2011) lists of key features. Baker (2011) emphasizes the optionality of the immersion programs, as also mentioned by Bingham Wesche (2002), the importance of meaningful communication in class and the social (also in Bingham Wesche), political, economic, and educational aspects of the immersion initiative. Conversely, Swain & Lapkin (2005) stress the importance of the immersion language being the language used for instruction, adding that for some students it is the L3. These authors refer to the confinement of the immersion language to the classroom, which can be related to Baker's (2011) third feature that describes students' liberty to use other languages outside the class. Lastly, Swain & Lapkin (2005) emphasize the need to acknowledge the home cultures of all students in the class.

In summary, we find that in addressing the same issue, all sources mentioned above delimit a number of key features that identify Canadian immersion programs and are essential in the evident achievements of the program. Despite some differences, the agreement in identifying defining features of

immersion programs points to the validity and strength of these features and to the fundamental role they play in consolidating and garnering favor for immersion education in Canada.

Additionally, Cummins (2014a) further highlights certain features that were first identified in Canadian immersion programs' and that have also contributed to the continued success of said programs; these are not included in the characterizations above. The fact that some of these characteristics have become core principles of models such as CLIL demonstrates how valuable these are. This value has been made evident due to the permanent examination that these programs have undergone, which Cummins (2014a) describes:

> The Canadian French immersion programs, however, were the first to specify some of the instructional strategies (now generally referred to as 'scaffolding') that are necessary to ensure that students comprehend the meaning of what is being communicated by the teacher. These strategies include the use of concrete demonstrations, visuals, and verbal paraphrases. Also, within French immersion programs, reformulation in French of what students say in English is always possible because virtually all teachers are reasonably fluent in English […] Additionally, French immersion programs were the first bilingual programs to be subjected to intensive long-term research evaluation, although some large-scale research had been undertaken in other contexts prior to the Canadian experience […] (p. 4-5)

This brief review shows some of the essential goals and features of the Canadian Immersion programs. Several other programs have tried to replicate the results of Canadian programs. Given the sound accomplishments associated with this model, this is only understandable. Another key feature follows: parental involvement in Canadian immersion programs.

2.1.3. Canadian Parents for French (CPF)

Parents have played a much more prominent role in Canadian Immersion programs than in any other educational movement. Just as when the French immersion programs originated, parents today who send their children to these programs do so with the intention of providing their offspring with as many

advantages as they can. For Roy (2008) "[m]ost parents who continue to send their children to French immersion programs do so because French is one of the official languages in Canada and because it provides their children with future opportunities: cognitive, social, and economical" (p. 397). Parental involvement has played a critical role in Canada. These parents' efforts are reflected in constant commitment of the association of Canadian Parents for French (CPF) to strengthen the French language. This commitment is particularly meaningful when one considers that many of the most common advocates are volunteer parents (Gibson and Roy, 2015). For Gibson and Roy (2015), CPF can be described as "[…] a grassroots, non-profit association established to promote French-second language learning in Canada [… their] efforts have contributed to the advancement of Canada's official language policy" (p. 218). A closer analysis at the work done by this association serves as an example of how shareholders can make a difference when it comes to seeking, implementing and maintaining fruitful educational programs in a given community.

According to Gibson and Roy (2015), the association was originally formed as a result of an initiative by the first Commissioner of Official Languages, Keith Spicer, in 1977. Spicer brought together representative parents who had a leading role in the promotion of French across the country. The mission of these participants was to report information back to stakeholders in each one of their regions. By request of Spicer, these parents participated in a Conference where the parents themselves, researchers, specialists, language consultants and other interested parties came up with a series of recommendations for the government. These recommendations ranged from curricular issues to cultural reinforcement in French communities; from teacher and staff training to student exchanges and funding. One of their key recommendations was the creation of a national association of parents. By October 1977, the association had defined clear objectives, had an established executive committee, and had developed a strong structure that featured representatives at the national, provincial or territorial, and community level. Gibson and Roy (2015) list three main objectives endorsed by the, then, newly created association. Namely,

> 1. to assist in ensuring that each Canadian child have the opportunity to acquire as great a knowledge of French language and culture as he or she is willing and able to attain;
>
> 2. to promote the best possible types of French language learning opportunities;

> 3. to establish and maintain effective communication between interested parents and educational and government authorities concerned with the provision of French language learning opportunities. (Canadian Parents for French, 1978, p.1) (p. 224)

Gibson and Roy (2015) argue that CPF efforts have focused, through the years, mainly on three fronts. First, they have ensured permanent distribution of key information among the different participants and the general public who are interested in immersion education. Second, they have advocated for French as a Second Language programs that are effective and that satisfy the students' needs. Third, they have promoted a wide variety of activities outside the classroom to enhance the knowledge that students acquire in the classroom setting. Gibson and Roy (2015) offer multiple examples of how the CPF has traditionally achieved these duties. In terms of distribution of information, the CPF association is very concerned with providing quality information to other parents and to their communities. With this intention, they have developed a *Handbook for parents*, distributed all over the country, with information about French second language instruction. Moreover, they offer diverse publications directed to reducing anxiety or feelings of weakness in parents and students in immersion programs. They produce materials for students; i.e., booklets, videos, surveys, activity books, collections of good practices and others. Finally, they have launched several multimedia campaigns to inform the population in general about Canadian immersion education practices.

For Gibson and Roy (2015) their second main aim, advocacy for effective programs, is mirrored in their constant search for funding for FSL programs. Whenever one of the funding agreements expires, CPF joins forces to present facts and make recommendations that reflect any new priorities and necessary adjustments in immersion programs. Second, CPF has made evident the necessity of offering opportunities for students to continue developing language skills at the postsecondary level. They advocate for permanent teacher training as well. They have worked publishing newsletters, organizing conferences, and publishing articles on this topic, thus attracting a great deal of attention to the issue of teacher training. Third, they have held many conferences on the topic of attrition in immersion programs. Fourth, they have shown great resistance to opponents of bilingualism and immersion programs in particular. Through cooperation with researchers and educators, they are skilled at presenting evidence that disqualifies their opponents' criticism. Fifth, they are in constant contact with decision-makers

and have willingly presented reports at the governmental or provincial level to argue in favor of immersion education. As part of their work, they have offered multiple consultations and symposia on the subject.

Regarding their third focus of work, extracurricular activities promoted by the CPF, Gibson and Roy (2015) offer many examples. These activities come in the form of summer programs for students in immersion programs offered at the local or provincial level. They organize family camping weeks where school-age participants get involved in French related activities. Moreover, they arrange public speaking competitions and "speak-offs," events where participating students come from core French, French Immersion, or French as a First Language groups. They have implemented national contests in partnerships with sponsor organizations. They have also promoted competitions, dances, carnivals, field trips, parties, and youth conferences among many other activities.

Finally, Gibson and Roy (2015) point out future challenges that the association faces. These have to do with greater inclusion of academically challenged students, appropriate provisions for students with varied linguistic backgrounds, adequate measurement and assessment of language proficiency on a life-long fashion, dealing with a shortage of teachers for core-French and French-immersion in some regions, developing a more articulated educational policy that reflects research findings and that serves as a directing guide for school boards, and postsecondary programs that develop FSL proficiency. While these challenges vary depending on the region, many regions would benefit from appropriate treatment with any or all of these.

As can be concluded from the previous paragraphs, the role of parents and stakeholders in language education is fundamental to ensure success. The CPF association serves as an excellent example of how a grassroots organization can be heard and can find the ways to fulfill the needs of a community or an entire country. This goes to show that top organizations do not always have to be in charge of initiatives for them to prosper. The achievements of CPF can be summarized in the following quote from Gibson and Roy (2015):

> The most obvious impact of the efforts undertaken by the grassroots members of Canadian Parents for French is the spread of the French immersion option from a limited number of major centers in 1977 to hundreds of communities large and small in every province and territory except Nunavut, and from a program mainly

> available at the elementary level to variations on the immersion approach from preschool through postsecondary. (p. 231)

2.1.4. Program structure

In terms of organization of the programs, to think of Canadian language programs as following only one structure would be a mistake. Great variety can be found across the country, its institutions and its language programs. Lazaruk (2007, p. 607) mentions the following types of bilingual programs. The first is *Core French*, the most common type of program offering students "a basic level of proficiency in French." It starts at some point between kindergarten and 5th grade, with 20 to 40 minutes of instruction per day, with the intention of developing students' language and cultural skills. The second is *French immersion*, in which French is used as the language of instruction, more than just as the subject of the class. These probably are the programs that have been studied the most and are the focus of the present section. Third, *Extended French* is a type of program available in a few provinces only. It adds a couple of core subjects taught in French to the French Language Arts courses in students from Grade 4 to secondary school. Finally, *Intensive French* is described by the author as the newest type of program in which students spend 70% of the day learning the French language in a FSL program that focuses on language use and which offers "an intensive period of French instruction covering half of a school year, generally in Grade 5 or 6."

Swain and Lapkin (1989) describe the basic distinction between the most common immersion programs: early vs. late immersion; they add that the program that has been studied the most is the early total immersion program. They describe the program simply as: "'early' because it begins with the beginning of school, 'total' because *all* instruction for the first few years is in French" (p. 151, emphasis in original). The other option, they say, is what is known as *late immersion*. Swain and Lapkin (1989) describe this program as follows:

> It is called "late" because it begins at about grade six, seven, or eight. Instruction in French may consume all class time or be as little as fifty percent of it. Prior to entering the program, students will have had daily periods of French as a second language ("core" FSL classes) for at least a year. Following the program, which may be for one or more years, students are again usually able to take several subjects per year in French if they so choose. (p. 151)

In addition to the starting age of students entering the programs, further characteristics contribute to the differences among the diverse types of programs. In consequence much variation is found within immersion education. Swain (2000b), concentrates on what the programs have in common rather than describing their distinctive features. She explains that although Canadian French immersion programs take on several forms, the element that they share is that students receive subjects such as history in French. She insists that during at least 50 percent of the school day, students use French while at the same time they are learning the language. Genesse and Jared (2008) describe how, originally, children in this type of language program initiated their literacy and academic instruction in their second language, French, before they learned to read and write in English. Students were instructed in French from kindergarten to Grade 2, when they started to be partly instructed in English as well. Makropoulos (2009) maintains that since the 1980s different types of immersion programs have become available to the public. In the following description Makropoulos (2009) aptly depicts the features of the different programs:

> Early French-immersion (EFI) programs begin at the onset of elementary school and provide equal instruction time in both official languages after initially exposing students to more French. Middle French-immersion (MFI) programs are sometimes offered from the onset of grades 4 and 5 to students from regular English programs. Late French-immersion (LFI) programs are offered at the intermediate level (grades 7 and 8), and provide up to 75% of the instruction time in French. Students from EFI, MFI, and LFI programs can enroll in secondary French-immersion programs that are typically offered from grade 9–12 (ages 14–17) in Canada. (p. 318)

Moreover, Baker (2011) refers to the diverse types of immersion bilingual education. For Baker, the program can be classified based on two components that are basic: *age* and *time*. According to the students' age, *early immersion* includes students at the kindergarten or infant stage; *delayed or middle immersion* describes settings where children start around ages nine or ten; and *late immersion* refers to programs that start at secondary school. On the other hand, based on the amount of time devoted to French instruction in the second language immersion program, *total immersion* starts with 100% immersion in the second language and then in periods of around two or three years it can be reduced progressively until it gets to 50% at the end of junior schooling. Furthermore, it is *partial immersion* when it offers,

from the beginning, around 50% immersion through both infant and junior education (239).

According to Cummins (2014b), in Canada, "[c]ore FSL programs typically teach French for 30–40 min each day. Starting grades vary from province to province, and within provinces school boards typically have some discretion regarding the starting grade level" (p. 2). He adds that over 90% of the student population in Canada follows this education option and insists on the need to turn to more successful bilingual programs such as immersion programs.

As stated previously, variation is a common aspect of immersion education. The different studies that inform the following section describe immersion programs with different structures. Certain terms discussed in this section (i.e., early, late immersion, core education) will be used below.

2.1.5. Research findings

Canadian immersion programs represent the type of bilingual education programs that have been most thoroughly investigated for the longest period of time. Due to the success of the programs and given that they have been implemented for an uninterrupted number of years, these programs have offered an ideal opportunity for researchers to carry out studies in a setting that offers ample possibilities for research. This research has yielded large amounts of outcomes and it has served for introspective reflection in many different areas. Some of the outcomes of this research are discussed in the following subsections.

2.1.5.1. Advantages

As a result of the success and rapid spread of immersion programs throughout Canada, these programs have received quite a bit of attention from researchers, educators, and specialists who have devoted a lot of time carrying out investigation in immersion classrooms. This research focus has led to an enormous amount of studies over more than five decades of immersion education. Commenting on the first studies carried out on immersion programs during the 1960s Lambert (1981) asserts:

> To our surprise, our bilingual youngsters in Montreal scored significantly higher than did carefully matched monolinguals on both verbal and nonverbal measures of intelligence; they were further advanced in school grade than were the monolinguals; and they performed as well or better on various tests of competence in French (the language of schooling) than did the monolingual controls, at the same time as they outperformed the controls by far on all tests of competence in English. Furthermore, their pattern of test results indicated that they, relative to monolinguals, had developed a more diversified structure of intelligence and more flexibility in thought, those very features of cognition that very likely determine the depth and breadth of language competence. (p. 10)

Findings such as those described by Lambert have been replicated in numerous studies; for a review of these studies, see Genesse and Jared (2008). Some studies that add evidence to describe the results of Canadian immersion education will be mentioned in the following paragraphs.

Swain and Lapkin (1989) present two main findings in their paper. First, they provide evidence to show that older students (enrolled in late immersion) can be as successful as younger learners (early immersion students) in learning particular properties of the language, and that older learners can be more effective, time-wise, than younger learners in doing so. They point to the fact that older learner's advantages are more evident in literacy skills (reading and writing) while younger learners perform better in listening and speaking tasks. The latter exhibit boosted confidence and lower degrees of anxiety as well. Swain and Lapkin (1989) attribute the older students' advantages in reading and writing to the implementation of visual, linguistic and cognitive strategies that they use in their L1 and to the cognitive maturity that allows them to process information on the basis of their L1 experience. The second argument presented by Swain and Lapkin (1989) has to do with methodology. They insist on the integration of grammar and content during instruction. According to these authors, "[…] grammar should not be taught in isolation from content. But then, neither should content be taught without regard to the language involved. A carefully planned *integration* of language and content, however, holds considerable promise" (p. 153, emphasis in original). This notion has been echoed by other authors and is the basis in teaching methodologies such as Content Language Integrated Learning, for example. As research continues, major support is being given to the idea of an adequate integration of grammar and content.

Swain (2000b) provides an avid description of the contributions derived from French immersion education, in particular toward the fields of Applied Linguistics (AL) and Second Language Acquisition (SLA). In reference to AL, she identifies three main contributions. The first is the awareness derived from French immersion programs toward acknowledging minority and majority groups, and within them, the influence that the social conditions of these groups exert in the outcomes of immersion classrooms. While French immersion acknowledges these societal conditions, it shows how both groups can benefit from the model. For Swain (2000b), the second input is related to the idea that "the contributions of disciplinary knowledge and related research paradigms to applied linguistics have been amply demonstrated in the FI research literature" (p. 200). For Swain, fields such as those of linguistics, sociolinguistics or education have found a space for academic practice through experimental, observational or ethnographic studies that have shed light on the method and results of immersion education, all the while strengthening each of the fields in the process. Third, immersion education has affected many subfields of Applied Linguistics (i.e., education, language policy, second language pedagogy, among others) in a more straightforward fashion, providing analysis and growth to the different fields.

Regarding specific contributions to SLA, Swain (2000b) lists the following six main areas. First, studies on French immersion have been the greatest contributors to a new, enhanced view of the role of *output in SLA*. Swain highlights the role of output (not just input) in the process of language learning. The importance of output for Swain is clear in her *Output Hypothesis*, discussed in section 1.1.1.4. Second, FI has provided much analysis and evidence toward the types and uses of *negative feedback* in SLA. Swain (2000b) refers to negotiation of form as an ideal type of feedback allowing students to reevaluate their output and come up with ways to repair their production. Third, the notion of *focus on form in relation to SLA* is linked to the previous ideas of output and negative feedback. Swain (2000b) claims "there is value in focusing on language form through the use of pre-planned curriculum materials in the context of content-based language learning" (p. 205). Fourth, for Swain the contributions in connection with the *role of L1 and SLA* have shown that the L1 is used, in immersion settings, mainly with three purposes: to advance in the completion of tasks, to talk about the L2, and to create and maintain interpersonal contacts in class.

Fifth, on the issue of *age and SLA*, studies suggest that there is a difference in the cognitive abilities that younger and older students put into practice, and that older learners may be more efficient than younger learners in the development of

literacy skills. Finally, regarding *language testing and SLA*, Swain explains that studies show that there is usually a disparity between what the tests measure and what students have actually learned. She insists that, as researchers, "if we are to measure the learning that occurs as a result of the research 'treatment,' [… we should] tailor our tests to what happens during that treatment" (p. 206). In summary, Swain (2000b) enumerates an important list of contributions that have favored and strengthened the development of SLA as a science. Swain explains how the *controlled* environment in which French immersion programs develop, serves as an ideal setting for research to take place. This is reflected in the amount of research that has been produced in this context.

Along these lines, Bingham Wesche (2002), maintains that the benefits of French immersion programs can be separated into three large categories: language skills, academic achievement, and language attitudes. First, in terms of French language skills, immersion students enrolled in early, middle, or late programs, exhibit a better functional command of the French language than students enrolled in core French classes. Within-group comparison also shows that students in early immersion programs have a better command of French than students in middle immersion and these, in turn, perform better than students in late immersion programs. Second, regarding academic achievement, assessment in math, science, and social studies has shown no detrimental effects of immersion in general achievement in English. Assessment in French seems to lag a little behind in early levels. The same occurs with late immersion students who have a limited French language background and who show initial, temporary, lower scores. Despite successful results in all types of programs, early immersion programs have important advantages when compared to all other programs. The third area of advantages comes in terms of language attitudes. The very positive attitudes reported by current and past immersion students supplements the learning experience and is very likely responsible for high levels of motivation and involvement.

Moreover, Bingham Wesche (2002) highlights that the immersion experience has consistently offered some general findings. The experience shows the advantages associated to early starting immersion programs. Children that start the program with a well-defined L1 (that progressively continues developing) benefit the most from immersion programs offering a robust communicative emphasis. Second, intensive instruction in immersion serves better purposes than reduced instruction that is extended in time. Even brief programs that offer very intensive instruction have proven to be more beneficial than programs with long

duration but limited instruction. Third, meaningful content learning results very significant as it represents a key vehicle for learning. Language learning through content instruction results in ample communicative opportunities with peers and, possibly, native speakers of the language. It offers opportunities for language analysis, feedback, communicative production and enhanced accuracy. For Bingham Wesche (2002), these characteristics in combination with the appropriate contextual factors, result in multiple benefits for immersion students.

Bialystok, Peets and Moreno (2014) set out to determine how much L2 instructional experience was necessary before benefits in metalinguistic awareness and executive control appeared in emerging bilingual children in immersion settings. On the premises that bilingualism accelerates the development of metalinguistic awareness, the authors studied how many years in immersion education were necessary to identify these metalinguistic benefits. Bialystok et al. (2014) found that two years of instruction were enough for immersion students to outperform non-immersion students in several tasks, regardless of low proficiency levels in the immersion language. They found that the metalinguistic advantages increase after 5 years of instruction, although not to the level of bilingual children.

In describing their findings, Bialystok et al. (2014) claim that immersion participants in the study outperformed non-immersion students in a morphology task (Wug Test). Immersion and non-immersion students originally performed equally in a grammaticality, sentence-judgment task, although more years in the program resulted in better performance for immersion students. Finally, for verbal fluency tasks, at a younger age receptive vocabulary levels were the same for both groups of students while productive vocabulary was higher for non-immersion students. As immersion children got older, they performed equivalently to monolingual students. The authors observed that students with a more organized semantic system and a larger vocabulary were able to come up with more words in these tasks, and that, regardless of the program, students produced more words as they grew older. Bialystok et al. (2014) conclude that greater experience in language use results in greater benefits and more similarity to the pattern found in bilingual students. As experience modifies ability, the authors say, the role of immersion classrooms becomes fundamental in helping children become bilingual as they incrementally enjoy more of the benefits brought about by bilingualism.

Jim Cummins has invested most of his professional career analyzing several aspects of French Immersion. In studying the overall outcomes of immersion education Cummins (2014a) comments:

> The overall outcomes of French immersion programs can be summarized as follows (see Baker, 2011; Genesee & Lindholm-Leary, 2013):
>
> – Students acquire reasonably good receptive skills (listening and reading) in French but their productive skills (speaking and writing) are limited with respect to grammatical accuracy and range of vocabulary.
>
> – Teaching through L2 entails no adverse effects on L1 literacy development.
>
> – In early immersion programs (starting in Kindergarten, age 5), students are able to develop functional decoding skills in French despite the fact that their French proficiency in the early grades is very limited.
>
> – A large majority of students spontaneously develop English decoding skills in Grade 1 and 2 with no formal instruction in English reading; many are reading much more fluently in English at the end of Grade 1 than in French.
>
> – Immersion appears appropriate for a wide variety of students – not just an academic elite. (p. 6–7)

Cummins (2014a) goes on to explain that there is still room for improvement if we look at these results. He insists on revising common pedagogical assumptions and using the vastly accumulated research evidence to implement instructional strategies that lead to even better results. Furthermore, along the same lines, Cummins (2014b) summarizes years of research in the following ideas:

> A common finding from L2 immersion programs across a variety of contexts is that students gain a reasonable level of fluency and literacy in L2 at no apparent cost to their academic skills in the socially dominant language. In the Canadian French immersion context, students catch up in most aspects of English standardized test performance within a year of the introduction of formal English language arts. With respect to French skills, students' receptive skills in French are better developed (in relation to native speaker norms) than are their expressive skills. By the end of elementary school (Grade 6, age 12) students are close to the level of native speakers in understanding and reading of formal French (assessed by standardized tests) […] (p. 3)

In a different study, Lappin-Fortin (2014) presents very interesting findings involving French immersion students, core French students and what she terms core French plus students, to refer to those who have participated in some kind of immersion experience either in Quebec or France. Her results show that French

immersion students express themselves with more ease than core French students. Also, core French plus students did better in terms of sentence length than the core French students. However, her findings show, among other results, that French immersion students do not outperform core French students in terms of accuracy and even more interestingly, that the core French plus group is more successful regarding total word counts and accuracy than either of the other groups. This last group outperforms both the French immersion students and the core French group in accuracy in simple sentence construction and one other variable under study: verb + infinitive use. She maintains that core French plus students possess greater knowledge of grammatical rules and verb morphology and they are able to use these rules accurately. The author attributes these differences to possibly higher motivational factors. These findings highlight the importance of participating in authentic immersion experiences as it evidences greater advantages for students who have enjoyed this experience over students who have followed French immersion and core French classes without spending time immersed in a French speaking culture.

Mady (2015) analyses the achievement results of 6th grade, Canadian born English/French bilingual, Canadian born multilingual, and immigrant multilingual students enrolled in immersion programs in the Ontario province (an English dominant region) of Canada. Mady found that the immigrant group outperformed the Canadian born English/French bilingual group and the Canadian born multilingual group in French speaking, reading and writing; integrative motivation and in oral willingness to communicate. The results were also significant in favor of the immigrant group for instrumental motivation when compared to the Canadian born bilingual group. For Mady (2015) these results are a clear indication of the greater advantage that immigrant students have in comparison with the other multilingual groups, despite the challenges to gain access to immersion education that this population faces. The results of Mady's (2015) study are evidence of the great effort that immigrant groups make to adapt to the Canadian community and its demands. Besides, they show that immigrants are not to blame for lower numbers of official language bilinguals in Canada, and further come as a good response to politicians who ask for high levels of language proficiency to select immigrants to be accepted in the country.

In addition, Hipfner-Boucher, Pasquarella, Chen, and Deacon (2016) investigated how French/English cognate awareness developed in immersion children and how this development related to reading comprehension. Their findings show that the immersion children showed an awareness of cognate

relations as early as first grade and that the development of this ability progressively develops through the first years of elementary school. Hipfner-Boucher et al. argue that cognate awareness is critical for lexical quality because it entails a much richer vocabulary in students and thus supports reading comprehension. They claim that their findings also substantiate the idea of cognate awareness as a metacognitive ability as it shows knowledge of words beyond their individual meaning. Furthermore, they found that this awareness of French/English cognate relations contributes positively (and could be used as a predictor) to French reading comprehension, both for students whose L1 is English and for those who have a different L1. This is an indication, they maintain, of the contribution of cognate awareness to reading comprehension even in cases where etymological connection does not include the students L1. Finally, Hipfner-Boucher et al. (2016) call for an early, explicit teaching of cognates in immersion education and for consideration of cognate awareness development as an additional metalinguistic skill in that context.

2.1.5.2. Challenges of Immersion Education

While the amount of benefits of immersion programs vastly outnumbers its limitations, the years that have gone by since the initial stages of Canadian immersion education have allowed for a very conscientious, in-depth analysis of the process. As a result, multiple ideas have been brought forward. Some of these concentrate on discussing problematic issues while others also intend to provide possible solutions for said shortcomings. Some of the challenges mentioned have to do with core features of the model while others have come up as a result of more current linguistic and social demands of the participants in the learning process. Makropoulos (2009), for example, argues that the present times offer more pressing issues. She maintains that stakeholders are aware of the changes in political priorities, the budget cuts affecting immersion programs, and how they have repercussions on accessing to these programs.

Cummins (2007), for example, points to a dire need to revise basic instructional practices that have become permanent assumptions in immersion programs. These notions are discussed by Cummins (2014a, 2014b), who finds that these assumptions lie on weak foundations and require immediate revision. He contends that these presuppositions are inconsistent with current theoretical findings and contradict evidence from areas such as Applied Linguistics and

Cognitive Psychology. The first erroneous assumption deals with the instructional use assigned to the TL and the avoidance of use of the L1 as a resource in the classroom. For Cummins (2007), while evidence supports the idea of extensive communicative TL language practice in the class, it further substantiates the cognitive usefulness of the L1 in the classroom. Cummins (2007) notes that students' prior linguistic and experiential knowledge can contribute to L2 advancement and he argues that it is only through resorting to L1 experience that prior knowledge is made available to learners. Cummins (2014a) argues that L1 and L2 academic skills are interconnected and that this is a resource that all students have equal access to, regardless of their linguistic backgrounds because "[t]he interdependence principle applies to languages that are linguistically distant from each other as well as to languages that have common roots" (p. 7).

The second assumption requiring revision is the idea that translation between L1 and L2 is not allowed in the immersion classroom. For Cummins (2007), translation is a powerful tool in the classroom and students must receive adequate guidance to access and deal with language information as efficiently as possible. For Cummins (2007), engaging students in writing projects could render translation a clear strength in the classroom. Cummins (2007) also mentions, as an example, products such as *identity texts*, developed by students in their two languages. Furthermore, Cummins (2014a) argues: "In the context of L2 immersion programs, translation in the instructional process has been regarded as highly counter-productive because students would likely *tune out* input in their weaker language, knowing that translation into their L1 was on its way" (p. 10). For reasons such as this, he insists, is that very little attention has been devoted to developing strategies that instruct students on how to transfer information across languages through translation. Cummins (2007, 2014a, 2014b) contends that, rather than avoiding translation, teachers should provide students with strategies to use translation as a powerful tool in their second language acquisition process. Besides translation, we should not lose sight, however, of the importance of input and output in L2. The use of L1 as a resource in the classroom should not interfere with the development of L2. Cummins (2014a) argues: "[i]t is certainly important within immersion programs to create largely separate spaces for each language. Extensive input (and ideally output) in the target language and active engagement with the language are prerequisites for acquisition" (p. 17).

The third assumption reviewed by Cummins (2007, 2014a, 2014b) is that one requiring strict separation of languages in the classroom, or what Cummins calls the *two solitudes assumption*. Cummins (2007) notes that this idea is

counterintuitive because students would naturally tend to make constant cross-linguistic connections. Cummins calls for a dynamic view of multilingualism instead: one that acknowledges the difference in the mental structures of monolinguals and bilinguals. For Cummins, if teachers instructed students to identify similarities and differences between their languages and guided them to develop strategies that entailed coordination across languages, students could learn the languages more efficiently.

On the other hand, Cummins (2007, 2014b) claims that rigorous separation of languages impedes activities such as directed teaching focused on French/English cognate relations, creation and distribution of multimedia language books, or projects such as sister exchange classes, all of which could contribute greatly to language development. Cummins (2007) concludes by pleading for a view of L1 as a tool to achieve superior L2 proficiency that can be achieved if the correct instructional strategies are implemented. Furthermore, Cummins (2014b) invites educators to make use of coordinated planning that could help them integrate objectives that are common to English and French. We can imagine how this joint learning of contents would strengthen the knowledge of features that are shared by both languages while facilitating students' comprehension and use of the languages.

Authors such as Howe (2014) defend that Canadian teacher education has traditionally provided pre-service teachers with a varied array of tools to succeed in their endeavors. Howe (2014) insists that characteristics such as: "effective leadership, multiculturalism, racial tolerance and global citizenship education, as well as comprehensive curriculum, teaching and learning, all figure prominently in teacher education" (p. 589). He maintains that multicultural teacher development has been a constant in Canadian teacher development. While this may be the case, ideas such as those put forth by Cummins (2014a, 2014b) certainly open the discussion for more and better ways to improve Canadian immersion programs from the perspectives of teacher training.

Furthermore, Makropoulos (2009) addresses the issue of class inequalities in school choices when certain parents lead the decision-making process to select schools, especially in the context of Late French immersion (LFI). She stresses the importance of offering options for bilingual education to older students, who may not have had that opportunity previously and who may not find this opportunity elsewhere. Makropoulos (2009) discusses the claims of students from immigrant and working-class families who did not enroll in early immersion programs either due to their parents' lack of knowledge and information about these programs or to

the parents' pressures of working life. The study shows that these students, however, later enrolled in LFI programs as a result of teachers' recommendations and their own personal interests. Makropoulos (2009) contends that middle-class family students actively tried to join the programs and showed a very clear interest in acquiring communication skills in French that could eventually provide them with economic advantages in the future. For Makropoulos (2009), working class families were far more concerned with employability issues, concentrating mostly on finding a way to join the labor market in the future.

Makropoulos (2009) points to a distinction between different groups of students in late immersion education. First, there are students belonging to a middle-class group, whose parents are English-speaking individuals and who receive encouragement from these parents to pursue LFI education. According to the author, parents' involvement in this group sometimes includes enrolling students in programs even when students are not necessarily interested in them. On the other hand, there are students whose parents are immigrants or English-speaking, working-class Canadian and who rely mostly on processes that are initiated by teachers' recommendations for students' enrollment in FI programs. The students in this group also show a clear, personal interest in joining the programs. A third group is formed by descendants of at least one Francophone parent, who join FI programs by initiative of their parents and who follow an interest in maintaining their families' French language skills. This study informs how social class and family background play a determining role in the reasons for enrolling LFI programs and about the roles of parents and students in this process. We could take this study as an invitation to look for fair procedures enabling all students to have a clear option for immersion education regardless of their social groups.

In a different study, Makropoulos (2010) explores the subjects of students' interest in FI programs and its connection with aspects such as class, culture and race. He found that students who had shown themselves to be engaged had related the immersion curriculum with their language abilities and needs. These students claim that the French skills acquired in secondary immersion classes provided them with the ability to succeed at the level required for university work. A distinction is made in that study between students with or without a Francophone parent. Engaged students with no Francophone parent claimed to face difficulty dealing with the linguistic demands in classes taught in French (their second or third language). Conversely, engaged students coming from Francophone families did not find problems facing linguistic demands since they relied on their previous

linguistic background, which, though varying in degrees, offered the comfort necessary in the situation, which was classified as easy for them.

On the other hand, disengaged students in Markopoulos' (2010) analysis claimed to perceive a disconnection between what the curriculum offered to them and their abilities and interests. Regardless of the students' achievements in terms of language and content and of how well they were doing in the program, these students believed that the immersion program was not something worth investing their time on. For disengaged students, one possible contributing factor identified by the author is whether students had a Francophone parent or not. As was also the case in Makropoulos (2009), the existence of a French linguistic background facilitated or hindered the processing of challenging linguistic information. For those students who do not have a Francophone parent, managing challenging linguistic information could serve as an additional factor for disengagement. In summary, this study shows how students' attitudes are shaped by how they see their own realities and needs related to linguistic and academic situations. The school and students' personal experiences contribute to their perception and attitudes toward the immersion program they were enrolled in. These findings represent a challenge for immersion programs in offering a type of language program that maintains students engaged for as long as possible.

Roy and Galiev (2011) describe the challenges that immersion students encounter in being recognized as bilinguals, despite the social and cognitive benefits associated to bilingualism in immersion contexts. The authors maintain that discourse on bilingualism shapes how students and others around them perceive their own language competencies. According to the authors, there is a held belief in the Canadian francophone society that sees bilingualism as the competent use of English and French where code-switching and transfer are not present. The authors found that while early immersion students are enrolled in the program by their parents, late immersion students make this choice by themselves because they need to know the country's two official languages. Students and parents who choose immersion do so because it offers cognitive advantages to learn subsequent languages, it creates better opportunities at the job and societal level, and because they could learn the two official languages of the country. Both parents and students do recognize that the language levels they reach might not correspond to their needs and that they might not suffice to compete for certain jobs. Immersion students do not consider themselves fully bilingual or having an adequate level of French, even after a number of years in the program.

Moreover, Roy and Galiev (2011) refer to the notion of what characteristics determine the appropriate variety of French and competency as a Francophone as contributing factors for students' self-evaluation of language proficiency. In this sense, teachers' decisions in the classroom, and their overt disqualification of certain language varieties, are said to have a strong effect on students' subsequent evaluations of themselves. The authors identified that Canadians see immersion students either as true bilinguals or as bilinguals who are not francophones. Canadians associate code switching with a lack of language proficiency, despite the fact that theory dictates the opposite. Moreover, sociocultural knowledge is also considered a distinctive feature of bilingualism by which French speaking learners of English are more likely defined as bilinguals than English speaking learners of French, simply because the former may have broader sociocultural knowledge.

The authors note that immersion education in Canada reflects the social and ideological perceptions of different actors in the society. Who is bilingual and what identifies her as such is shaped by the social and ideological characteristics deemed official in that society. Roy and Galiev (2011) conclude that French immersion education does not simply entail becoming a French-English bilingual, but "French immersion education is about gaining cultural, linguistic, and social wealth and dominance in Canadian society and elsewhere and about becoming competitive in the bilingual job market" (p. 371). A reflection of the issues presented in this study leads us to recognize the challenge faced by immersion students. In a society where native speakers define who is bilingual and at the same time represent the standard against which learners are measured, immersion students face a tough challenge in speaking both English and French well, while at the same time being accepted as bilinguals by the Francophone society groups.

On another issue, Cummins (2014b) draws attention toward the lack of coherence that characterizes the policies that regulate Canadian language teaching. This incoherence results from, he says, the independent and largely disconnected provincial jurisdiction that creates different policies and provisions. Cummins (2014b) thus calls for a change in FSL policies finding support for this claim in what he characterizes as a critical failure of Core French as a Second Language (FSL) programs across Canada. He contends that the Core FSL program, despite enrolling over 90% of students in FSL programs in Canada, continues to result in constant, disappointing failure. More particularly, he discusses three areas of interest that have been analyzed in greater depth in recent years.

First, Cummins (2014b) describes the issue of English and French instruction to newcomer students. Cummins notes the lack of training and

preparation at pre-service or professional development stages for teachers working in primary and secondary levels. He insists that teachers must excel not only in knowledge of their subject of expertise but also in dealing with students who are learning the language while learning the content through that language in their classrooms. Cummins proposes this be done following two cost-effective solutions. First, he argues that schools could specify the type of knowledge and abilities that they expect in the professionals they will hire, so that they can be guaranteed to have the linguistic and content knowledge needed to deal with the features of new immersion classrooms. Second, he suggests the creation of intra-school, advancement-criteria that would require specific qualifications of school professionals. This practice would ensure that schools with a diverse population would have professors and administrative staff with formal or demonstrated knowledge of educational practices in environments with varied linguistic or cultural representations.

The second issue addressed in Cummins (2014b) is the apathetic attitude of the government toward promoting and implementing bilingual programs for heritage languages: "Canadian provinces have shown little interest in imaginative approaches to heritage language education" (p. 6). Given the lack of demand of these programs in the different communities, the government has remained uninterested in the subject. The third issue has to do with the practice of denying bilingual opportunities to deaf children. He criticizes the common practice of Audio Verbal Therapists who forbid patients with cochlear implants to receive American Sign Language input or instruction. Cummins maintains that this practice denies this population the opportunity for bilingualism and stops them from engaging in real communicative practices while condemning them to spend time learning to decode speech instead. For Cummins, the three above-mentioned practices show how assumptions that are not substantiated on evidence can determine common policies and practices.

Along that same vein, Cobb (2015) taps into the sensitive issue of availability of Canadian French Immersion programs for special education students, and this issue becomes more complicated when the inclusion of parents in the children's educational process is not seen as beneficial or even necessary:

> Support for students with special needs varies in intensity and delivery, and may unfold as an alternative program, modifications to curriculum expectations, and/or accommodations to classroom environment, instructional delivery, and/or assessment […] Access to support has remained a persistent issue for students with

> a variety of needs, including those who are gifted, as well as those with attention deficit/hyperactivity disorder (AD/HD) and attention deficit disorder (ADD). (p. 171)

Cobb (2015) argues that support for the gifted is recognized as a type of special education only in Ontario, but not in the rest of Canada. Moreover, ADD, AD and HD are still under revision to determine whether they should be recognized as special education. Cobb mentions that while many children diagnosed with these conditions in Ontario could access bilingual education, this is not recognized as a right. Thus, fighting for access to these programs becomes even more difficult for special education students. In addition, Cobb draws attention toward the lack of support for special education children and their families. He notes that adequate support is crucial during the second language acquisition process for any student in general, and for special education students, in particular. For him, this includes gifted students who often face underachievement and frustration as a result of different experiences in the classroom, including facing insufficient challenge. Cobb insists that attention and support are necessary for special students and he suggests these be provided through prompt referral of students for assessment and by monitoring the process of students who are waiting to be assessed.

Along with the issue of accessibility, Cobb (2015) discusses the issue of exclusion that takes place when students are relocated in different learning programs that restrict their learning opportunities in some way. The process can also exclude parents (an integral element if exclusion is to be avoided) when they are not welcomed as part of the decision-making process. Cobb recommends different ways of minimizing exclusion in immersion programs. Namely, the development of inclusive processes where students are referred for assessment under supervision of both parents and teachers; for mentoring and professional development of teachers in identifying and providing support to special needs students; and for developing awareness of inclusion and conflict resolution techniques in immersion teachers to foster more productive relations with parents. For Cobb (2015), addressing the issues of accessibility and exclusion could guarantee a more equal education for children, especially in immersion programs. It could contribute to lowering the attrition rates that are prominently higher in the case of children and adolescents with learning disabilities.

Finally, Roy and Schafer (2015) explore the role of reading, in French immersion programs, and how this is commonly assumed to be a language skill rather than a social practice, thus largely limiting reading to curricular requirements

rather than associating it with leisure and entertainment. They add that the issue becomes more complicated when immersion students' reading abilities, and the cultural competence required for them, are measured against those of native speakers. Students may be asked to read native-like texts set in political and social contexts that are unfamiliar to them. The authors argue that cultural references and uses of French or Quebecois varieties unknown to students make texts more difficult for students and may discourage them from appreciating these texts. According to these authors, strong ideologies held by teachers and society in general are imposed on students who end up accepting and reproducing them. Expectations of native speaker-like competence and cultural knowledge can be harmful for students. Roy and Schafer insistently call for a wider view of literacy (especially in bilingual programs) that goes beyond simple acquisition of decoding and encoding skills. Roy and Schafer (2015) note:

> If we look at literacy practices, learning to read and write does not only mean that students need to read the words, acquire more vocabulary, understand different varieties of French and succeed during tests. It also means that literacy practices are co-created in many ways by the people inside and outside of the community. (p. 532)

They invite teachers in immersion programs to teach reading, not only to succeed in formal testing, but also as a possibility of using reading to examine and learn ideologies related to the language and culture.

While the number of challenges that immersion education faces may seem extensive, they should only be taken as a sign of a reflective, critical attitude on the part of researchers genuinely interested in continuing to improve the process of immersion education. This process of reflection is supported by over 50 years of experience and a steady growth in number of students and programs. Features such as the research and analysis derived from immersion programs have guaranteed the undeniable success associated with these programs.

2.1.5.3. Other findings

Bourgoin (2014) explores the predictive effects that L1 literacy has for intra- and inter-language predictions of reading skill abilities. Bourgoin departs from the

premise that reading skills represent a critical determinant of academic success. She insists that lack of consolidation of reading skills is even more detrimental for students at risk for reading difficulties and that these difficulties are greater for students with literacy limitations who are learning an L2. However, these limitations, she says, could be eradicated through adequate targeted instruction; hence, the need for timely identification and intervention. Bourgoin (2014) claims that phonological awareness (especially recognizing the first sound of a word) and alphabetical knowledge in early L1 literacy skills are significant predictors of L2 reading. Similar findings are claimed regarding L2 phonological awareness, and alphabetical knowledge and reading in the L2. These effects could be identified even in very early stages of French immersion. Bourgoin notes that special consideration should be given to the idea of using these initial literacy predictors to identify individual differences very early in the process of immersion so that students can have access to the benefits of well-developed reading skills. Bourgoin (2014) stresses the possibility of an early identification of students at risk for reading difficulties in L2 by administering appropriate tests before they start learning the L2 or once they have started learning the L2.

Cummins (2014b) provides a very detailed list of activities that can be implemented in the classroom to ensure the presence and acknowledgement of linguistically diverse populations and their languages in the second language classroom and in the school. He describes the idea that is being put into practice across Canada by a number of educators and researchers. He terms these practices *teaching through English as an additional language (EAL) lens* and *teaching through a multilingual lens*. He mentions simple activities that can be implemented to increase the presence of the students' languages in the classroom and in the school itself. Some of these activities include presentations of one new word by a different student each day; learning and using simple greetings in the different languages of the classrooms, welcoming students by greeting them in different languages; addressing participants in different languages at school assemblies and displaying students' work around the school.

Cummins (2014b) also lists activities that can be used to motivate students in the use of their L1 to develop reading skills, do research or take notes during class. All these activities could facilitate access to students' background knowledge and further it by accessing information in their L1 via Internet or other sources. For Cummins, use of their L1 could ease comprehension and facilitate transfer of information between languages. Students could be encouraged to use their L1 when they are planning or preparing for projects that will later be presented in the class

language. Moreover, students could read, tell stories or ask parents to tell them stories in the L1. School libraries should give them access to L1 dual language books. Teachers and principals could invite community members to class to read or tell stories. Students could research current topics using their L1 and then discuss these topics from a perspective of their L1 culture or ideology. In addition, Cummins (2014b) argues that teachers and schools could make use of technology to raise awareness regarding language, geography or intercultural realities. *Google Translate* can be used to develop rough versions of papers that can later be revised by more advanced speakers of that language. Using *Google Earth* can provide a closer look at specific regions and then promote healthy discussions about other countries and/or compare those countries to the Canadian reality. Finally, students can be part of a *Dual Language Project Work* that could be distributed online once it is ready or could even be built in cooperation with other schools via web.

Addressing a similar issue, Lyster and Tedick (2014) propose three ways to enhance students' metalinguistic awareness in immersion pedagogy. They find that the strengthening of metalinguistic awareness is a core goal for immersion education, and thus recommend that it be done through the implementation of the following practices. First, a stronger *focus on form* that allows awareness and practice activities targeting form either implicitly or explicitly. This practice would allow students to achieve not only better accuracy but also stronger academic literacy, as reflected in higher levels of oral and written proficiency. Second, a focus on *interaction and corrective feedback* between teachers and students is suggested. The authors note that how teachers and students interact in corrective feedback is a key feature in immersion pedagogy.

For Lyster and Tedick (2014), this interaction offers multiple possibilities for language development that can be instantiated through questioning and scaffolding which eventually lead students to greater engagement with language and content material. Through corrective feedback, teachers could draw students' attention to language forms to improve accuracy. Third, strengthening of *cross-lingual pedagogy and teacher collaboration* could reinforce greater vocabulary development and facilitate the use of L1 as a cognitive tool for L2 learning. These ideas are similar to those discussed by Cummins (2007, 2014a, 2014b). Lyster and Tedick (2014) strongly emphasize the power that teachers, as ultimate decision makers, have on directing students toward successful language learning outcomes. For these authors, strengthening metalinguistic awareness could contribute to the improvement of grammatical accuracy, lexical variety and sociolinguistic competence in immersion education students.

2.1.6. Concluding remarks

The movement that began as an experiment for barely over 20 students has served as a beacon for teaching practices for over five decades. Although still evolving as a social and educational movement, Canadian French Immersion Programs have served as a guide for educational practices around the world. With minor modifications, the original objectives and teaching practices and pedagogy still seek and reach the desired attainment of second language skills. The role of CPF is worth noting, as it provides an example to follow if we want a program to receive as much attention outside the classroom setting as it receives within it. The objectives pursued in the French Immersion classroom are echoed in the society through the efforts made in this organization. The different perspectives discussed by Baker (2012) in terms of a combination of elements at the political, social, educational and economical levels seem to be present in this context. The undisputable accomplishments of these programs, thus, continue to inspire second language models like those discussed in the following sections.

2.2. DUAL LANGUAGE EDUCATION: THE CASE OF THE U.S.

In this section, we will explore the characteristics of Dual Language Education (DLE) within the context of the U.S., where DLE represents another form of strong bilingual education in the sense of Baker and Wright's (2017) notion; that is, that DLE seeks attainment of functional bilingualism and biliteracy. After giving background information, we begin by discussing some of the distinguishing features of the model. A list of key factors that have been identified as determining elements of success across different programs will be mentioned. Then both positive effects, as well as problematic issues associated with the model, will be brought forward. Finally, we will present what different authors consider pressing issues for future research as well as some concluding remarks.

2.2.1. Background information

A review of the history and development of Dual Language Education in the United States reveals that this is a subject of permanent discussion and constant,

politically charged, modification (see Baker, 2011, for a historical review of bilingual education in the U.S.). In the literature on Dual Language Education, multiple names are used to designate this teaching model: Dual Language Education, two-way bilingual education, bilingual immersion education, two-way immersion education, developmental bilingual education (Freeman, 2000), dual immersion education, enrichment education (Torres-Guzmán, Kleyn, Morales-Rodríguez & Han, 2005) and dual language enrichment education (Parkes, 2008). In this study, we will use the term Dual Language Education to describe this phenomenon. Although DLE is often associated with the United States, this type of program also exists today in Macedonia, China, and the South Pacific (Baker, 2011); and Canada, Israel, Ireland and Germany (de Jong, 2014). The research discussed here, however, is centered on the conditions, characteristics, and evolution of these programs in the U.S.

Howard, Sugarman, Christian, Lindholm-Leary, and Rogers (2007, p. 1) define DLE as "any program that provides literacy and content instruction to all students through two languages and that promotes bilingualism and biliteracy, grade-level academic achievement, and multicultural competence for all students." According to Baker (2011) DLE started in Dade County, Florida in the United States in 1963. He describes how parents from the Cuban community in the area, with the hope of eventually returning to Cuba, were interested in their children conserving Spanish proficiency while being able to function in the English-speaking society they were immersed in. For Baker (2011), since the dual language program began to be implemented, English-speaking children from middle class parents were enrolled in it, showing interest from parents in foreign language instruction. As the phenomenon evolved in the U.S., it continued to mix students from majority and minority groups across the country. In this context, when DLE programs were first implemented, they were seen as an option that would use the experience gained in the immersion programs in Canada while at the same time having a chance of reaping greater benefits due to the presence of native speakers of both languages in the classroom (Valdés, 2013), something the Canadian programs did not always enjoy.

Valdés (2013) mentions that DLE appeared to be promising from several perspectives. To begin, it offered instruction in the first language to students from Mexican minority groups; and second, it brought together majority and minority students, to help diminish the historical segregation toward the minority group. These favorable characteristics that initially described DLE education have only been strengthened throughout time. Baker (2011, p. 231) maintains: "[d]ual

Language education attempts to effect social, cultural, economic or political change, particularly in strengthening the weak, empowering the powerless, and working for peace and humanity in the midst of conflict and terror." It can be deduced from these ideas that DLE programs are identified with a strong sense of social change that seeks for a more equal society. Parkes (2008) insists that both children and parents in this model differ substantially from participants in any other kind of model. Palmer, Martínez, Mateus, and Henderson, (2014, p. 762) consider that implementation of DLE programs that foster social integration represent a triggering element for a change in school demographics where more "white native English-speaking students" enroll in DLE programs such as the one they describe in their study.

Both in its origins, linked to Cuban immigrants, as in its historical development, often linked to Mexican immigrants, Spanish has dominated the landscape of DLE representing the most frequent counterpart language to English. More recently, however, languages such as Korean, Mandarin and French (Tedick, 2014), are also part of this landscape. It is evident why the presence of minority and majority populations is often referred to in descriptions of DLE; this feature stems from the core of the model. According to Lindholm-Leary (2012), in general, what a DLE program does is combine English Language Learner (ELL) students that share one of the minority languages in the area and native English-speaking children in a single educational setting where they would receive academic instruction through both languages. These features establish the context for an enriched setting where two groups of students that bring with them contrastive backgrounds and histories enjoy high quality education in an environment that promotes equal opportunities for both populations. For Torres-Guzmán, et al. (2005), these rich characteristics of the population in dual language programs "can translate concerns about language development and cultural pluralism into an equitable linguistic and cultural education for all students […]" (p. 472).

More recently, a new form of DLE model has been gaining popularity in the U.S., the *Sheltered Instruction Observation Protocol* (SIOP). Polat and Cepik (2016) describe SIOP as a planning tool that offers a framework that guides teachers to deliver, assess, and evaluate those "instructional practices that can help ELLs [English Language Learners] attain English proficiency and achieve academically in content areas" (817). According to Daniel and Conlin (2015) this approach integrates language support into the English content classes rather than offering support in the weaker language outside of class. For Daniel and Conlin (2015):

> The goal of the SIOP model is to prepare teachers in helping ELLs to navigate the dual challenges of learning subject-area skills and content and learning language through building students' background knowledge, making content comprehensible, and attending to other key components of sheltered instruction. (p. 171)

Daniel and Conlin (2015) citing Echeverria, Vogt, and Short (2008), list the main components of SIOP as the following: content and language objectives are reflected on meaningful activities and material; students' background knowledge is built on the basis of prior knowledge and vocabulary growth; comprehensible input is provided; scaffold techniques and learning strategies play a key role in the classroom; students' interaction is encouraged; clear application and put into practice of content and language knowledge is crucial; well-paced instruction and student engagement are necessary; and vocabulary and content review sessions, as well as student comprehension assessment, are crucial. In general, we find that the model attempts to provide full access to knowledge to all students in the classroom. All this is done in a context that remains true to the basic ideals of dual education that will be described in the following sections.

For SIOP, while some argue that there seem to be challenges to overcome in terms of interpretation of the approach (Daniel and Conlin, 2015) or put into practice of "theoretical and operational foundations" (Polat and Cepik, 2016), there is evidence that improvement is possible and is already taking place, at least at the level of teacher's attitude toward students in this kind of programs (Song, 2016). SIOP, as a newer addition to Dual Language Education, serves as evidence that the model continues to evolve and grow.

Considering the background information presented above, we move now to analyzing further characteristics and basic tenets that have served as cornerstones for DLE as it has developed through history.

2.2.2. Goals and characteristics

As mentioned above, one of the highly distinguishing features of DLE programs is that it targets students from two different segments of the society or specific community where it is implemented. These participants have two different language backgrounds. On the one hand, the program engages learners from

language minority groups (i.e., Spanish, Korean) who typically represent minority populations in the area. On the other hand, it enrolls students from the majority language (i.e., English) who tend to embody the dominant social group and dominant language in the area. Both groups of students are integrated into a classroom where they receive literacy and content area instruction through both of their languages. In this same vein, for de Jong (2014), a DLE program "distinguishes itself by enrolling native or fluent speakers of each of the two languages in the program and by emphasizing the benefits of social and academic integration of diverse student populations […]" (p. 241). The social benefits of DLE education tend to be highlighted as a distinguishing factor for this model. Furthermore, Tedick (2014, p. 160) insists that an equally distributed student body in the educational setting is ideal to stimulate "meaningful student interaction between the two groups of learners, provides exposure to peer 'native' language models, and promotes positive intergroup relations (de Jong & Howard, 2003)." This contact with ideal language speakers that serve as language models also favors DLE programs when compared with alternative methods of bilingual education.

Baker (2011) offers a detailed description of what Dual Language bilingual schools may offer to their students:

> The **mission** of Dual Language bilingual schools may also be couched in terms such as 'equality of educational opportunity for children from different language backgrounds', 'child-centered education building on the child's existing language competence', 'a positive self-image for each child', 'a community dedicated to the integration of all its children', 'enrichment not compensatory education', 'a family-like experience to produce multicultural children', and 'supporting bilingual proficiency not limited English proficiency'.
>
> The mission of all Dual Language schools (compared with mainstreaming) is to produce **bilingual, biliterate and multicultural children**. Language minority students are expected to become literate in their native language as well as in the majority language. At the same time, majority language students should make 'age-relevant' progress in their first language and in all content areas of the curriculum. (p. 225) (emphasis in original)

While all the benefits mentioned above are likely to take place in dual language settings, some of them are more frequently mentioned than others. The DLE model pursues specific, clearly identified objectives. Across the literature, three major goals are traditionally listed as fundamental in DLE as it brings together

children from two different language backgrounds. Namely, first, it gives students the opportunity of achieving high quality academic knowledge in two different languages. Second, it offers children the opportunity of becoming bilingual and biliterate as they advance through the different grade levels. Third, it promotes and further develops cross-cultural understanding and positive attitudes that fight racism and foster social change while developing multicultural appreciation (Freeman, 2000; Torres-Guzmán et al. 2005; Parkes, 2008; Lindholm-Leary, 2012). These goals, the first two in particular, represent a great historical breakthrough for students from minority groups who have traditionally been in disadvantage but are, now, offered highly prestigious instruction through DLE programs.

2.2.2.1. Language separation

Certain common characteristics serve as bases to describe dual language programs. First, DLE promotes strict language separation, what Baker (2011, p. 226) calls "language separation and compartmentalization." This implies a strict division of language use depending on different factors. First, Baker (2011) describes the language division according to time. According to this division, the model alternates languages between lessons, days, weeks or semesters. Special attention should be directed to carefully dividing time so that students are equally exposed to each language and can achieve the set goals in each of these languages. When alternating languages, care should be given to a balanced distribution and a precisely timed switch over to the other language when the phase is due. These options are possible as long as a clear-cut boundary is drawn between the two languages.

Second, language separation, as implemented by bilingual teachers, is also required. Bilingual teachers are expected to maintain the language students identify them with by avoiding switch between languages. It is believed that this would motivate students to respond to the teachers according to the language these commonly use. In the event of shortage of bilingual teachers, pairing of teachers could take place. In this case a Spanish-speaking teacher and an English-speaking teacher use their respective language when teaching their subject (i.e., Science, Art) to the same class. According to this idea, mixing of languages is to be avoided and switching languages within a lesson, especially by these bilingual teachers, is seen as having detrimental effects on students.

Third, language separation by content curriculum takes place when certain subjects are taught in specific languages. A possible alternative is to follow *language day* patterns in which given days of the week are associated with a given language (i.e., Spanish is used on Mondays and English is used on Tuesdays). While this alternative does not fix a language to a subject it warrants a rotation of all the subjects in relation to the language, ensuring exposure to a great variety of language forms across different subjects. The *language day* usually alternates between weeks so that specific subjects are not always taught in the same language. This language separation could be done by assigning a given language to a particular subject (i.e., social studies in Spanish, mathematics in English). Nowadays, some flexibility regarding these language boundaries is expected and called for by several authors. Baker (2011, p. 229) criticizes DLE that enforces a very strict language separation and characterizes it as "dated, difficult and unreasonable." Palmer et al. (2014, p. 759) condemns this "monolingual notion of bilingualism." They insist that this strong language separation "is founded upon the outdated notion of languages as separate systems, and bilingualism as dual monolingualism" (p. 759). As discussed in section 1.1.1.2, there are well-supported theories that argue in favor of cross-linguistic transfer of information; namely, Cummins' Common Underlying Proficiency (1980, 1986, 2008) and Interdependence Hypothesis (1979, 1986, 2000, 2005). On account of these theories and considering the existence of a base of cognitive and academic knowledge that underlies academic performance and which is evident through transfer of proficiency from L1 to L2, such strict separation of languages is uncalled for and may even result in detrimental effects for students.

2.2.2.2. The 50:50 and 90:10 models

A further language division according to time is actually another fundamental, key characteristic of DLE. Baker (2011) insists that this is a feature that distinctively characterizes each DLE program on its own. Two models, nonetheless, are typically present across the U.S.: the 50:50 and the 90:10 models. For de Jong (2014, p. 242) "[b]oth models can be effective when they are well-implemented [...]" These two models are described by Baker (2011) as follows:

> The two main models in the US are 50:50 and 90:10. In the 90:10 model, 90 % of the instruction is in the minority language in the kindergarten and 1st grade, with

> 10% to develop English oral language proficiency and pre-literacy skills. Over the remaining elementary grades this ratio changes to 50:50 (e.g., by the 4th to 6th grade). Students often begin formal English reading in the 3rd grade. But from the 1st grade, they are exposed to English literacy more informally.
>
> In the 50:50 model, a 50%-50% balance in use of languages is attempted in both early and later grades. Variations between 90:10 and 50:50 are possible where the minority language will be given more time (60%, 75%, 80%), especially in the first two or three years. In the middle and later years of schooling, there is sometimes a preference for a 50%-50% balance, or occasionally more accent on the majority language. (p. 227)

Lindholm-Leary (2012) acknowledges these two possibilities for models and argues that, besides this distribution of languages used for instruction, another major difference of DLE lies in the language used to teach reading. Lindholm-Leary (2012) explains that, in the 90:10 model, reading is first taught in "the partner language" (i.e., Spanish, Korean) for both native speakers of this language and English-speaking youngsters. For the 50:50 model there is a lot more variation. While in some schools children learn to read first in their corresponding primary language (Spanish and English respectively), at other schools students develop reading skills in the two languages simultaneously. Authors like Castro, Paéz, Dickinson, and Frede (2011) identify reading as a basis for academic success, thus the importance of a close follow up on the reading process in DLE. Castro et al. defend the critical connection between phonemic awareness and decoding skills, key elements to support oral language proficiency, which is, in turn, basic for subsequent reading ability. Castro et al. (2011) also characterize the process of language and literacy development in dual language learning as one that "involves the integration of component skills (e.g., sound-symbol awareness, grammatical knowledge, vocabulary knowledge), as well as more elusive sociocultural variables. Bilingual learners can and do develop second language literacy while acquiring second language oral proficiency" (p. 16).

2.2.2.3. Program coverage

One more distinctive feature of DLE has to do with the extent of the program in each school setting. That is, whether the program is a *strand* (or stream) program or if it is a *whole-school* program, the former appears to be a lot more

common, according to the literature, than the latter. Torres-Guzmán et al. (2005) describe school-wide implementation of DLE as rare, as most of the schools in their study implemented strand-DLE. Palmer (2010) describes a strand program as one that

> […] is situated in an English-language mainstream public school much as a transitional bilingual education program might be, with one classroom out of two or three at every grade level dedicated to Spanish–English dual-language instruction, and the other classes conducted entirely in English. (p. 95)

The previous citation raises a delicate issue by comparing DLE with transitional programs, which are known to aim at preparing students to become part of mainstream groups, while many times sacrificing their L1. This is clearly an idea that works opposite to what seems to be the true purpose of DLE, adequate development of two languages. Moreover, on the subject of strand vs. whole school programs, de Jong (2014), for example, attributes this *design issue* of strand *versus* whole school implementation as carrying important effects on the outcomes of the program in general. Also in her study, only 97 out of 441 programs were school-wide programs. For her, the overall environment of the school may affect the mission and organization of the program as well as the relationships of the different programs within the school. Further rationale is provided by Hernández (2015) who argues "[t]he most common programs exist as TWBI [Two Way Bilingual Immersion] strands within English-medium public schools, while others may offer a school wide approach" (p. 102–103). As can be gathered from the last set of ideas, the issue of the extent of the program in a given setting is another recognizing feature of DLE.

2.2.2.4. Other key characteristics

Other sensitive issues that may exert more influence on DLE than in other sorts of bilingual programs have been identified. Freeman (2000), for example, points to context as a very influential factor in DLE: "The particular structural, sociolinguistic, and ideological context in which the dual-language program is situated influences how this bilingual program functions on the local level and challenges the dichotomous thinking that characterizes most discussion of bilingual

education" (p. 202–203). She describes how the socio-economical background of students is not always limited to the typical dichotomy of majority vs. minority population; or how the variety of levels of proficiency changes from one program site to the next. According to Freeman, features such as these give context a more prominent role in DLE when compared to other types of bilingual education programs.

Furthermore, Torres-Guzmán et al. (2005) direct attention to the decision-making parties involved in processes that affect DLE programs and how the way these individuals value the student members of the speech communities in the program may be reflected upon the programs that are offered to this very population. The issue of value and perception of students is critical in this type of education, considering that minority groups, often representing segregated groups, are a fundamental part of the program. Torres-Guzmán et al. (2005) insist:

> […] the inherent flexibility in decision making can be used to respond more sensitively to the different educational needs of students within a given community and can serve as a space where the transformational values held by school personnel, administrators, and teachers may be brought into the schools. (p. 455)

Further information is provided by Parkes (2008) about the characteristics of the parents participating in these programs. In his study, Parkes (2008) set out to determine the type of families that select DLE programs and the reasons behind this decision in the Southwest of the U.S. He determined that, from the participant families in the study, the majority use Spanish with their children (54.6%) while the second group communicates mainly in English with them (45.4%). Also, 34.2% of the parents are Spanish dominant, 24.7% are English dominant and 40.5% are bilingual parents. As of the educational level of parents, 50% have high school education while 32.1% have undergraduate and 17.9% graduate degrees. These results evince interesting parallels between a clear Spanish or bilingual speaking caretakers and an evidently limited higher education experience.

Second, in terms of objectives; Parkes' results show that the parents' main reason for choosing this model is that they want their children to speak, read and write in two languages. Other reasons include a desire for their children to succeed in globalized societies and at the school level and for children to relate comfortably with different people. All parents (Spanish speaking, English speaking and bilingual parents) in this study expressed a strong interest in their offspring to

become bilingual and biliterate. For Parkes (2008), the distinctiveness of family characteristics and the particular expectations they have on DLE programs deserve special attention when designing this type of program. Based on this study, we can conclude that the make-up of the family and community in this program, as it is in other DLE programs, is very different. These characteristics certainly affect the programs and their outcomes.

From the above we can gather that DLE is affected by very particular factors that may not play a prominent role in other types of bilingual programs. These distinctive features contribute to their essence and uniqueness.

2.2.3. Key factors for successful programs

Lindholm-Leary (2005) (also in Howard et al., 2007) identified a list of key factors that are consistently found in Dual Language educational settings and which are associated with successful outcomes in these programs. She argues that these features identify effective language programs, in general, and dual language programs, in particular. Lindholm-Leary further insists that each one of these eight factors, as key elements determining the singularity of each program, should be analyzed through the optic of the context in which each program is situated. She maintains: "[u]nderstanding these features can help young programs mature and more experienced programs develop into a program that promotes more successful outcomes in students" (Lindholm-Leary, 2005, p. 44). These factors are described in depth in Lindholm-Leary (2005) and will be briefly summarized in this section.

2.2.3.1. Assessment and accountability

For Lindholm-Leary (2005), the idea of assessment is consistent with the notion that all students who participate in DLE should reach a high standard of education. From this, it follows that assessment is essential to measure the progress in students' performance and the achievement of the objectives of the different programs, mainly dual language programs. Following assessment results, programs, teachers, students and other actors involved in the process would be held accountable for their performance. Lindholm-Leary (2005, p. 10) argues: "[d]ual language programs require the use of multiple measures in both languages to assess

students' progress toward meeting bilingual and biliteracy goals along with the curricular and content-related goals." She adds that the data collected through assessment should undergo scientific analysis for these results to provide accountability and to allow for improvement. Moreover, she recognizes that assessment and accountability should portray a number of essential features. Namely, be consistent and systematic, monitor program effectiveness, be aligned with true standards, be geared toward the goals of bilingualism, biliteracy and multiculturalism; have multiple measures in both languages; and be scientifically interpreted. Furthermore, for the author, the results that can be derived from these data should report on students' progress, track students' progress over time, aim at staff advancement and be distributed among stakeholders for subsequent analysis. Assessment and accountability are, thus, seen as valuable tools that can effectively inform about the progress of a program as well as appropriately mark the way toward improvement in any given setting.

2.2.3.2. Curriculum

For Lindholm-Leary (2005), both the features of curriculum itself as well as those of the planning that goes into curriculum are crucial for the design and implementation of a program. As such, the following features of curriculum are of critical importance. Curriculum should correspond with the criteria and assessment that are sought after in the program. It should challenge students; this could be attained through activities that demand higher order thinking skills from students, skills such as the ones described in Bloom's Taxonomy. It should integrate technology, and its topics must be meaningful and interconnected. It should correspond to an *enriched* type of program, one with a long-term effect on students. The idea of bilingualism, biliteracy and multilingualism must be reflected throughout the core features of the curriculum. Finally, it should further be a reflection of the students' cultural values and their characteristics. For Lindholm-Leary (2005), planning related to curriculum should involve vertical and horizontal correspondence across its components. It ought to take into account materials that are varied in genre and presentation (video, print, audio; referred elsewhere as *multimodal* materials); and like the rest of the curriculum, they should aid bilingualism, biliteracy, and multiculturalism goals for students.

2.2.3.3. Instructional practices

Lindholm-Leary (2005) provides a long list of important features in this strand and insists: "good instruction is even more complicated in dual language programs because of the added goals of bilingualism, biliteracy, and multicultural competence, and the constant need to integrate and balance the needs of the student groups" (p. 14). The features that describe this factor are concentrated in the qualities that instruction should have in the program, in general, as well as in those describing input in the classroom. For Lindholm-Leary, the program should target a variety of learning styles and proficiency levels., while fostering positive interactions and genuine dialogue between peer students and between professors and students of the different social groups interacting in this setting.

The instructional practices should promote cooperative learning where group work shows shared, common, objectives; and individual work seeks social equity and accountability while both aim at bilingualism at the same time. In terms of language input, instructional practices should include *sheltered techniques* (visuals, modeling, various presentation strategies) that allow negotiation of meaning. They should be challenging, plentiful, interesting and relevant. Students should be offered controlled as well as flexible tasks that aim at highly proficient oral skills. Moreover, program activities should address all students' needs equally; and while students are integrated for the majority of instruction time, lessons will be delivered in only one language at a time. If we analyze this factor, we can conclude that here is where planning meets reality. Instructional practices are the contact point where planning (at macro and micro levels), and students in the classroom meet in the process. Carrying such weight justifies the key role that Lindholm-Leary assigns to these instructional practices.

2.2.3.4. Staff quality

The characteristics under this factor mainly describe the role of teachers in the model. Lindholm-Leary (2005) contends that well-qualified teachers are a chief component of a successful program. To begin, teachers should be certified, and they should excel at their curricular subject as well as at knowing the features of the type of model they are immersed in. Along these features, they need to possess good techniques for classroom management and instruction. In addition, they should be

familiar and have experience with the characteristics of their educational setting and their students' backgrounds. They should have bilingual teaching credentials and be savvy on language acquisition's best practices. Teachers are expected to be native speakers or show native-like proficiency in the language of instruction, being bilingual and fully biliterate. Lindholm-Leary (2005, p. 21) warns us that the latter should never be the sole reason for a teacher to be part of a program, as "one cannot assume that because a teacher has a bilingual credential that s/he has current knowledge, understands, or supports the dual language program." This type of language proficiency is required to ensure "cognitively stimulating instruction" (p. 21), but is not, by any means, the only distinctive feature of teachers in dual programs. Lindholm-Leary (2005) maintains that monolingual, English-speaking teachers are often part of DLE programs. These teachers should have the ability to understand *non-English* language coming from students, especially at initial levels of the program, with the purpose of meeting the students' needs from the beginning.

2.2.3.5. Professional development

This factor describes features not only of teachers, but of other members of personnel as well. Training for administrators and teachers should come from various fronts. For Lindholm-Leary (2005), to begin, personnel should be knowledgeable in dual language models, their tenets and distinctive features. They need to be familiar with theories that deal with bilingualism and second language development, as well as biliteracy. They should also know how to deal with content subjects that contain and reflect the basic goals of bilingualism, biliteracy and multiculturalism of the model. Moreover, training in equity matters is crucial to promote high achievement expectations, especially in the group that represents the minority or disadvantaged population. It is essential to keep in mind that "[a]long with the training of teachers, training of staff is an important component to a successful program. An effective program cannot have office staff who only speak English if a significant number of parents do not speak English" Lindholm-Leary (2005, p. 24). Given that many times parents' first contact the program through administrative channels, it is necessary for parents to have a way to communicate and learn about the program in cases where they are not fluent in English. We can see the value of this factor as it demonstrates that if all members of a program are targeting a similar objective, they all should be equally involved in pursuing it. A

program would be perceived as stronger if the public can recognize that they represent a united front within the school.

2.2.3.6. Program structure

For Lindholm-Leary (2005), the features conforming this factor are divided into five segments. First, a strong program fosters a clear view of the DLE model and sets the goals in a way that they permeate all activities in the school. The program structure has a clear focus on bilingualism, biliteracy and multiculturalism, evident through high achievement expectations for students. Second, students, parents and teachers are treated equally. They enjoy an organized and safe environment that facilitates learning in an active community interested in the dual language learning process. Participants also enjoy the resources and professional support that encourage additive bilingualism as an outcome of the process. Third, leadership comes from the school principal, the program coordinator and the team in charge of managing the model at the local level. This means that there is permanent communication with the central administration, which is in charge of tracking the development, planning and coordination of any given program. These leaders supervise staff unity, collegiality and development as well as funding.

Fourth, the language educational model supports second language development, theory and research based on bilingualism. A fruitful program structure fosters appropriate instructional and classroom practices and commits to these Dual Language educational practices. Finally, for Lindholm-Leary (2005), an ongoing planning program should be focused on developing the basic goals of dual education and they should be reflected in all curricular areas. We can conclude that a permanent planning program supports practices that guarantee proficiency levels across all linguistic areas. Different actors are involved in the design and implementation of a good program structure; active communication and support seem to be key to obtain the desired outcomes.

2.2.3.7. Family and community involvement

According to Lindholm-Leary (2005), dual language education programs ensure family and community involvement through the program itself and through

liaison parents. From the perspective of the program, it implements several activities that warrant home/school connections that create a welcoming environment for parents. DLE programs posit great value on bilingualism, biliteracy and multilingualism. This is also reflected in practices such as making announcements and posting signs in the languages that are part of the program and in the hiring of staff that speak the non-English language. As to the liaison parents, these are bilingual speakers who maintain contact with parents from both language groups, organize parent-training sessions, know about theoretical issues of dual language programs and contribute to other parent-related issues when required. The key is parent involvement because "[w]hen parents are involved, they often develop a sense of efficacy that communicates itself to children with positive academic consequences" (Lindholm-Leary, 2005, p. 40), which of course translates into beneficial consequences in the program itself. We can assume that, given that the mere existence of the program reflects the needs and interests of the community, this factor is of capital importance for an effective program to take place and for the program's subsequent accomplishments.

2.2.3.8. Support and resources

The impact that support can have on a program can be one element that really makes a difference in the effects that any given program can have. Lindholm-Leary (2005) maintains "[t]he support a school receives influences its funding, materials, teacher training, program model, planning, parent involvement—and thus ultimately student achievement" (p. 42). For Lindholm-Leary (2005), support that guarantees a successful program comes from various sources, at various levels. First, at the administrative level, administrative support derives primarily from the school district and the local Board of Education. This source of support is fundamental for funding allocation and structural as well as functional guidance for the program. A second source of support derives from the community and the school administrative authorities. They need to show a welcoming attitude and perception toward bilingualism, in general, and dual programs, in particular. Without their support, programs are not likely to last long, that is, if they start at all. The third source of support lies at the school level. The principal plays a key role in integrating every member of the school as an active participant in the program. Making sure that the staff understands and supports the model, finding and allocating the resources that the program needs to function accordingly, and

creating a thriving atmosphere at the school setting are some of the roles of the school's principal. He needs to ensure acceptance of the program and the personnel that implements it, and he should make an effort to provide materials that lead to the attainment of the goals and objectives. Finally, the last source of support is located at the family level. Family support is essential as a source of critical advocacy that helps maintain the program working in the school, even at times when the program may seem inadequate for other stakeholders or authorities in the area. Once more, we are reminded that the interest in dual language learning springs mainly from the families in the communities.

While acknowledging the undeniable singularity of each program and the effect that context and language specific features have in determining the distinctive features of any given DLE program, Lindholm-Leary (2005) thoroughly characterizes successful programs from a basis of academic research in the area. The elements described above, summarize decades of experience with successful DLE programs and serve as a guide that ensures future positive attainment of programs if these are implemented correctly.

2.2.4. Alleged benefits of DLE

Many positive characteristics have been ascribed to DLE programs. In this section, we will present several studies that refer to undeniable benefits that have been associated with DLE. To begin, Freeman (2000) makes reference to the social effect that this type of education can have. For Freeman (2000) a positive attribute of DLE is that it challenges the traditional view of monolingualism by which minority language students were expected to learn English and become part of mainstream education to receive high quality education. Freeman (2000) argues that by promoting minority language learning at school, DLE raises the status of the minority languages that are part of these programs and "contest the legitimacy of monolingualism in Standard English as the unquestioned norm for students in mainstream U.S. schools" (p. 207). Torres-Guzmán et al. (2005, p. 455) further credit this model as *socially worthy* given that it directly promotes a more *inclusive* society that fosters bilingualism as well as diversity among its members. More importantly, for Torres-Guzmán et al., it challenges the traditional constraints assigned to the social status of minority groups while broadening the possibilities of *equity and inclusion* of these individuals as valued members of the social group.

A number of academic studies have evidenced the effects that DLE has on students. Lindholm-Leary and Block (2010), for example, contend:

> Data are consistent in showing that Hispanic students participating in dual language programs in predominantly Hispanic/low SES schools achieve at similar or higher levels compared to their mainstream peers in tests of English. In addition, students achieve above grade level in assessments in Spanish. (p. 43)

These authors further maintain that, in their study, both groups of students (Spanish dominant and English proficient) appear to exhibit advantageous results in areas such as language arts and mathematics. Not only are these students performing well in English, but they are also excelling in Spanish tests. Lindholm-Leary and Block conclude that this evidence points to the idea that DLE programs are a contributing factor to closing the academic performance gap between Hispanic and non-Hispanic groups, even in segregated settings.

Furthermore, Lindholm-Leary (2012) offers a summary of findings from different studies where she enumerates the following benefits of DLE. First, DLE students show equal or above grade performance on English standardized tests on reading and mathematics. Second, when compared to peers across state, they show similar scores at around grade 5-7 or earlier. Third, achievement gaps between English language learners and their (English-only class) English native speakers counterparts close at around 5th grade. Fourth, when measured in the partner language, DLE students obtain equal or above grade levels in reading and mathematics. She insists that these findings apply to groups whose counterpart language is Spanish, Chinese and Korean and that these same types of results could be found at secondary levels. In addition, Lindholm-Leary (2012) maintains that students attain high levels of proficiency on both languages; that students who are English language learners are typically assessed as proficient in state tests; and that these results are extensive to populations coming from "different types of communities (urban, suburban, rural) and socio-economic backgrounds (high, medium, low income communities); and with students of different ethnic, linguistic, socio-economic, and special education needs" (Lindholm-Leary, 2012, p. 258). These ideas find support in Marian, Shook, and Shroeder (2013), who insist that DLE brings multiple benefits to educational programs. Marian et al. (2013, p. 182) found that both, minority and majority language, students "show improved math and reading performance on standardized tests in English" while they "stand

to gain proficiency in both languages of instruction." They identify the latter as a desirable qualification in today's globalized society.

On a different note, Ruiz (2012, p. 153–154) offers examples of how the "high expectation curricula" that takes place in DLE is an effective practice that promotes "high level of interaction, authentic communicative contexts, a focus on comprehensible input and output, primacy of students' background knowledge, and early introduction of print […]" Moreover, she describes interactive journals as a tool that has been used to put these elements into practice from very early stages of the learning process in DLE classrooms. In her study, Ruiz (2012) insists that dual English language learners in the U.S., particularly the ones from minority language groups, have an array of features (derived from their life experiences) setting them apart from monolingual English students; thus, they should not be compared to the latter. For Ruiz, faced with the task of learning English while performing accordingly in content areas, dual language learners need and do learn at a faster pace and demonstrate superior gains within a limited number of years, which leads them to succeed in the process. Ruiz (2012) concludes that this is achieved through enriched programs that implement high expectation curriculum; one of these programs being DLE, which provides "instruction tailored to the unique resources and needs that English Learners bring to our classrooms" (p. 159).

Palmer et al. (2014), based on close analysis of successful dual language teachers, describe a number of behaviors that, they found, favor a translanguaging pedagogy (see section 1.1.1.5. above) that, in turn, supports authentic bilingualism. First, the effective teachers in their study served as models of dynamic bilingualism. These teachers implemented language practices that supported language and content learning in students. Second, regardless of the language competencies of students, efficient teachers treated students as competent bilinguals since the beginning stages of their language process. By treating students as bilingual speakers who are academically competent, these teachers reassured bilingual competency development in students. Third, effective teachers enthusiastically acknowledged and emphasized students' metalinguistic remarks associated with cognates, grammar, and structure, for example. By so doing, they assigned value to the students' knowledge and allowed this knowledge to become part of the growth in understanding of both, English and Spanish, in the classroom.

This strategy once more supports Cummins' theories (1986, 2007, 2009) in regard to L1 transfer of knowledge into L2 on the basis that there is a Common Underlying Proficiency that functions interdependently to support the development of both languages. Palmer et al. (2014) conclude that these practices evince how

teachers are crucial in allocating a key role of students as bilingual resources in the classroom. They insist that these practices promote and develop strong bilingual personalities that strengthen dual language practices. As this excerpt illustrates, good end results in a program are due to a combined source of efforts, one of which is the key role of teachers in the classroom, as contended by Lindholm-Leary (2005), mentioned above.

Lindholm-Leary and Genesse (2014) present the following results that tap into effects of various forms of bilingual education; special emphasis is given to the outcome of DLE programs. First, students enrolled in bilingual programs attain the same or better results than students enrolled in mainstream programs. Second, majority students (i.e., the largest group of students regardless of their L1) develop regular levels of proficiency in the L1 and advanced levels of proficiency in their L2, L3 or L4. Third, students in two-way immersion and developmental bilingual education programs, while becoming highly proficient in their L1, are likely to be highly proficient in the dominant language at a rate similar to that of their counterparts (students in mainstream programs). Fourth, students from various, ethnic, socio-economical, and learning-challenged groups enrolled in bilingual programs benefit (academically and in terms of L1) equally or more than students in mainstream programs. Fifth, while intensive exposure to the dominant language does not translate into higher proficiency and achievement in that language for minority of majority students, intensive exposure to the minority language results in higher levels of proficiency for both groups of students. Finally, in general, there is a positive connection between bilingualism and academic achievement for minority and majority language students.

In a similar vein, de Jong (2014) provides a summary of outcomes that are frequently associated with two-way immersion programs. She maintains that the 90:10 model programs appear to be more beneficial to develop Spanish (oral) language abilities. Second, she argues that students in two-way immersion programs perform just as well or better in English tests than students enrolled in mainstream plans. Third, involvement in two-way immersion programs that start at the elementary level exert a long-term, positive impact that can be traced even to secondary school. Fourth, students with low socio-economic status, African-American students and students with special needs perform equally or significantly outperform non-DLE groups.

The studies reviewed show some of the results that have been found after analyzing the effects of dual language programs. Some ongoing research confirms

the results mentioned above and seeks solutions for the problems identified, a selection of which are discussed in the following section.

2.2.5. Challenges of DLE

As is the case with other forms of bilingual education, despite the sounding success of DLE, a number of issues represent future challenges for the model. For instance, Torres-Guzmán et al. (2005) bring forward a critical issue that may have devastating effects on the very core principles of DLE and thus on its main objectives; namely, the erroneous allotment of language to time distribution. In the sample studied by Torres-Guzmán et al. (2005), they found that:

> […] there is a widespread belief that a 90-10 model is implemented as 90% instructional time in English and 10% in the LOTE [Language Other Than English]. In other words, teachers do not know the basic tenets of dual-language programming and are making decisions about medium of instruction in ignorance of what they ought to be doing under the dual-language label. (p. 467)

According to Torres-Guzmán et al. (2005), while teachers are partly responsible for this problem, the authors assign greater responsibility to district, program and administrative parties in charge of creation, labeling and implementation of the programs in the area. These actors, the authors insist, are the ones who should set the basis and constantly assess the programs that are in process in their area. These authors also found that a number of programs labeled as *dual programs* were actually second language and heritage enrichment programs. This situation, Torres-Guzmán et al. (2005) sustain, may compromise the future development and permanence of dual programs. The authors also found that a few teachers (those who work in programs where there is a stronger presence of the minority language) exhibit a better understanding of the foundations of the model. The problem mentioned above calls for better teacher preparation, as one way of dealing with the proper implementation of the model.

Palmer (2010) raises issues of segregation and even racism (in the form of personnel's attitude) present in the dual language program in her study. She describes situations where African American students and their parents have faced undermining comments in relation to their participation in the program, and

students' suitability for it, from personnel at the school. She claims that social workers, in the institution in her study, are credited with comments that discourage African American students from participating in dual programs on the basis of their alleged inability to learn languages. This fact fosters exclusion of this population from such programs while at the same time promotes racism, creating an inequitable learning environment. For Palmer, some teachers in this setting argue that the program does, in fact, exclude students that do need the program, Latino students included. The author insists that the staff's attitudes affect the overall academic efforts of black students in the school. She argues that the situation contributes greatly to behavioral problems in the classroom, leaving students with a clear lack of motivation toward learning in general. Palmer (2010) concludes that for DLE to succeed, the interaction of race and power variables in each school must receive careful attention. These should be dealt with the utmost care so that racism and inequities are eradicated from the learning environment.

Further rationale is given by Paciotto and Delaney-Barmann (2012), who point to the challenges raised by sociocultural and economic factors in rural areas. In particular, they put forward the issue of human resources and teacher availability in these regions. They illustrate this as follows: "[b]ecause of the lack of financial resources and ELL education knowledge, and geographic isolation, rural districts have to 'adapt' the preexistent human resources to fit the needs of changing demographic contexts" (p. 20). Personnel, they insist, go through a learning process; they lack effectiveness, and experience trial and error practices mostly based on their own motivation and personal interest toward their student population. Paciotto and Delaney Barmann (2012) add that a lack of state support results in a shortage of prepared teachers and administrators, an impossibility of language policy implementation, and a concentration of bilingual teachers where the focus of the migratory group is located, all of which are common in underfunded rural areas. We can see how pressing the issue of qualified teachers is in those areas. The situation probably adds this difficulty to others that are commonly found in these contexts, such as school funding or teaching material limitations.

Feinauer and Howard (2014) draw attention to the problem of the lack of accountability established in connection to the "third goal" (cross-cultural competence of dual language programs) and they also offer ways of solving this problem. They contend that, when compared to the other goals of academic achievement and language and literacy, cross-cultural competence has received limited attention and is seemingly viewed as a less important objective. These authors argue that to develop intercultural understanding and cross-cultural

abilities, students need to possess a strong feel of their own cultural distinctiveness. They maintain that this issue may be addressed by being knowledgeable about how the notion of student identity would help understand the students' advancement in terms of cross-cultural competence.

Feinauer and Howard (2014) further assert that knowledge of students' cultural identity development may be approached from different perspectives. First, a developmental perspective can show how the program shapes the students' identity through time. Monitoring whether the program offers opportunities for students to explore their identities over time could do this; the fact that DLE programs are usually long-term programs also facilitates implementation of this type of technique. Second, identity formation may be explored through a sociocultural perspective. Since linguistic interactions have a powerful effect on how we perceive the world and ourselves, these interactions play a fundamental role in explaining identity formation in this sociocultural sphere. If we consider that minority and majority students in DLE programs enjoy different opportunities (language and education wise) than students in mainstream settings, we could predict that they are also faced with different identity choices. Finally, a post-structural perspective views identity as a dynamic, changing phenomenon in which language plays the role of defining and delimiting identity. In the context of DLE, students' choice of language, students' positioning of themselves and the students' positions assigned to them by others may serve as a reflection of students' identity. The authors conclude by insisting on the need for students to develop a strong self-identity as a basis for subsequent attainment of cross-cultural competence.

Furthermore, de Jong (2014) points to the idea that the traditional focus on a dichotomous group as representative of student population in DLE no longer suffices because it leaves out groups such as those who are already bilingual, students with special needs or African-American students enrolled in dual education in urban U.S. She also refers to organizational issues of DLE; in particular, the fact that most of the programs in her study (344 of 441, p. 248) are strand programs rather than school-wide programs. She insists that strong leadership is required to transform ways of thinking and turn strand programs into school-wide programs that support the goals of dual language learning.

Hernández (2015) addresses another sensitive issue in DLE (which she refers to as *two-way bilingual immersion*); that is, the challenges of student-student interaction and teachers' practices in these settings. She draws attention toward the power struggle that takes place in the classroom due to the perception of English as

the language of power in the U.S. English, she continues, is internationally linked to:

> […] status, power, and wealth, largely due to its status as a global language in science, technology, medicine, entertainment, sports, and so on […] the status of language in TWBI [Two Way Bilingual Instruction] settings may be influenced by Spanish- and English-speaking students desiring to conform to the dominant language that is associated with prestige and power in U.S. schools and communities. (Hernández, 2015, p. 106)

If attention is given to the ideas associated with the prestige of the English language referred to above, and to the varied linguistic and stratified society where this interaction is taking place, these findings (Hernández, 2015) are not surprising. She discovered that state and district assessment measures endorse English as the language of prestige and value. She found that both Spanish- and English-speaking children prefer to resort to English during their interaction in the classroom, delegating Spanish to an inferior position. She further determined that English-speaking students, who also do most of the talking, generally lead small-group interactions in the classroom.

Hernández (2015) concluded that whenever the language of interaction was Spanish, English-speaking youngsters would switch to English during these interactions and Spanish-speaking children would conform. For Hernández, teachers frequently reported facing challenges in dealing with students' interaction, particularly when those in the dominant group refused to work with students of minority populations. Teachers also acknowledged certain leniency toward English use. Hernández attributes this, possibly, to teachers' awareness of general policies favoring English as the prestige language and to the difficulty of the Spanish material. He searches for ways to encourage Spanish speakers to participate more actively in small groups and develop non-threatening activities that can empower this population. Better sociolinguistic skills would result in a greater use of Spanish in the classroom. There is also a need for district and state support of the tenets of dual language practices. Hernández (2015) concludes by emphasizing the necessity of consciously strengthening the status of Spanish as the minority language in this scenario. We can see how this plea is justified; this idea alone would have clear effects on the other issues raised in this study.

2.2.6. Future research

As a phenomenon that is alive and evolving, DLE serves as a source of future studies. Several authors have brought forward ideas that can nourish new research. For instance, Castro et al. (2011) argue that further analyses are needed in areas such as language development of English language learners in pre-school; and influence of English immersion programs on children's L1 development and native language development and literacy. Marian et al. (2013) call for studies covering cognitive functioning in areas other than reading and math; and studies exploring the connections between the academic knowledge gained and its transfer to other languages and non-academic contexts. Moreover, de Jong (2014) calls for future longitudinal research that could provide information on contrastive ways of learning for students from different populations; studies on graduation rates, course placement, higher education and career choices of students; and studies on racial, ethnic and linguistic multiplicity as well as on the impact that they have on dual language programs design and operation.

Lindholm-Leary and Genesse (2014) posit many areas in which research could serve to clarify unsettled issues. First, from a methodological standpoint, they say, instructional time patterns at each school setting must be defined more clearly. Programs should also be defined in terms of length and instructional practices to allow for more accessible means of comparison across programs. Second, regarding assessment, instruments that measure students' achievement (in language and non-language areas) need to be created to have access to more reliable, valid information, especially regarding minority students. Third, studies that explore results of students in Chinese and other Asian languages are necessary. Fourth, further research that explores outcomes of students with special needs in dual language programs is required. Fifth, these authors call for research to determine the features that identify high-quality programs and the impact that other types of programs have on students. These examples show how the DLE field has a long-lasting projection toward a very active future.

2.2.7. Concluding remarks

The history of DLE programs shows that the development of the model has not been easy, as it has evolved in a "racialized context" (Palmer, 2010, p. 110),

immersed in a "political context of opposition to bilingual education" (Torres-Guzmán et al., 2005, p. 455). The model, nevertheless, continues to grow, not only in the U.S. but in other regions as well. It has branched out to forms such as the SIOP approach mentioned at the beginning of this section. The evidence shows that when implemented correctly, DLE yields beneficial results in its population. As such, this model "holds promise for replacing segregated, assimilationist, and academically ineffective education in schools with many Spanish-speaking ELLs" (Paciotto & Delany-Barmann, 2012, p. 19). With these ideas of growth and change, we can watch for what the future years of design and implementation of DLE programs will bring. As mentioned above, a new form of DLE, SIOP has emerged more recently. This demonstrates that DLE continues to change and adapt to the needs and challenges of students and society in general.

2.3. Content and Language Integrated Learning: The CLIL phenomenon in Europe[3]

This section provides a description of the *Content and Language Integrated Learning (CLIL)* movement in Europe. This is the last type of *strong bilingual education* (Baker and Wright, 2017) program analyzed here. Information is provided on how the model originated and on the political, educational and grass-root interests in its origin. Then, a series of definitions are presented to illustrate the different perspectives that have informed the model. Key concepts that serve as bases for the model are discussed; namely, the Four Cs Framework (Content, Cognition, Communication and Culture); the Triptych Approach (Language *of*, *for* and *through* learning); and the scaffolding practices that best identify the model. Subsequently, a description of the benefits that are associated with the model is provided. This is followed by a characterization of CLIL in contrast to Content-Based and Immersion instruction. Finally, future challenges that have been identified regarding the model are presented to show possible future lines of study within this model.

[3] As part of the research carried out for this doctoral thesis, earlier versions of several parts of this section have been published in Castro-García, D. (2017). Bilingualism, CLIL, and immersion programs: The case of Costa Rica. In G. Nieto Caballero (Ed.), *Nuevas aportaciones al estudio de la enseñanza y aprendizaje de lenguas* (pp. 123–136). Cáceres: Universidad de Extremadura.

2.3.1. Origins of CLIL

Marsh (2012) provides a thorough, historical description of the steps that led to the establishment and implementation of CLIL in Europe. From the 1958 European Economic Community Council regulation of official status languages within the European Union—although at the moment CLIL was not mentioned as such—to the 2011 European Commission Working Papers that describe CLIL as a key element in education, different participants have set the basis for CLIL's ongoing role and growth today. The process has involved actors at the supra-national level, namely The European Parliament, The European Commission, The European Council and the Education Council. These, relying on their hierarchical position, have developed resolutions, recommended initiatives, produced declarations and formulated objectives that have resulted in the creation and application of CLIL practices in the different Member States of the European Union. According to Marsh (2012), The Maastricht Treaty (1992), another supranational initiative, serves as referent for the focus on education, training and languages that has served as a basis for CLIL. The Eurydice Network, has helped to provide detailed reports on CLIL functioning, implementation and operational factors which are, in turn, based on reports of the different participant States. Working Groups, such as the Group of Intellectuals for Intercultural Dialogue, have contributed different proposals to support the CLIL process and Civil Platforms, like the Civil Society Platform on Multilingualism have also come up with recommendations to help promote multilingualism in the European Union. These different participants paved the way in the forms of national and supra-national resolutions, regulations and proposals that have resulted in what CLIL stands for today.

Additionally, also at the supranational level, Marsh (2012) mentions a number of programs that have provided support along this process. The LINGUA program (European Council, 1989) sought implementation of an improved quality of language learning, initiating the process of search for a more apt teaching model. Also, the LEONARDO DA VINCI program (European Council and European Parliament, 1994) appeared as an action program that reinforced vocational training, particularly on language teaching and learning. The SOCRATES program (European Parliament, 1995) came to encourage forms of mobility and linguistic development through exchanges in the European region, while the CULTURE 2000 program (European Parliament and European Council) sought to preserve Europe's cultural heritage, giving emphasis to teaching and learning of its languages. Finally,

two important documents should be mentioned, given their impact on this process: the White Paper (European Commission, 1995) and the Green Paper (European Commission, 1996). The former's plea for an early start of foreign language teaching and learning as well as for greater flexibility and improvement in education facilitated the development of innovative approaches such as CLIL.

This paper also refers to the 1 + 2 principle, by which EU citizens are called to be proficient in their mother tongue plus at least two other languages of the community. Moreover, the Green Paper's insistence on learning at least two more community languages to have access to opportunities at the occupational and personal level in the EU context also reinforced the idea presented in the White Paper and thus also invited implementation of CLIL-like practices. It can be deduced from the brief description above that there have been many participants in the creation and implementation of CLIL and that while the process seems to have started at the top (institutional) level of society, there has been plenty of opportunity for participation at the parental and school level, and more importantly, yet, that the whole process responds to the need for language instruction provision at the society level.

The CLIL movement finally appeared in Europe in 1994 (Marsh, Maljers & Hartiala, 2001). According to Marsh (2012), the 1993 already existing *Content and Language Integrated Classrooms (CLIC)* concept was transformed by several stakeholders into *Content and Language Integrated Learning (CLIL)* with the purpose of designating and assigning much-deserved importance to the methodology rather than to the context introduced in the former concept. Dalton Puffer (2008) insists that the driving forces behind the implementation of CLIL practices in Europe derived from two very distant focal points: local grass-roots (teachers, parents, specific institutions, and communities) and top-down policies implemented in the European Union like those above. The middle-level groups regulating local education policies were left to find a way to mediate between the rather disconnected focal sources.

2.3.2. Definition of the model

CLIL literature presents many definitions that help construct an idea of the key features that constitute the model. These definitions vary in terms of the emphasis they give to certain features of the model. Some concentrate more on "the skeleton" of the concept (its general structure or stronger characteristics). Others try to depict a more real image of the concept or even its implications in the

classrooms' dynamics. Others yet offer a hint at the political shades that may influence the model in some contexts. Some of these definitions will be presented now. An early description by Coyle (2006) combines the first two emphases more commonly present in definitions across the literature. Some of CLIL's general features and certain possible classroom functions result in a fairly complete definition put forward by Coyle (2006):

> CLIL is a lifelong concept that embraces all sectors of education from primary to adults, from a few hours per week to intensive modules lasting several months. It may involve project work, examination courses, drama, puppets, chemistry practicals and mathematical investigations. In short, CLIL is flexible and dynamic, where topics and subjects —foreign languages and non-language subjects— are integrated in some kind of mutually beneficial way so as to provide value-added educational outcomes for the widest possible range of learners. (p. 3)

While Coyle's definition does not provide specific details of the implementation of CLIL, it helps us create an idea of the levels at which the model could be implemented as well as the form it could take in a classroom, the flexibility it offers for topics and activities, and the types of students it addresses. Hence, following the definition above, readers can develop a pertinent, although general, idea of the concept.

However, as was mentioned above, not all definitions highlight the same features of the model. There are a good number of definitions that concentrate more on delineating the general structure of CLIL and thus offer only a broad idea of the term. See the following examples to illustrate this assertion:

> The term CLIL […] functions as an umbrella not only for a wide array of educational practices but also for an even wider array of terms tied to specific lingua-cultural, national, educational and disciplinary traditions […] the term has acquired some characteristics of a brand-name, complete with the symbolic capital of positive ascriptions: innovative, modern, effective, efficient and forward-looking [… the intention being that] 'CLIL' retains its open nature as an umbrella term for many realities of non-language content teaching through an additional language. (Dalton Puffer et al., 2010a, p. 3)

> CLIL, understood as an approach that integrates language and content, coexists with a plethora of terms that range from the bilingual integration of language and

> curricular subjects, to content-based language teaching, theme-based language teaching, or content-enhanced teaching. (Lasagabaster & Sierra, 2010, p. 367-368)

> The CLIL 'generic umbrella' includes many variants. Some of these may be considered as primarily language teaching. Some can be seen as mainly content teaching. The essence of CLIL leads to it having status as an innovative educational approach which transcends traditional approaches to *both* subject and language teaching. (Marsh, 2012, p. 308)

> [A]n umbrella term broadly covering the central part of [the] continuum between content-driven and language-driven teaching approaches. (Juan Garau & Salazar Noguera, 2015b, p. 3)

All the above definitions seem to, purposely, have the intention of generally describing the scope of CLIL without really providing specifics on how it comes into play in an educational setting. This wide-ranging form of describing the model through expressions such as "an umbrella term for many realities," "coexists with a plethora of terms," "[t]he CLIL 'generic umbrella' includes many variants," "an umbrella term broadly covering" allows for many interpretations and provides partakers with an endless list of possibilities that they could define as CLIL practices. As mentioned above, other definitions focus on more practical aspects of the model inside the classroom context, to an extent that concentrates on the function given to the second, foreign or additional language, depending on the definition, and the role this language is to play in CLIL. These also appear to have the intention of guiding practitioners (to different extents and at different levels) on the implementation of this CLIL subject-specific language in the classroom. The following are some examples of this type of definition:

> Content and language integrated learning - the use of an L2 in the teaching of non-language subjects- […] (Dalton Puffer, 2008, p. 1)

> CLIL is a dual-focused educational approach in which an additional language is used for the learning of both content and language. (Mehisto, Marsh & Frigols, 2008, p. 9)

> The term 'Content and Language Integrated Learning' (CLIL) was adopted in 1994 (Marsh, Maljers & Hartiala, 2001) within the European context to describe and further design good practice as achieved in different types of school environments

> where teaching and learning take place in an additional language […] CLIL is an educational approach in which various language-supportive methodologies are used which lead to a dual-focused form of instruction where attention is given both to the language and the content. (Coyle, Hood & Marsh, 2010, p. 3)

> Content and language integrated learning (CLIL) can be described as an educational approach where subjects such as geography or biology are taught through the medium of a foreign language, typically to students participating in some form of mainstream education at primary, secondary but also tertiary level. (Dalton Puffer et al., 2010a, p. 1)

> CLIL is here understood as an educational model for contexts where the classroom provides the only site for learners' interaction in the target language. That is, CLIL is about either foreign languages or lingua francas [...] CLIL can be described as an educational approach where curricular content is taught through the medium of a foreign language, typically to students participating in some form of mainstream education at the primary, secondary, or tertiary level. (Dalton Puffer, 2011, p. 182-183)

> In the original terms of reference CLIL and EMILE refer to any dual-focused educational context in which an additional language, thus not usually the first language of the learners involved, is used as a medium in the teaching and learning of non-language content. (Marsh, 2012, p. XIII, XIV)

> An inclusive term, particularly used in Europe, for bilingual or multilingual education in which a second or later language is used for learning subject content, and where both language learning and content learning occur simultaneously with an emphasis on their integration. (Baker, 2014, p. 235)

> CLIL is a teaching approach in which an additional language is used for the teaching and learning of subjects with a dual focus on language and content. (Heras & Lasagabaster, 2015, p. 71)

All of the definitions in the latter group aim at establishing a clear distinction between the role of the second, foreign or additional language in the CLIL setting in contrast to the traditional role assigned to this language in conventional, mainstream classroom contexts. These definitions revolve around one same function for the additional language. They either describe it as used for content teaching and learning or in non-language subjects; that is, the second, foreign or additional language is used as a means, and not as the final goal in the classroom,

and it is definitely not seen (at least when it is the CLIL language) as the subject of the class. Finally, other definitions try to convey that socio-political function that the CLIL model serves in some countries or areas inside the European Union. Given the historical, social or political region where CLIL is implemented, CLIL may also serve as a bridge in multilingual territories to create spaces for trilingual regions as well as to introduce languages that the present social requirements demand of citizens.

The definition presented by Lorenzo, Casal & Moore (2009) hints at this function that also applies to CLIL: "CLIL serves as an umbrella term embracing all scenarios and whatever combination of regional, heritage, minority, immigrant and/or foreign languages they involve; providing for a highly diversified language curriculum" (p. 419). Regardless of the focus imprinted by the author defining the concept, CLIL represents a very strong teaching force in European countries and it is now beginning to be considered in the educational contexts of South America as well (see Curtis (2012) in reference to Colombia and Banegas (2016) for Argentina, for example). Since its origins, it has sought to find the best way to form bilingual citizens who can deal with the linguistic requirements set by globalization. Like the rest of the world, when the CLIL proposal first appeared, Europe was looking for a way to respond to the demands of our modern society in terms of educating bilingual people who would be prepared to interact with citizens from diverse linguistic backgrounds. This process still continues today. There are certain characteristics of CLIL that permeate the different definitions listed above. The word *variation* condenses essential features that are part of CLIL. Dalton Puffer et al. (2010a) discuss this variation in relation to how CLIL is implemented in terms of *length* (short or long) or *intensity* (low, medium or high) including all the possible combinations that are derived from these.

From the definitions above we can also abstract that CLIL caters to a varied range of age groups, it houses programs that offer various numbers of hours, it provides opportunities for different types of learners and it allows for the implementation of varied practices and methodologies. Mehisto et al. (2008) insist that CLIL is a flexible approach offering a new perspective which has tried to synthesize and put into practice the knowledge gained through many years of experience in approaches such as immersion, bilingual education or multilingual education. Juan Garau and Salazar Noguera (2015a) add practices such as Content Based Language Teaching and Content Based Instruction to the former list.

In addition, another feature found frequently is the idea of the *dual* focus served by CLIL, in the sense that, as Coyle et al. (2010, p. 35) state: "it is not a question of whether to focus on meaning or form, but rather that it is fundamental

to address both." Appropriate CLIL practices try to find a balance between content learning and language learning in a way that these would only benefit from one another while at the same time avoiding any possible detrimental effects caused by CLIL implementation. This occurs because since CLIL was first implemented, many have been concerned of students not learning the content of the class. For those familiar with CLIL, there is no debate on whether a class should concentrate on language itself or on its specific content, but a clear balance should guide the two. In sum, CLIL is "an umbrella" that protects many methodologies, activities, practices and realities be these personal, political or social; and according to Dalton et al. (2010a, p. 3), also "cultural, national, educational and disciplinary traditions." All of these underlie the constant reference to flexibility and variability in the definitions.

Along these same lines, a word of caution is necessary. Above, variability and flexibility were mentioned as imminent descriptors for CLIL, and although they are indeed appropriate, those features should still be treated with care. Coyle (2007) already warned us:

> Given the diversity, I would argue that such a flexible inclusive approach to CLIL is both a strength and potential weakness. The strength of CLIL focuses on integrating content and language learning in varied, dynamic and relevant learning environments built on 'bottom-up' initiatives as well as 'top-down' policy. Its potential weakness lies in the interpretation of this 'flexibility' unless it is embedded in a robust contextualized framework with clear aims and projected outcomes. (p. 546)

To prevent vague interpretations, Coyle (1999) readily discussed the 4Cs framework and its principles. These are to be kept forefront to prevent misconceptions in the practice. Definitions such as that by Lorenzo et al. (2009), presented above, should be interpreted with care.

2.3.3. The four Cs framework in CLIL

Coyle (1999) developed the 4Cs Framework to represent the intertwined connection that exists between *content, cognition*, *communication* and *cultural awareness* in CLIL-like practices. For Coyle (1999), "it is through **progression** in the knowledge, skills, and understanding of the content, by **engagement** in

associated cognitive processing, **interaction** in the communicative context, and a deepening awareness and positioning of cultural self and otherness, that learning takes place" (p. 53, emphasis in original). The framework originally tried to guide teachers and planners toward "effective teaching and learning in content classrooms" (Coyle, 1999, p. 46) and it cursorily became one very important cornerstone in CLIL. Today it is still used as a planning framework that helps teachers ensure an appropriate balance of the 4C elements in every class. According to Coyle (2006):

> Teachers, learners, trainers and researchers are collectively exploring the interrelationship between subject matter (content), the language of and for learning (communication), the thinking integral to high quality learning (cognition) and the global citizenship agenda (culture) – which constitute four Cs […] From this perspective, CLIL involves learning to use language appropriately whilst using language to learn effectively. (p. 9)

Along these lines, the framework clearly represents a key strength in acceptable CLIL practices given that it serves different aims in the classroom. If we consider that the framework can guarantee that in CLIL classrooms learners can learn content and language at the same time that they communicate effectively with their peers while also developing cultural sensitivity and awareness, then the importance of following the principles underlying the 4Cs framework are undeniable. Coyle et al. (2010) contend that this contextualization of the 4C Framework reflected in careful planning and which shows the evident integration of the 4C elements in the teaching and learning process is what distinguishes CLIL from other approaches or other forms of bilingual education. Coyle et al. (2010, p. 12) argue that "CLIL is not simply education *in* an additional language, it is education *through* an additional language."

2.3.4. The language triptych approach

The quote in the preceding paragraph takes us to another important CLIL principle put forward by Coyle (2006, 2007), known as the *triptych approach*. According to Coyle, the different types of communication and interaction that students implement in a CLIL classroom demand to be viewed differently if these are to be compared to the ones in a regular foreign language classroom, for

example. Coyle (2006, p. 10), insists "if the content determines the language needed in CLIL, then language **of** learning, **for** learning and **through** learning is a more relevant analytical approach to determining the language to be taught in CLIL classrooms […]" (emphasis in original). These characteristics of language in CLIL classrooms may have an effect on the structure of the curriculum because the linguistic needs of the students are different in CLIL contexts. Students require grammar forms and structures that enable them to discuss and interact with peers in ways that are not in vogue in regular foreign language classrooms and that do not necessarily follow the same order used in those classrooms. Coyle (2007) describes the CLIL classroom language as follows:

> Language *of* learning is based on an analysis of the language needed for learners to access basic concepts and skills relating to the subject theme or topic […] an analysis of the language needed to scaffold content learning will lead to a complementary approach to learning progression […] Language *for* learning focuses on the kind of language which all learners need in order to operate in a foreign language using environment. It foregrounds metacognition and learning how to learn. I would argue that the development of teaching strategies to scaffold learning, as well as the development of independent learning strategies, must take into account the language required for both these processes to operate successfully […] Language *through* learning is predicated on the sociocultural tenet that learning cannot take place without active involvement of language and thinking […](p. 553-554)

Coyle et al. (2010) expand on this idea by saying that the language *of* learning is that language required to have access to the information discussed in class. The language *for* learning is the language required to function in the second language, to discuss ideas or interact with other learners, for example. Finally, the language *through* learning is the one that is needed to carry out active involvement both in the language itself and it is also the one required for analyzing both content and language and for *higher order thinking*, by which, according to Bloom's Taxonomy (Bloom, 1956; Anderson and Krathwohl, 2001), students engage in cognitive processes that allow them to use knowledge more dynamically. This vision of language is not limited to how students make use of it inside the classroom, as is usually the case in traditional, foreign language classrooms. It goes beyond the classroom limits by allowing students to take this language with them

outside the classroom, as they get involved in thinking practices that may encompass analyzing concepts and elaborating ideas even after the class is over.

This last function of language in CLIL-like views widens the scope of the use of language that takes place in a traditional foreign language classroom in which, if any, the language *of* learning and the language *for* learning are sought for and used. This stepping away from curricular grammatical progression may demand more careful planning on the part of the professors as they must be ready to anticipate students' linguistic demands and to provide the language forms that students may need to function on different subjects in the CLIL classroom. This is where the idea of *scaffolding* becomes absolutely necessary.

2.3.5. Scaffolding: A key element in CLIL

Scaffolding techniques and scaffolding language embody another key feature in CLIL. Given that students face more linguistic and even content related challenges in this type of educational context, scaffolding becomes a key player in the classroom field. CLIL teachers, thus, need to facilitate access to students to either the new language or new content that students are receiving in class. In these settings, it is the teacher's role to create a connection between what students know and the new information they are receiving. According to Baker (2011), who defines scaffolding following the ideas of *zone of proximal development* put forward by Vygotsky (1962) and Bruner (1983), scaffolding can be described as the type of language that is present in a teacher-student, cooperative-classroom relation where "the teacher supports the student by a careful pitching of comprehensible language […] This is achieved by the teacher moving from the present level of understanding to a further level that is within the child's capability" (Baker 2011, p. 294-295).

Meyer (2010) also describes multiple benefits brought about by scaffolding techniques. Besides the evident access to classroom material, balanced scaffolding provides students with access to different types of input; it can be regulated or reduced as students' language levels advance; it enables students to carry out tasks in the classroom; it supports classroom production by helping students verbalize their ideas; it enables weaker-language students to benefit more from the class. Overall, he concludes, scaffolding enhances students' proficiency in the cognitive academic language. According to Meyer (2010, p. 14), "[t]o successfully deal with multi-modal input, students need to have a wide variety of study skills at their

disposal which makes the scaffolding of language and learning a key component of successful CLIL teaching."

Moreover, having access to language and learning techniques that implement scaffolding allows CLIL students to activate metacognition and engage in more advanced, critical thinking (Coyle et al., 2010). For Marsh (2012), academic language deserves particular attention in CLIL settings as it should be made part of the lesson through scaffolding to ensure that the learning outcomes are the best possibly expected. Harvey, Tihinen, Määttä, and Uusiautti (2013) argue that scaffolding holds a special relation to the cognition principle of CLIL. For these authors, scaffolding allows students to engage in more elaborate thinking, which would eventually give students access to deep levels of analysis and understanding that would translate into cognitive rewards.

Being such a critical component of a CLIL setting, scaffolding needs to be addressed with very careful attention; especially if we consider that there are usually multiple language levels in a classroom and they all deserve to be acknowledged. If CLIL practices really aim at providing all students with equal access opportunities to learn the foreign language, teachers should pay special heed to students with lower linguistic proficiency so that they receive the language they need through scaffolding techniques that would enable them not to miss out on learning. As Harrop (2012) puts it, weaker learners may be put in a vulnerable position if scaffolding does not meet their needs. As can be deduced from the discussion above, CLIL attempts to do quite the opposite, and several studies (see Dalton Puffer, 2008, 2011, and others mentioned in the next section) present evidence of the benefits that CLIL brings to the process.

2.3.6. Alleged benefits of CLIL

In this section, I will refer to some of the evidence that has been gathered in connection to the beneficial effects of CLIL. As can be seen below, there are multiple benefits that have been linked to CLIL practices, these range from a more fluent use of the language to stronger vocabulary, from higher levels of motivation to intercultural awareness. To begin, Coyle (2006) contends that if qualitative class time is spent in a foreign language, this would result in increased linguistic competence (see Marsh 2012, Juan Garau & Salazar Noguera 2015a). For Coyle (2006, p. 6), CLIL offers students the opportunity to engage in "problem-solving, risk-taking, confidence building, communication skills, extending vocabulary, self-expression and spontaneous talk […]" through the combination of content and

language learning. She argues that the teachers and learners' sense of belonging to a learning community is just one of the outstanding outcomes of CLIL.

Dalton Puffer (2008) refers to several benefits identified in CLIL. For Dalton Puffer, CLIL students equate or even outperform non-CLIL students' results on content tests; they are more persistent to get tasks completed and are not as easily frustrated, thus developing stronger procedural competence; and their communicative competence also evinces their higher levels of L2 learning. She also notes that the naturalness of the classroom and its reduced focus on form allow students to enjoy a more relaxed learning environment. This author insists that while some language areas seem to be unaffected by CLIL instruction (i.e. syntax, writing, informal/non-technical language, pronunciation and pragmatics), receptive skills, morphology, risk taking, vocabulary (especially for technical and semi-technical terms when addressed explicitly, see Dalton Puffer, 2008), creativity and fluency quantity are greatly benefited (see Baetens Beardsmore, 2008) in reference to these last two). In a later study, Dalton Puffer (2011) stresses that CLIL students' strategic competence enhances them to convey content notions accurately, even at early stages of the process and despite possible limited linguistic resources.

Furthermore, this author identifies two advantages of CLIL educational settings: "the didactic nature of the interaction and the cultural familiarity with the domain of use and its rules" (Dalton Puffer, 2011, p. 195). This predefined, pre-planned language learning that seems to be disguised in a content learning class that takes place in the students' L1 culture, is what for Dalton Puffer results in students building their very own, confident use of the language and a clear appropriation of it. This is for Dalton Puffer (2011, p. 196) the "most striking outcome of CLIL programs." CLIL students whose language-learning aptitude and interest are average have been known to benefit from CLIL-like practices as well. This is not commonly found in a traditional language classroom. If we consider the above principles of CLIL, this last concept reinforces the CLIL quality of serving a wider range of types of learners.

Lasagabaster (2008) contends that CLIL programs

> […] help prepare students for internationalization […], boost the affective dimension […], help improve specific language terminology […], enhance students' intercultural communicative competence […], foster implicit and incidental learning […], trigger high levels of communication among teachers and learners [… and] is also believed to improve overall language competence in the target language, in particular oral skills. (Lasagabaster, 2008, p. 31)

He adds that L2 usage boosts motivation in all students while creating a facilitative atmosphere that helps students advance at their own pace. In his research, Lasagabaster, (2008) shows that CLIL groups outscored non-CLIL peers in all areas tested (i.e. grammar, listening, speaking, writing and overall competence). He also found that lower level CLIL-students surpassed non-CLIL students from upper school levels in overall foreign language competence. Finally, when analyzing sociocultural students' status, he found that CLIL exerts the same beneficial effect on students regardless of the parents' sociocultural status. In a later study Heras and Lasagabaster (2015) further investigated motivation between CLIL and non-CLIL students and found that in their study, although CLIL students obtained higher scores in the tests, the differences found turned out not to be significant. They further analyzed the results for gender groups and found that although female students exhibited a higher mean, the differences were non-significant either. Non-significant results were also found for their vocabulary tests and gender groups. Although the results of this study seem to contradict findings of other studies, the authors attribute these differences to sample size and degree of intensity of the CLIL program.

Following the same line of benefits described by other authors above, Lorenzo et al. (2009) also provide evidence for CLIL learners outperforming non-CLIL counterparts and showing greater gains than monolingual students. Lorenzo et al. (2009, p. 427) elaborate on the benefits discussed by Dalton Puffer (2008) analyzed above and extend these benefits to "structural variety and pragmatic efficiency, hence encompassing language growth at lexico-grammatical and discourse levels." These authors describe later start learners' competence as comparable to that of early start students, and they also curiously point to CLIL as a facilitator for cohesion between schools as well as a generator of greater inter-departmental collaboration. Lorenzo et al. (2009) also find great benefits from the fact that each one of the intervening teachers in a given CLIL classroom focuses on a particular area of language specialization, which results in students having access to a wealth of language learning possibilities.

Coyle et al. (2010) insist that CLIL stimulates linguistic competence, cognitive flexibility (see Juan Garau & Salazar Noguera, 2015b), and intercultural awareness while offering learners of various ages an opportunity for language learning which builds on a complementary experience where language learning and subsequent language acquisition develop through content learning. Dalton Puffer et al. (2010a) also refer to high-quality language use taking place in CLIL classrooms. This language allows students to interact more actively through the use of discourse-pragmatic strategies that serve a wide range of functions in more

challenging discussions than the ones present in traditional language classrooms (see Harvey et al., 2013 for a reference to cooperative action and student activation in the classroom). In Dalton Puffer, Nikula, and Smit, (2010b) attention is brought again to the interactional ground facilitated in CLIL, where the communicative intentions that are present clearly distant themselves from those of other learning settings. They argue that the lexical and morpho-syntactical resources that students develop are evident through complex structures in students' written production, which also shows greater pragmatic awareness.

Maillat (2010), who says that CLIL promotes acquisition in an environment that favors a lowered affective filter, also finds evidence for pragmatic strength in CLIL. Maillat (2010) proposes that the facilitative, low affective-filter, low-anxiety (see Harrop, 2012, for the latter), environment that fosters L2 use in CLIL gives rise to a *mask effect* that "liberates" the spoken production of students who feel that the use of the L2 as a medium of instruction and not as a subject itself provides them with a setting that is less constrained by the emphasis on form typical of other learning situations. Overall, for Maillat (2010), CLIL provides the pragmatic strategies that enhance oral production and hence language learning. Furthermore, Lorenzo and Moore (2010) point out that the symbiosis between language use and content goals is evident in students that have the ability to handle academic content even if they possess limited language development. Lorenzo and Moore (2010, p. 30) argue that this is possible through the idea of *notion*, which "can be represented in different language forms (with different degrees of success) at different competence levels." We can link this idea to the scaffolding practices discussed above which play a fundamental role in CLIL. Given the adequate scaffolding language and techniques in the classroom, students can have access to the information they need to obtain the cognitive knowledge that is sought.

Marsh (2012) insists that CLIL offers a series of benefits that range from learning of both language and content to serving political policies and goals moving through social, cultural, professional, economical and psychological benefits carried by bilingualism. He lends special importance to the fact that CLIL caters linguistic opportunities for a much larger number of individuals than the previously existing options due to its presence in mainstream education. Harrop (2012) also refers to this possibility of accessing content and cognitive knowledge and creativity for students that are not necessarily at the top of the class while they also enjoy the benefits of enhanced cultural awareness that allows them to learn more about themselves and others. Nicolay and Poncelet (2013), studying 5-year-old children in a Belgium CLIL context, describe the importance of phonological awareness in the learning and the development of vocabulary. They conclude that

as vocabulary develops, students refine their phonological awareness, which is then used to represent novel, unfamiliar sounds, which, in turn, results in learning new words. Nicolay and Poncelet (2013) also identify phonological awareness in L1 as a possible predictor of vocabulary learning variance in L2, especially in connection with productive vocabulary.

To conclude this brief summary of the numerous benefits of CLIL-like instruction, it is also important to point out that researchers also talk about equal access to internationalization for all students (Pérez-Vidal, 2015), deeper, more intense cognitive processing (Prieto-Arranz, J. I., Rallo-Fabra, L., Calafat-Ripoll, C., & Catrain-González, M., 2015), increase in students' active participation with peers (Nikula, 2010) and clear difference in long-term development of "written complexity, accuracy, and fluency in writing" in relation to non-CLIL learners (Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J., 2015, p. 145). As it is the case in any developing process, limitations and deficiencies have also been identified as part of the CLIL model development. These will be referred to in section 2.3.8 below.

2.3.7. Delineating differences: CLIL is neither Content Based nor Immersion

Some authors have lent importance to establishing a distinction between CLIL and other educational models particularly *content based teaching* and *immersion programs*. Coyle (2006, p. 2) insists that "[w]hilst CLIL shares certain aspects of learning and teaching with these, in essence, it operates along a continuum of the foreign language and the non-language content without specifying the importance of one over another." Lasagabaster and Sierra (2010) discuss the difference between CLIL and immersion programs and identify seven key differences. First, the language of instruction in CLIL does not coincide with the language spoken in the community. Second, CLIL teachers are not usually native speakers of the language used for instruction. Third, the starting age at CLIL is usually around secondary school age. This creates a clear difference in terms of the amount of exposure for students participating in CLIL and Canadian immersion programs, for instance. Fourth, the teaching materials used in CLIL settings are usually adapted for the particular settings as opposed to the ones directed to teach native speakers in immersion settings. Fifth, the language objective of CLIL does not aim for native-like proficiency, as is many times the case of immersion programs. Sixth, immigrant students rarely participate in CLIL programs (particularly in the Spanish context described by these authors), most often they

enroll in immersion programs. Finally, given the relatively new age of CLIL, the amount of research in the field does not yet compare to that existing in connection to immersion settings.

As regards immersion programs such as those of Canada, Dalton Puffer (2008) also argues that whereas CLIL teaches another prestigious language and the initiative is derived partly from parental support housed inside the mainstream education sector, it differs from Canadian education programs in that the language is not official in the countries where CLIL is taught, nor are the teachers native speakers of this language. Dalton Puffer and Smit (2013) agree with some of the elements mentioned by Lasagabaster and Sierra (2010) and offer the following reflection:

> - CLIL is about using a foreign language or a lingua franca, not a second language. Students will encounter the language of instruction mainly in the classroom, since it is not regularly used in the wider society they live in. The dominant CLIL language is English, reflecting the fact that a command of English as an additional language is increasingly regarded as a key literacy feature world-wide.
> - CLIL is usually implemented once learners have already acquired literacy skills in their mother tongue. CLIL teachers are normally non-native speakers of the target language and are typically content rather than foreign-language specialists.
> - CLIL lessons are usually timetabled as content lessons (biology, music, geography, mechanical engineering etc.) while the target language normally continues as a subject in its own right in the shape of foreign language lessons taught by language specialists.
>
> In short, CLIL can be seen as a foreign language enrichment measure packaged into content teaching. (p. 546)

Regarding *Content Based Instruction (CBI)*, Dalton Puffer (2008) insists that CBI programs, such as the ones that take place in the United States, develop second language skills

> through the teaching of curricular content that is not typical of language classes per se […] such programs have commonly been developed in situations where education systems have to cope with large numbers of immigrant speakers who have yet to learn the official language of instruction in the system. (p. 2)

Along these same lines, Gabillon and Rodica (2015) state that CBI original purpose was two-fold: it provided extra support to tertiary students in the language used as the medium of instruction and it prepared immigrant pupils to move on to mainstream education. The ideas that originated CBI when contrasted with the ones behind CLIL origins make it abundantly clear that a sharp differentiation between the two is necessary.

2.3.8. CLIL's future challenges

To conclude this brief reference to CLIL, it is important to mention a number of concerns raised by researchers in connection with this model. Although some shortcomings have been mentioned and references to limitations have arisen since CLIL first started developing, criticism toward CLIL practices has become a lot more evident in recent years. For instance, Bruton (2011) raises very serious concerns in relation to CLIL results, its implementation as well as the research that has derived from it. He argues "there is every reason to believe some students may be prejudiced by CLIL, and that not only academic, but also institutional, interests may be taking precedence over some students' interests in the state educational sector" (523). In his discussion, he questions research studies reporting on CLIL and he argues that these present biased results that respond to researchers' interest; that these studies are limited and questionable; that CLIL groups are already more proficient or motivated before they are part of the CLIL process and that this fact alone yields quantitative and qualitative results that do not necessarily favor CLIL approaches. All along, he insists that there is not enough control of intervening variables (i.e. number of hours, additional English classes, existing initial differences, lack of pretests, socio-economic differences, methodology and type of instruction, among others), and that these variables may be influencing results that are commonly associated with beneficial CLIL effects in the literature reporting on CLIL results.

Furthermore, Bruton (2013, 2015) insists on the idea that there is a lack of evidence of CLIL real benefits and that "if there are any possible CLIL successes they are probably attributable to selective measures and contrived supportive conditions" (Bruton, 2015: 119). Bruton (2013) questions CLIL from different perspectives, from its novelty to its effectiveness in language development, to its teachers' motivation. Bruton (2013), furthermore, insists that the current, constant concentration on CLIL takes attention away from the many non-CLIL students and/or the problematic situation of FL teaching in mainstream education in the

Spanish context, the one that does not involve CLIL practices. He concludes that there are political, educational and parental factors that come together to facilitate the adoption of CLIL practices and he calls for redirecting attention toward the less favored groups in the educational stream.

More recently, Fernández-Sanjurjo, Fernández-Costales, and Arias Blanco, (2017) have reported on a study that evaluates competence in Science in students enrolled in CLIL and non-CLIL programs. In their report, involving primary students, they discuss two important findings. First, non-CLIL students statistically outperform their CLIL counterparts on competence in the subject of Science in their L1. Second, they claim that, when considering the socio-economic status of students, students belonging to high and medium status obtain similar scores and show no statistically significant differences among these groups. However, students with lower socio-economic status obtain lower scores and those are, in turn, statistically significantly lower when compared to the group formed by the high and middle status students. These results, of course, deserve a lot of attention given that they contradict one of the main tenets of CLIL. They seem to suggest a disadvantageous position for some CLIL students regarding content learning and they also hint at the effect of the socio-economic status of students in the program. Further studies in these areas are deemed extremely important.

Although several concerns have surfaced since the beginning implementation of CLIL, there is no doubt that the issues raised by Bruton (2011, 2013, 2015) and Fernández-Sanjurjo et al. (2017) are worthy of delicate attention. Hütner and Smit (2014) have challenged and accurately provided clarity to some of the assumptions presented by Bruton (2011, 2013, 2015). They concentrated particularly on the issues of CLIL complementing, not replacing, FL teaching and in explaining the known variability that conforms CLIL. However, not all Bruton's ideas have received detailed attention. Important consideration should be given to Hüttner and Smit's (2014) call for contemplation and analysis of national, regional or institutional language policies. It is these language policies that can ascertain appropriate CLIL implementation and its subsequent beneficial results.

Pérez-Cañado (2016) also makes reference to this upsurge in criticism toward CLIL. She acknowledges that after a period of cheerful enthusiasm toward CLIL and its results, a more conscious scrutiny has settled, and more critical analysis is surfacing now. Pérez-Cañado (2016) identifies three main challenges surrounding CLIL, namely its characterization, its implementation and its research. Regarding the difficulty of pinning down CLIL limits, Pérez-Cañado (2016) suggests that the pedagogical and research community can find a common ground that recognizes the variety of programs encompassed within CLIL as well as the

multiple results and effects that these multilingual programs yield in order for this community to benefit from them. As for the implementation of CLIL, Pérez-Cañado (2016) argues that this could be achieved through targeting program diversity and ensuring equal access to content and languages to all students, regardless of their personal differences. Finally, in terms of research, Pérez-Cañado (2016, p. 18) calls for basic tenets of "unbiased, balanced and methodologically sound research" to provide us with accurate information regarding CLIL. According to Pérez-Cañado, these ideas may serve to pave the road that may aid CLIL in producing the results that the present society demands.

Along with the crucial issues mentioned before, the following concerns have appeared in the literature throughout different studies during the last almost ten years. Lorenzo et al. (2009) find the integration of content and language a current challenge for CLIL practitioners. These authors also speak of the need for more research to investigate the status of technical, content related, productive and receptive vocabulary; as well as more longitudinal studies that also include gender variables. Lorenzo and Moore (2010) point out that while the teachers' variation in how they approach the use of language in the classroom can be beneficial for students, these teachers' views can be widened to offer even more benefits to learners. Dalton Puffer (2011) calls for more clear-cut goals and objectives for the model (see Gabillon & Rodica, (2015) concerning insufficient professional knowledge). She argues that although more conceptualization, guiding frameworks and material have been developed, there is still need for more commitment from national education systems in Europe. Meyer (2010) mentions the need for guidance on how to develop quality materials that implement the 4Cs framework. Dalton Puffer (2011) asks to direct attention to language forms and usages requiring improvement, such as use of null subject, negation, suppletives, more relevant subject-based language, and better ability in describing subject specific concepts.

Marsh (2012) mentions the challenge faced by teachers in matching the age of the students (as well as their language competence) with the cognitive demands of the model, a fact made evident by the knowledge and skill heterogeneity of the CLIL classrooms. He also points to the need to increase higher order thinking skills in the classrooms. Harrop's (2012) concerns initially have to do with the possible disadvantageous position of weaker language students in the classroom, which may be worsened if the appropriate treatments are absent. Her biggest concern is much stronger; she asks to accept the fact that given the additional difficulties present in a CLIL type of setting, overachievement will inevitably be limited. This will bring with it the consequence of major skepticism and limited interest in CLIL programs. Harrop (2012) also refers to theoretical and methodological deficiencies of CLIL

that are observed in the model favoring some skills over others. She concludes by analyzing the costly demands of the model in relation to both financial and human resources as well as for its ultimate execution.

Dalton Puffer and Smit (2013) also point to the need of attention in managerial issues in CLIL contexts. They make a compelling case in favor of more research in CLIL and more importantly they provide specific lines of research to do just that. They propose this be done by giving continuity to earlier research and by making action-research an active component of CLIL programs. They mark seven lines of research in relation to goals examination, investigation on perceived success, examination of stakeholders' views of CLIL in connection to various languages, investigation of actual presence and use of the five language skills in the classroom, investigation of academic language function of CLIL in the classroom, investigation of CLIL and non-CLIL pedagogies' similarities and differences, and finally investigation of explicit language teaching in CLIL settings.

2.3.9. Concluding remarks

There is scientific evidence of the multiple benefits that can result from the correct implementation of CLIL practices. As can also be concluded from the last set of paragraphs and as is certainly the case in any movement that is alive and evolving, there is evident room for improvement in CLIL practices; more importantly, many stakeholders are already working on finding solutions to some of the problems that have been identified. Stakeholders are certainly interested in continuing to reap the benefits from CLIL while researchers are actively seeking for results and possible solutions to the problems they identify. There is significant evidence of the benefits derived from CLIL when the model is correctly implemented, but only time will tell us the ways CLIL finds to continue its evolution and improvement. This process will secure the position that the model already enjoys as a referent in the language education arena.

CHAPTER 3

BILINGUALISM AND BILINGUAL EDUCATION IN COSTA RICA

This chapter begins by briefly discussing the variety of bilingualism that is currently present in Costa Rica. The first part of the chapter presents a discussion of the key theoretical aspects that were presented in Chapter 1. The Costa Rican context is, thus, analyzed in light of those theoretical concepts. Moreover, the chapter presents the working definition of the term *bilingualism* that will be used as a basis for the analysis of the Costa Rican context. It concludes by discussing the type of foreign language instruction that is common practice in this country and contrasts it with what specialists define as true bilingual education. It outlines the efforts that are being made by the National Board of Education with the intention of providing students in the public sector with high quality bilingual education.

3.1. BILINGUALISM IN COSTA RICA[4]

Nowadays, when people in Costa Rica hear the word *bilingual*, they are bound to associate the idea with a native speaker of Spanish who also speaks English (Spanish being the official language of the country). Although other forms of bilingualism exist, they are usually not considered part of the equation. For instance, indigenous languages, derived mostly from the Chibchan family, which prevailed long ago, do not typically come to mind when people think of bilingualism and bilingual education in this country.

However, the last census attests to the presence of at least eight different indigenous groups (Costa Rica-INEC, 2017). These peoples amount to 104,143 individuals and represent less than 2.5% of the total population. We should also bear in mind that many of the participants in the census who identify themselves with an ethnic group do not necessarily speak the language of their ancestors or live in the indigenous area; the real number is thus much lower than the official one given above.

According to Costa Rica-INEC, from the indigenous groups in the country, the Bribrí people (18,198), Cabécar people (16,985) and Guaymí people (9,543) are the largest groups whose languages still enjoy some vitality, for the rest of them are already extinct or on the way to becoming so. In these groups, we can still find

[4] As part of the research carried out for this doctoral thesis, earlier versions of several sections of this chapter have been published in Castro-García, D. (2017). Bilingualism, CLIL, and immersion programs: The case of Costa Rica. In G. Nieto Caballero (Ed.), *Nuevas aportaciones al estudio de la enseñanza y aprendizaje de lenguas* (pp. 123–136). Cáceres: Universidad de Extremadura.

native speakers of each indigenous language. Some of them continue to wear their traditional clothing, and many of their children learn the indigenous language as their first language, most often along with Spanish, which is also the official language at school.

Yet another instance of bilingualism in Costa Rica is the language spoken by the Afro-American population in the Caribbean region, a mix of English and Spanish that has resulted in a different dialect spoken by many people in the region. Although this one is not discussed here in depth, it should be recognized as well. In sum, other types of bilingualism are present in Costa Rica, aside from Spanish/English bilingualism. Nonetheless, English-related bilingualism and in particular, English bilingual education represents a strong, moving force in this country, one in which the government is investing most of its time and resources. Therefore, English bilingual education will be the focal point in the remainder of this section when discussing bilingualism and bilingual education in Costa Rica.

3.2. BILINGUALISM: THEORY INTO PRACTICE

Among the concepts and hypotheses to be discussed below, to provide a more precise view of the current situation in Costa Rica in light of the theory of bilingualism discussed in Chapter 1, attention will be given to the following:

- Concepts of additive and subtractive bilingualism (Lambert, 1977, 1981)
- CALP skills (Cummins, 1979, 1980, 2008, 2014); Interdependence Hypothesis (Cummins, 1979, 1984, 1986, 2000, 2005, 2009); Common Underlying Proficiency (CUP) (Cummins, 1980, 1986, 2005, 2007); Threshold Hypothesis (Cummins and Swain, 1986)
- Input Hypothesis (Krashen 1982, 1991, 2003, 2013)
- Output hypothesis (Swain, 1995, 2000a), Languaging (Swain, 2006)
- Translanguaging (García, 2009; García and Wei, 2014)
- Bilingualism, including Baker (2011), Grosjean (2013), Mackey (2005), and Edwards (2006)
- Working definition of bilingualism (Grosjean 2010, 2013)
- Current definitions of bilingual education: García (2009), Cummins (2014a), Baker (2011), Bialystok (2016).

Regarding the theoretical aspects analyzed in Chapter 1 on bilingualism, relevant concepts will be applied to the Costa Rican context. From Lambert's (1977, 1981) distinctions between additive and subtractive bilingualism, the Costa Rican context portrays characteristics of additive bilingualism as both Spanish and English are highly valued and there is no intention of replacing Spanish with English. On the contrary, the National Board of Education supports Spanish language proficiency and its use in the classroom, and at the same time seeks to provide quality English classes to all citizens through the public educational system. This system, as we will see, is clearly in need of reform and improvement; this is a goal, the government insists, that it is working on. The additive characteristic of bilingualism is seen here as covering a wider scope. The positive attitude toward the benefits brought around by knowing English along with Spanish is recognized at the national level. Given the negative features associated with subtractive bilingualism, having an additive orientation is a good starting point for the Costa Rican context. The negative experience associated with subtractive bilingualism toward most of the aboriginal languages that once existed here is still present in the cultural memory of the country.

3.2.1. Application of Cummins' hypothesis

We will begin by analyzing how Cummins' (2008) fundamental theories apply in this context with relation to bilingualism. We cannot help but admit that these are absent, to a fairly large extent, from the Costa Rican educational setting. For the most part, practitioners of bilingual education in Costa Rica still see the first and second languages as two separate entities. This fact alone affects the implementation of Cummins' theories. Bilingual education is usually limited to students' use of BICS in simple tasks that, to a great extent, are not cognitively demanding and do not require real communication among the students. CALP skills, which are necessary to transcend the immediate context, are not usually required. Given their importance, these skills should be more directly present in the curricula to aid students develop the literacy knowledge that they would need outside of the classroom setting. These ideas, also present in the categories of objectives in Bloom's taxonomy, may very well be discussed in teacher training sessions, both at pre-service and in-service sessions, but they do not necessarily make their way into primary and secondary classroom settings.

In relation to another of Cummins' (2009) key concepts, the *Interdependence Hypothesis* should also be considered as central in every educational context. Given the fundamental role that the L1 has in the development of the L2 in this hypothesis, the development of L1 in Costa Rica should be analyzed. Whereas Spanish is one of the core subjects in the curricula across primary and secondary education, there seems to be indication of serious limitations in the development of Spanish at certain points of the students' language development. The following studies show how the panorama changes as students get older. According to these studies, the vocabulary of students evolves from promising results during preschool years to a critical situation at the university level.

Murillo Rojas (2009) reported on the lexical richness of 101 Costa Rican, preschool children from 4 to 6 years old. She evaluated the diversity of words through the analysis of lexical content words (adjectives, nouns and verbs) in oral narratives with the intention of determining whether these children displayed normal language development. In her study, Murillo Rojas used the formula Pd50, which entails the analysis of the variety of words in the first 50 statements produced by the children in an oral narrative. This study concludes that the lexical variety of this group of children is above average. That is, children in this age group have a higher than normal language development. The mean for the lexical diversity index in this group is 0.65 (1 is the highest index) while the average that is considered normal in the literature is 0.50, claims Murillo Rojas. The author also concludes that gender does not exert influence in the number of words that children know, as both genders produce similar word counts.

In a similar study, using the same methodology and procedures, Araya Ramírez (2012) measured vocabulary in primary school. The vocabulary of 1st–3rd grade students and of 4th–6th grade students was evaluated in the production of narrative and explicative oral texts. The index of lexical diversity was established at 0.57 for the former and at 0.56 for the latter in narrative texts. That means that from the perspective of narrative texts students show a normal development (this ranges from 0.51 to 0.61). In the case of explicative texts, the former obtained a mean of 0.61 and the latter of 0.62. The first three grades have a normal index of lexical diversity and the last three have a superior index of lexical diversity. Araya Ramírez, however, argues that it is not possible to identify a logical, sequential growth of vocabulary during primary school that would reflect the positive influence of the educational process. She concludes that the school appeared to have little influence on vocabulary development as she found no signs of clear, constant

vocabulary growth as students advance in school. She argues that the development of oral skills seems to be assumed to be a responsibility of family and society and is missing from the school curricula. This study, however, still supports the idea of normal-to-superior development of vocabulary during childhood. Whereas that is the case for preschool age children, the situation is different for teenagers.

In a recent article in *La Nación*, the country's most important newspaper, Cerdas E. (2017) reports on a study carried out by researchers from Universidad de Costa Rica, which determined that the Spanish vocabulary of secondary students is *basic*, contains only *common* words, and is set at 2200 words. Five thousand words would indicate a more refined and appropriate vocabulary for this age group according to this article (Cerdas E., 2017). That is, whereas at the preschool level vocabulary knowledge seems to exhibit a normal development, during high school years the normal growth in vocabulary size seems to weaken.

In another article, Madrigal and Vargas (2016) analyzed written compositions of Costa Rican university students to determine their vocabulary in Spanish. The count of lemmas in these compositions showed that a mean of 56.97 lemmas in the first 100 words of each composition places these students at an inferior level in terms of lexical diversity. For these authors, a mean of 66.5 to 70.5 would mean lexical quality of a superior level, 60.5 to 64.5 represents a central level and 54.5 (or lower) to 58.5 represents an inferior level. The authors also claim that there is no lexical variety in the texts that were analyzed and that male students tend to produce texts with a more sophisticated variety of words than female students. Madrigal and Vargas (2016) find that the students use repetitive vocabulary in their composition and that the ideas are segmented and lacking cohesion. These authors conclude that this problem with lack of vocabulary derives from teaching practices that continue to analyze Spanish language through the study of isolated segments of the language without allowing students to create links between contents that would allow them to use their knowledge appropriately at the discourse level. Sánchez Avendaño (2005) offers similar results and argues that university student compositions lack written discourse strategies. He concludes that the compositions in his analysis are evidence of problems in morphosyntax, lexicon, punctuation and spelling, and to a greater extent in cohesion, coherence and presentation of information. The problem has then been documented at the university level as well.

As a professor at Universidad Nacional (Costa Rica), I have experienced these problems myself while working in the classroom with first year university students who display a very limited control of vocabulary in Spanish, and often

appear not to be familiar with the concepts represented. It is common to hear students "explain" a concept through its features, uses, and functions rather than using the specific label that delimits the concept: the vocabulary word. They also show limitations with basic concepts such as word, sentence and paragraph structure. As expected, these problems are carried over when students attempt to acquire skills in English as an L2. As Madrigal and Vargas (2016) argue, many of these problems are associated with the limited reading ability that students possess. Reading has not traditionally been a common leisure activity in Costa Rica, and the presence of technological devices has only worsened a situation that was already critical. Whereas the problem does not appear to be serious during childhood, the situation becomes increasingly complex as students get older. Although the Board of Education claims that they have begun to implement changes in the Spanish curricula, the problems are still latent. A clear objective should be set to strengthen Spanish skills and abilities, which in turn would increase the possibilities of transferring this knowledge over to English in the second language classrooms.

Considering the previous information as a point of departure for students' SLA as well as my personal experience in the educational system, I believe that the possibility of application of the Interdependence Hypothesis is rather limited. The weak foundation of the students' L1 depicted in the studies mentioned above also affects the optimal performance of the Common Underlying Proficiency (CUP) (Cummins, 1980, 1986, 2005, 2007). If the cognitive and academic proficiency for each of the languages in the CUP benefits from the experiences in each of these languages, then a weak performance in Spanish as the L1 will provide little benefit for the development of English, the L2. Although we know that development of an L2 strengthens the students L1, a weak L1 is not ideal for a good development of the L2. Cummins (2007) assumes that the CUP enables students to transfer academic, cognitive and literacy knowledge from one language to another, and that is how the students' real language proficiency surfaces. This takes us to the language level attained by the students, and that would account for what is stated in the Threshold Hypothesis (Cummins and Swain, 1986). According to this theory, students require a certain level of L2 language development to enjoy the benefits of cognitive growth and linguistic competence brought about by bilingualism. For the most part, given the characteristics of language attainment of the Costa Rican students, they seem to be placed at the second level (if not at the bottom level, as limited bilinguals) of this hypothesis where students are not balanced in terms of their knowledge of both languages.

As unbalanced bilinguals, there is no evidence that students would experience negative or positive consequences in connection to bilingualism (Baker, 2014). At this second level, students' thought patterns are similar to those of monolinguals (Baker, 2014); needless to say, this is not the level of bilingualism we would want our students to have. If Costa Rican students are currently placed at the lower or middle level on the Threshold scale, we would want students to advance to the top level so that according to the Threshold Hypothesis, they would enjoy all the benefits that bilingualism provides.

3.2.2. Application of the input hypothesis

With respect to the input hypothesis (Krashen, 1982), it would be safe to say it has received attention in the country but, similar to the situation in other countries, its interpretation might not have been the most appropriate. During the 1980s and 1990s, the input hypothesis received greater attention at the university level in teacher training programs. This focus on input was transferred to the primary and secondary classrooms in Costa Rica although mostly in reading; listening skills developed through a strong input program are not common. The fact that students finishing secondary education must take an English reading test at the end of their schooling also supports the idea of input, mainly through reading, to be the strongest practice in SL education contexts.

As discussed in section 1.1.1.3, Krashen (1982) lists characteristics that input should have for it to be optimal. Namely, it should be comprehensible, relevant, not grammatically sequenced, and sufficient in quantity. While I find the presence of all of these characteristics unlikely, the last feature is the one whose presence seems to be more affected in the Costa Rican setting, as input in this context is mostly limited to short reading passages that are read merely with the intention of later completing exercises that test reading comprehension. This input, then, does not appear to be ideal to promote a more active use of the language for communication and interaction, for example. There is a clear need to follow the idea of Krashen (2003, 2013) and Krashen and Bland (2014) regarding the importance of recreational reading. These authors insist that reading can bring benefits associated with vocabulary building, spelling, and literacy. If input through reading is only limited to the brief encounters with the language that take place during an English class just to carry out completion tasks, then that input does not suffice to provide its potential benefits, but this appears to be the case nationally

across schools. However, the basis is there, and students are already used to working with written material. Parents and teachers and any other member of the educational process have the mission of motivating students to expand their concept of leisure activities so that these include reading.

3.2.3. Application of the output hypothesis

As for the output hypothesis, as described by Swain (1995, 2000a), much attention was directed to students' output when the communicative approach began to be widely implemented at all educational levels in Costa Rica. The idea of students communicating in *real* scenarios received enthusiastic support in this educational setting. This *real communication* practice, however, was poorly monitored and resulted in students conveying poorly constructed messages containing structures and patterns of Spanish that were disguised by English-like features. While students and instructors had the idea of communication taking place, this communication was and is still today in need of serious monitoring and correction for this practice to have a positive effect in the SLA process. Part of the foundation of the Output Hypothesis is that output allows students to process language knowledge at a deeper level. I believe, nonetheless, that students must be guided in this process, at least initially, to find a way to analyze and monitor their own speech. This guidance, however, seems to be absent from regular teaching practices in Costa Rica.

Swain (2000a) also argues that students need to be able to process and analyze grammar in order to produce language accurately. In Costa Rica, many instructors have sacrificed grammar teaching and a clear focus on form under the impression that these foci limit communication. They do not realize that grammar and focus on form are necessary as they serve as groundwork for production in the L2. It is only with the right kind of knowledge that students can apply the *noticing/triggering function* (the first function of the hypothesis) that alerts students of the gaps they have in their knowledge. Production, in the form of output language, lead to hypothesis testing, giving students the chance to develop and test their own grammar rules. For Swain (1995), hypothesis testing takes place while students are facing communication challenges. If the learning environment does not offer students the chance to communicate and discover what they need to learn to succeed in a communication task, then language development will not take place. The third function of the hypothesis, the metalinguistic function, the use of

language to reflect on language is lacking from regular, mainstream teaching settings in Costa Rica. It may be present at the university level in some courses that require students to analyze their production, but at the primary and secondary level, this practice is not common.

The last feature leads to *languaging* (Swain, 2006), the idea of producing language to understand language. Again, this metalinguistic form of language can only take place if students have the base knowledge to analyze and discuss their output and that of their classmates, thus advancing their SL knowledge. Whereas students can hypothesize while trying to solve linguistic problems, they need either access to material or professional support, or enough opportunities to continue testing their own hypotheses in order to confirm or reject them. As mentioned above for the metalinguistic function of the Output Hypothesis, this type of practice is not common in the Costa Rican context; the output that students produce often goes unmonitored and hence may not provide all the benefits that it could bring under the right circumstances. The *collaborative dialogue* that Swain (2000a) calls for, is often absent from these settings or may even take place in Spanish. I believe that the fact that it takes place in Spanish would not be detrimental as long as students do it with a clear intention of advancing in second language acquisition. The problem is that students often switch to Spanish to avoid using English, thus hindering the SLA process. Swain (1995) admittedly recognizes the importance of input along with output. The right balance between input and output would be ideal to provide students with a well-rounded experience and allow them to fully embrace the learning process and use diverse opportunities to advance in the development of English as a SL.

3.2.4. Application of translanguaging

Translanguaging, the use of varied language practices to deal with the bilinguals' world is not common either in the Costa Rican context. Given that this practice requires a more developed type of bilingualism, and that students for the most part do not have this characteristic, the Costa Rican context does not serve as an ideal setting for this practice. I completely agree with Baker (2011) that this practice should be limited to bilinguals whose languages are well established or are developing strongly. For García and Wei (2014), translanguaging is also present in the educational domain. If put into practice in the Costa Rican context, it would require instructors to ensure that it really serves the functions of developing oral

communication, enhancing understanding, promoting bilingualism and biliteracy and facilitating home-school cooperation, as Baker (2011) proposes. It should not be confused with a type of code-switching used by the students simply to avoid using the L2.

3.3. UNDERPINNINGS OF BILINGUALISM IN THE COSTA RICAN CONTEXT

We will begin by saying that, using Baker's (2011) differentiation, the Costa Rican context portrays *individual* rather than *societal* bilingualism, as it is the case that most Costa Rican individuals attain bilingualism through formal instruction. Bilingualism in Costa Rica may also be supported by regular contact with English-speaking tourists and by individuals traveling to English-speaking countries, but it is not a feature of the society *per se*. In addition, following Baker's (2011) characterization of what elements are necessary to describe bilingualism, we will analyze how these concepts apply to the Costa Rican reality. If we recall our discussion in section 1.1.2, there is a lot of agreement as to what features are required to describe the term *bilingualism*. Readers may want to refer to that section to review Baker's (2011) concepts, which were analyzed along with notions of authors such as Grosjean (2013), Mackey's (2005) and Edwards (2006). First, regarding *ability*, the Costa Rican bilingual education has traditionally favored a more *receptive* type of bilingual ability in the sense of what Baker (2011: 3) calls *passive bilinguals*. It is easier for most Costa Rican students in the public sector to understand what they hear or read in English, than it is for them to speak or write in the language. The more recent, increased contact with native speakers of the target language has led to an improvement in productive skills favoring speaking abilities over written ones. The presence of many transnational companies has also contributed to the linguistic development of individuals in post-educational environments.

In terms of *use,* the use of the language is typically limited not only to the school setting but to the classroom setting. Despite the strong presence of English in the media and on the street (thanks to tourism), most people limit their use of English to the classroom setting, until they are practically "forced" to use language in the workplace. In this dimension Baker (2011) argues that it is important to consider *where* individuals learn the language because this evidently affects how they use it. As discussed for the previous feature, most individuals learn the L2 language in formal settings, although several individuals may support formal

instruction with frequent practice with native speakers in the surrounding environment. As for *how* people use it, for the most part individuals limit the use of English to the school or the workplace. Family and society settings favor Spanish, with the common exceptions of intercultural marriages where, if one family member speaks English, the family may alternate the use of languages at home.

When it comes to the *balance of the two languages*, Baker's (2011) third feature, there is clear dominance of Spanish over English in general. There is even a clear influence of Spanish, both in the accent and pronunciation of many speakers as well as in various other language skills. Equally competent bilinguals are scarce and as speakers reduce the use of English either at school or at work, Spanish prevalence becomes even more apparent. As for the fourth feature (*age*), *consecutive or sequential bilingualism* is the norm given that the contact with the English language takes place well after the age of 3; in most cases after the age of 13 for students limited to the public arena, and after the age of 5 even if students attend private education. Simultaneous bilingualism is rare, and it is limited to families with at least one native speaker of English or to those who have access to certain private schooling where English is a priority.

As far as *development* is concerned, it is *incipient* for most of the secondary school population; consequently, students graduate from secondary school when their English is still in the very early stages. Baker (2011) describes incipient bilingualism as that form of bilingualism where one language is poorly developed while the other is well developed. Keeping in mind what we discussed in section 3.2.1 about L1 development, we must be aware that Spanish development also faces certain challenges in this context. When students leave secondary school, they are still in need of further support to increase L1 skills and vocabulary; this, as we have discussed, also represents a challenge for L2 growth. Moving to the fourth feature, when dealing with *culture*, the Costa Rican society has traditionally been very receptive to the English-language culture, particularly in relation to the United States. Due to geography and tourism, contact with this foreign culture has resulted in appreciation and many times adoption of customs, practices, and even traditions from the North American countries. As a result, it can be said that there are many instances of biculturalism amongst the Costa Rican society, or at least that there is very good reception and appreciation of the foreign language culture represented by the U.S.

Regarding *context*, *additive bilingualism* is clearly present in Costa Rica. The government and different institutions strive for the population to learn English

while retaining their Spanish skills. Whereas there is no immersion in the target language culture, the interest in learning English and the possibility of contact with the language through media and in person provides a facilitative context for second language learning. Finally, in terms of *elective* or *circumstantial* bilingualism, Costa Ricans could be described as experiencing both, as many Costa Ricans are clearly choosing to learn English while others have been faced with this learning as a necessity to fulfill the demands imposed by their careers. The factor of *electivity* is evident as people often select a school because of the language offered and the level of foreign language achievement associated with the school. As for circumstantial bilingualism, the growth in the presence of now over 400 transnational companies attests to how important bilingualism has become nationwide. Firms such as Intel, HP, Hilos A&E, Abbot, Amazon, Kimberly-Clark, and Procter&Gamble require the expertise of functional bilinguals to service consumers in many different international markets. While bilingualism represents a choice for many, it also has become indispensable if one desires to compete successfully in the job market.

3.4. WORKING DEFINITION OF BILINGUALISM

As discussed in section 1.1.3, a clear-cut definition of the concept bilingualism and of the individual bilingual represents a great challenge as many factors could be considered for either one of them. Authors seem to agree on the idea that there are several possible levels of proficiency and that speakers are fairly proficient in different language skills (see Edwards, 2006; Skutnabb-Kangas & McCarty, 2008). Authors also agree that full mastery of both languages is hardly ever attained, but this does not stop people from communicating and being completely functional in both languages (Wei, 2005; Myers-Scotton, 2006). For the purposes of this study, we could recall Grosjean's (2013, p. 5) definition of bilingualism as "the use of two or more languages (or dialects) in everyday life," and that of bilinguals as "*those who use two or more languages (or dialects) in their everyday lives*" (Grosjean, 2010, p. 4, emphasis in original)."

These definitions allow for flexibility of use and level of proficiency in individuals, while at the same time aiming at bilinguals who could attain language faculties enabling them to function in a wide array of situations and domains of daily life. Although Costa Rica is a Spanish-speaking country and Spanish is, therefore, the official and most frequently used language in society in general, there

is a clear interest in learning English, supported at the parental, school and governmental levels. This support, in connection with the imminent presence of English speaking tourists, transnational companies, and a permanent influx of mass media information and entertainment from the U.S., makes English contact readily available to those interested in becoming bilingual. English is not an official language, but many people strive to reach proficiency levels that would allow them to use English in everyday life. It is in this context that this study takes place.

3.5. The Concept of Bilingual Education

The main characteristics of bilingual education and how it is implemented in the country will be described in the following section. The greatest challenge for bilingual education in Costa Rica is the fact that it is seen, still today, through the lens of foreign language teaching where language is taught primarily as a subject, rather than used as a medium of instruction. García (2009) insists that if languages are taught following traditional teaching in the form of foreign language instruction, then, that type of education should not be labeled bilingual education. For García, a language other than the students' native language should be used to teach content for bilingual education to take place. Cummins (2014a), along the lines of García, argues that the target language should be used as medium of instruction for bilingual education to actually take place. In addition, even though Bialystok (2016) offers a definition that embraces more possibilities for bilingual education programs and calls it an umbrella term, she still notes that for bilingual education to be present languages should be used to teach non-language academic subjects.

Baker (2011) defines the type of education program present in Costa Rica as a weak form of bilingual education. He labels it "Mainstream Education with Foreign Language Teaching." As will be discussed later on in this chapter, the teaching of non-language subjects in English represents the challenge that policy makers, teachers, curriculum designers and students must deal with. Inability to move toward more modern approaches to teaching and learning second languages would continue to restrict students' access to more beneficial forms of education.

3.6. COSTA RICAN CONTEXT

Traditionally, there has been a very active interest in English learning in Costa Rica. Through anecdotal articles, Reindorp (1943, 1944) offers a vivid description of the origins of English teaching in the country. Reindorp (1943) meticulously explains the procedure and methodology followed in an on-air radio program that broadcasted English lessons. He contends that this radio program tried to "satisfy in part the tremendous popular demand for English lessons. The demand for a knowledge of English was so widespread that it was decided to broadcast lessons over one or more stations" (p. 425). The interest in learning English and the response to the program were such that the 15-minute lessons became 30-minute lessons within a month of broadcasting. Pamphlets and complementary materials were also created to supplement this on-air instruction. Reindorp (1944) continues to describe his personal experience as a representative, among the many that were sent to Costa Rica to offer support in the teaching of the language and training of personnel. Reindorp depicts his work in assisting the Minister of Education, at that time, in the establishment of the Institute of Modern Languages in the National University. This institute intended to increase the availability of English instruction to satisfy the growing demand on the subject, which was severely limited by the lack of trained personnel. The institute also sought the creation of an international summer school of languages. In his depiction, Reindorp pinpoints the evident need for more and better-trained teachers as well as the conspicuous need for teaching materials.

Historically, Costa Rica has observed a number of educational reforms promoted by the state to respond to economic challenges surfacing at specific moments. Marín Arroyo (2012) speaks of a parallelism between English teaching practices and economic development in this country. In his article, Marín Arroyo (2012) depicts the steps through which English language teaching evolved throughout the 19th century. He describes each decade and pinpoints the major developments and changes in the field of EFL teaching in Costa Rica as well as the political and economic changes that gave birth to these reforms. The following paragraphs describe some of the main events described by Marín Arroyo (2012).

The connection between education and economy was first observed toward the end of the 18th century and beginning of the 19th century. At that time, the United States and England were buying more than 50% of the local coffee production. Coffee trading gave way to the settlement of companies from these English-

speaking countries in Costa Rica. Marín Arroyo (2012) asserts that as a consequence of trading, English language instruction becomes official in 1901. As a result, the Costa Rican government hires professors, mainly from the U.S., to teach the language. However, the following decades did not see a big impact in language teaching/learning in the country. In the 1930s, the government drives initiatives to establish cultural exchange programs among Costa Rica professors and professors from the United States through the *International Friendship League*. This initiative, along with the signing of a commercial agreement with the U.S, resulted in a significant increase of English instruction and a spring of private institutions offering English lessons. The cultural exchanges continued. In 1941 the Inter-American Summer University was created, and a wealth of books, pamphlets and movies were donated to the Costa Rican institutions. By 1944, English teaching at the elementary-school level was declared mandatory. Lack of trained personnel limited the application of this decree, however. Around this time, the Costa Rican-North American Cultural Center was created in connection to the U.S Embassy. The creation of this institution and the establishment of the English teaching major at the University of Costa Rica in 1958 provided a strong basis for English instruction in the country.

Several changes took place in the following decades. During the 1960s, the number of hours of English lessons increased from 15 to 18 hours a week. During the 1970s, a new language curriculum was developed. It sought to provide students with more meaningful learning and more communicative practices. The 1980s witnessed the advent of globalization and several changes took place then. Starting in 1987 a new series of books were used, more updated curricular and methodological practices were implemented, teacher training seminars were carried out and a new English Resource Center was created in the Board of Education.

The 1990s represented a leap in terms of access to language courses due to emphasis on English instruction in different majors at the four state universities. These institutions also offered free language teaching programs to the community at large, and the establishment of "Scientific Schools" and "Experimental Bilingual schools" was observed nationwide. In 1995 the Program of Foreign Languages for Development (*Programa de Estudios de Lenguas Extranjeras para el Desarrollo)* was implemented in the public sector with the intention of reducing the existing gap between the language outcomes at the public and private levels. Marín Arroyo (2012) discusses several short-lived initiatives that had been implemented at elementary school level without much success. At the present moment, however, an ongoing curricular program is being actively carried out nationwide in public

schools, and has been for over 20 years now. While this is mostly limited to vocabulary recognition and subsequent limited production in higher elementary school levels, it has been maintained through the years and it provides a good basis for improvement in near future.

As the previous paragraphs can attest, the history of English language learning in the country has undergone many changes. It seems, however, that since the beginning it has been widely accepted and it has traditionally been linked to the economical requirements of the country. This is as true today as it was during the last century. In the words of Marín Arroyo (2012):

> El desarrollo económico nacional ha estado siempre ligado a la educación y cada vez que el Estado pretende realizar un cambio en su estrategia económica el primer paso que da es promover la educación. Eso nos hace concluir que ella es, de manera irrefutable, el motor que anima todas las demás estructuras de la sociedad […] siempre que se ha generado un cambio en el estado, la enseñanza de la lengua inglesa juega un papel determinante y se convierte en un instrumento impulsor del desarrollo económico. (p. 10)

More recently, various Costa Rican institutions have been making strong efforts to provide a high-quality, public, bilingual education. In 2008, the Costa Rican government created the National English Plan (*Plan Nacional de Inglés*), ascribed to the Presidential Office. This plan is made up of several institutions whose purposes are to diagnose the needs that Costa Ricans have in terms of professional bilingual education; to provide higher education institutions with guidelines to promote English learning in children, adolescents and (young) adults; and to promote the involvement of public and private institutions in the development of language skills in the workforce, among several others (Arias Sánchez, Arias Sánchez, Mata Segrega, and Ruiz Gutiérrez, 2008).

As part of this same plan, the government launched the program called Multilingual Costa Rica (*Costa Rica Multilingüe*) in 2008. The main aim of this program is to improve the population's language skills to meet the current demands of communication in English in the country's workforce. Through this initiative, the government has tried to organize language learning and teaching practices in the Costa Rican public sphere. According to Quesada Pacheco (2013), this program has been developed on several fronts. It has created activities involving students from private schools who tutor students from public schools, and community

conversation groups, involving volunteers serving as language partners in different communities. Primary and secondary teachers have also received training in a large variety of workshops in different areas of specialization in the four state universities. In addition, community workshops are being offered to provide young workers access to jobs requiring English skills. And among other opportunities, university professors have been able to spend time in the United States for further training.

In Costa Rica, according to Article 78, Section VII of the National Constitution, public primary and secondary education (through ninth grade) is obligatory, free, and funded by the state, and has been since 1949. After ninth grade, it is free and state-supported, but no longer obligatory. In the public sector, English is taught at the primary and secondary levels to different extents in different institutions. Table 2 illustrates the distribution of levels in the national system and indicates the students' approximate age and number of classes taught.

Table 2. English instruction in the mainstream Costa Rican educational system

Levels	**Approx. Age**	**English Instruction**
Pre-school		
Pre-Kindergarten	4-5	None
Kindergarten	5-6	Four 40-minute classes per week (optional)
Primary education		
First through third grades	6-9	Three to five 40-minute classes per week, depending on school resources and conditions
Fourth through sixth grades	9-12	Three to five 40-minute classes per week, depending on school resources and conditions
Secondary education		
Seventh through ninth grades	13-15	Three 40-minute classes per week (mandatory)
Tenth through eleventh (or optionally twelfth)	16-17/18	Five 40-minute classes per week (mandatory)

According to the National Board of Education (*Ministerio de Educación Pública*, 2013), in primary school, each institution depends on the different conditions, budgetary restrictions, personnel availability, student conditions and needs, infrastructure and zoning conditions to decide on the time allocated to different subjects in the students' curricula. What is defined as *basic subjects*

(Spanish, Mathematics, Social Studies, and Science) is obligatorily taught following a strictly assigned number of hours and resources.

For the remaining subjects, known as *special subjects*, the number of hours may be adjusted to accommodate the particular needs of the institutions, that is the case of English subjects. English courses, as part of the Foreign Language scheme, may be reduced from five to three classes per week during primary education to adjust the number of total hours of the school to a specific official curricular plan. In addition, the National Board of Education notes that these might be adjusted to interactive-radio class modality in cases where teacher availability demands it.

In secondary education, all schools are expected to abide by the standard regulations and teach between three 40-minute classes (7^{th}–9^{th} grade) and five 40-minute classes (10^{th}–11^{th} grade) per week. This is the case in all secondary schools, partly because students are expected to take a compulsory national English examination at the end of 11^{th} grade. English classes are typically taught following Foreign Language Teaching methodology all over the country.

Table 3 presents the current objectives for 11th graders in mainstream education (*Ministerio de Educación Pública*, 2016) in the public sector as stated by the National Board of Education, based on the CEFR. Thus, these objectives apply to the vast majority of secondary students in the country. This information is available at <http://www.mep.go.cr/programa-estudio/ingles-1>. According to these objectives, students are expected to reach a B1 level.

Students in mainstream education rarely reach a level of B1. This is more likely to occur in private schools where more hours of instruction and different approaches and opportunities are offered. Students attending the Public Bilingual Experimental School are expected to attain a B2 level in the CEFR, but no students have yet graduated from an institution of this type because none of them are implementing a full program as it is described on paper. Several programs where the plan is being partially implemented have been in operation for over ten years now. More details about the latter type of school are discussed below.

Apart from the public instruction analyzed, numerous private institutions emphasize different foreign languages. Many of them focus on English teaching under different approaches. One of them is Content Based Instruction, as occurs in one of the schools in the present study. Regardless of the type of school, they follow the guidelines set by the National Board of Education for English instruction in terms of objectives and goals, and then they add the number of hours and activities of their choice.

Table 3. Objectives for 11th grade student in mainstream education (Ministerio de Educación Pública, 2016)

Listening	**Reading**	**Speaking**	**Writing**
• Can note the main points of oral texts at a normal speed related to public (transactional) and vocational (professional) domains. **INTEGRATION OF LANGUAGE SKILLS** • Can interact in conversations in order to share oral/written information (listening to speak/write). • Can listen to different opinions or points of view of native speakers of English in order to express argumentations (listening to speak).	• Can manage more complex texts in the public (transactional) and vocational (professional) domains within transactional (job applications/situations, rental forms). • Can apply strategies before, during and after reading a text in order to comprehend more accurately longer texts (narrative, expository). **INTEGRATION OF LANGUAGE SKILLS** • Can extract specific information in short media reports, straightforward factual and printed texts to share inferences (reading to write/speak). • Can interpret passage for explaining own strategies to convey meaning about controversial issues (reading to speak/write).	• Can maintain interactions related to public and vocational domains with emerging fluency while still relying on practiced language patterns. • Can ask for clarification, check for comprehension in less predictable situations. • Can explain a problem and propose a solution. • Can handle routines and familiar topics with spontaneous utterances. **INTEGRATION OF LANGUAGE SKILLS** •Can rehearse presentations supported by notes and illustrations and peer feedback (writing to read/speak). • Can express arguments/opinions when exchanging interactions (listening to speak). • Can interact in long conversations with peers and others about controversial issues using interaction and production strategies (listening to speak).	• Can write longer texts with more varied sentence structures and a wider range of vocabulary through a series of connected paragraphs within expository (advertisement, speech), narrative (biography, personal log) and transactional (question and answer) genres. Sentences connect with each other in a more cohesive manner. **INTEGRATION OF LANGUAGE SKILLS** • Can organize thoughts and opinions expressed by group participants for writing persuasive paragraphs (listening to write). • Can write about online or printed controversial issues for presenting meaningful information to audiences, e.g. Impromptu speech (reading to write/speak).

3.7. PUBLIC BILINGUAL EXPERIMENTAL SCHOOLS (PBES)

The bilingual public education experience did not begin in Costa Rica until 1995 when the first Public Bilingual Experimental Schools (PBES; *Liceos Experimentales Bilingües*) were created. As indicated by Granados Sirias & Chavez Soto (2018), nineteen different PBES were founded between 1995 and 2013. They cover fourteen different areas from the twenty-seven existing regions specified by the National Board of Education. They are located mainly in non-central communities with only a few in more central areas. These schools emphasize English-language learning and offer a total of 14 English classes a week, more than twice as many as students receive in a non-bilingual public school (only 3 per week in the first three years, and 5 per week in the last two years). In the PBES, on the other hand, classes are structured as follows: They offer ten 40-minute English periods, five of which emphasize listening and speaking skills, and the remaining five sessions focus on reading and writing. Along with these language-focused classes, they offer four 40-minute periods of Literature in the English Language. For those language sessions, groups are split into two sub-groups; for the latter, the Literature classes, the groups are not divided.

According to Granados Sirias & Chavez Soto (2018), the Higher Education Council (*Consejo Superior de Educación*) approved a new series of regulations for Public Bilingual High Schools in 2009. To comply with these new conditions, the PBES would eventually have to adjust to different regulations to qualify for a new category. Depending on the number of hours and subjects taught in English, the PBES would then be known as *Public Bilingual High School I, II, and III*. The above regulations are carefully described in the "Guidelines for the Development and Regulation of Bilingual Public High Schools of Costa Rica" (Granados Sirias & Chavez Soto, 2018). A few are given below:

- These schools must abide by the same laws and follow the same calendar and the same evaluation regulations that apply to all public schools, as well as the specific regulations given to them in that document.
- They should have the categories of "Model Institution" and of "Institution of National Interest," which would allow students from other communities to apply and attend these schools.
- They must represent a second option for schooling in the area; that is, there must be a regular (non-bilingual, public) high school in the same county.
- They must have no more than 750 students.

- The number of students per class will be no less than 20 and no more than 25.
- Students are required to take an admission test.
- They must emphasize the use of the English language in all subjects both when teaching and when evaluating.
- They must promote cultural interaction in English for both students and teachers.
- They must promote student involvement in extracurricular activities oriented toward the community.
- They must teach 25-60% of their curriculum in English.

Table 4 indicates the number of hours to be taught in English in each type of high school, but none is applying this new set of regulations. Only the original Public Bilingual Experimental Schools are operating, but for now without meeting the requirements to qualify for the new categories. This shows the Public Bilingual Education system still concentrates mostly on reinforcing language classes rather than using more innovative methods (CLIL or any type of instruction where English is used as a way to teach content). A literature class is taught, but it represents only 4 hours of class a week and it is still strongly oriented as an EFL course.

Table 4. Hours of instruction in each type of school (Granados Sirias & Chavez Soto, 2018)

Type of school/Hours per week, per level	**7th grade**	**8th grade**	**9th grade**	**10th grade**	**11th grade**
(Current) Public Bilingual Experimental High School	14	14	14	14	14
Public Bilingual High School I	14	18	18	20	20
Public Bilingual High School II	14	18	22	25	25
Public Bilingual High School III	14	22	27	32	32

Despite the efforts made, the public educational system has not yet been able to provide the type of bilingual education that this moment in history demands of Costa Ricans. Many people, therefore, have turned to private primary and secondary education to find the language skills needed to function in an ever more globalized world. Historically, Costa Rica has had different private schools concentrated on teaching different languages (German, French and mainly English). In the recent decades, given the socio-economic demand of an English-speaking work force, private schools have sprung up all over the country. The

province of Heredia is a clear example of this. According to the report on the Heredia School Region (*Región Educativa de Heredia*, 2015) in the last decades, over 75 private schools and high schools have emerged in this province in an attempt to respond to that need of foreign language proficiency. According to Granados Sirias (2016, personal communication), the total number of secondary schools in this country is 1,165 (951 are public, 18 are partly funded by the state, and 196 are private). In the province of Heredia alone, there are 27 private secondary high schools, but only a couple of them use English as a medium of instruction. Instead, they reinforce English-language classes; this is the most common way found in Costa Rica to promote bilingual education, in both the public and the private sector. In the last couple of years, two schools have started to function as Public Bilingual Experimental Schools, one in the region of Belén and recently one in Santo Domingo (still at a pilot stage). This panorama shows that bilingual education as described by García (2009) or Cummins (2014a) is not really present in Costa Rica. It also shows that there is a dire need for a change in the Costa Rican educational system.

3.8. Concluding Remarks

Aware of the many benefits that bilingualism brings to a bilingual's life, the Costa Rican society is now striving to offer new generations more opportunities to have access to bilingualism. Along the lines of Baker (2011), there is awareness that bilingualism brings cognitive, educational, social, economic and cultural benefits to those who enjoy this ability. The description provided in the previous sections provides a general idea of the current state of English-bilingualism and bilingual education in Costa Rica. Of course, no absolute generalizations are to be drawn from it; it is presented here as an overview of the context. What can be concluded from the references made regarding the Costa Rican context is that it does not yet enjoy the benefits that well-founded Immersion, Dual Education, and CLIL programs could bring to the process of bilingualism. An approximation to CLIL, in the form of Content Based Teaching is limited to a very few private schools, and it would appear to be almost absent even from the Public Bilingual Experimental Schools, a type of school that few students have access to in this country. At least on paper, it looks like these schools would eventually implement content-based instruction. With that idea in mind, we are working toward a more updated, non-traditional type of bilingual education in Costa Rica.

CHAPTER 4

VOCABULARY IN SECOND LANGUAGE ACQUISITION

This chapter addresses issues related to the general nature of vocabulary in the area of second language learning. Through the study of vocabulary, the present work analyzes the effects that bilingual instruction (allegedly a strong form of bilingual education through Content Based Teaching) and EFL instruction have on students' learning their L2. To establish a theoretical connection between the previous sections and the study discussed later on this thesis, we now set the basis that support a subsequent analysis. This section, thus, serves as the link between the theoretical concepts discussed before for bilingualism and bilingual education and their particular role in specific aspects of language learning such as vocabulary. The intention of analyzing SLA through the lens of vocabulary knowledge will lead to a better understanding and interpretation of the results obtained in the present study as discussed in the second part of the thesis. In so doing, this section lays the groundwork for the successive analysis of this study.

We will start this section by discussing the role that vocabulary knowledge plays. Beginning with some of the most important characteristics of vocabulary learning in SLA, attention is then given to key ideas and concepts put forward by some of the most renowned specialists in the field. The section on the distinction between breadth and depth of vocabulary follows, providing a more detailed description and analysis of the concept of vocabulary breadth with relation to vocabulary size; and of vocabulary depth in regard to types of vocabulary knowledge. The discussion continues with the notion of intentional and incidental learning, the difference between them, their application, and their effect on vocabulary development. The next section on vocabulary teaching provides some insight into what research has shown to be the best ways to deal with vocabulary in the classrooms. This analysis stresses the role of vocabulary within SLA. Finally, a reference is made to the mental lexicon and its semantic networks with the idea of illustrating a possibility of how vocabulary is organized in our brains.

4.1. VOCABULARY

Learning vocabulary is an essential part of being proficient in a second language. As learners start the second language acquisition process, vocabulary learning is one of the first tasks that they embark upon. As they progress, they will still continue to learn new vocabulary, even well after they have become fluent in the language and can use the second language proficiently in different tasks involving listening, speaking, reading and writing skills. This applies to second

language learners and to native speakers alike. Dóczi and Kormos (2016, p. 12) argue that "among all the aspects of language, it is lexis that is prone to the greatest change and development during one's life; therefore, the mental lexicon is in constant flux."

The vocabulary learning process, described as *daunting*, *challenging* or *a hurdle* (Schmitt, 2008, 2010a), is one of the aspects of language that has attracted the attention of learners, teachers and researchers for a long time. The study of vocabulary in second language learning has received an upswing in attention from researchers and other specialists in recent decades. According to Nation (2011), over 30% of the L1 and L2 vocabulary research that took place over the last 120 years was concentrated between 1999 and 2011; the steady growth in vocabulary studies after 2011 continues to support this renewed attention to the field of study.

Laufer and Nation (1999, p. 33-34) acknowledged the original upsurge of vocabulary studies: "[t]his increased interest can be interpreted to mean that there is considerable value in gaining knowledge about specific parts of language learners' proficiency because it can be used effectively for diagnostic, placement and curriculum-design purposes." Several other researchers have also recognized the importance of vocabulary studies. Nation (1993, p. 132) insists: "If learners do not develop a sufficiently large vocabulary then skill in language use and the knowledge that comes from that skill will not develop." This points to the importance of vocabulary development as a basis for the proper progression of language skills in particular, and cognitive development in general. For Read (2000, p. 14), "vocabulary is not just a set of linguistic units but also an attribute of individual language learners, in the form of vocabulary knowledge and the ability to access that knowledge for communicative purposes."

Along these lines, Schmitt (2010b, p. 4) asserts: "vocabulary has strong relationships with the language skills." Moreover, Coombe (2011, p. 122) states: "[a]lthough once viewed as an enabling skill of secondary importance to the four main language skills, vocabulary instruction has now come into its own." From this, we can gather that vocabulary no longer stands as a separate component of language attributes, but rather, it is acknowledged as a cornerstone for building and developing language capacities instead. Aitchison (2012, p. 53) highlights the importance of vocabulary by saying that "[w]ords […] are precision instruments which should be used with care and accuracy." All these quotes point to the key role of vocabulary in language development. This new perspective on vocabulary has granted this language aspect the attention deserved and has resulted in more in-depth studies on the field.

Nation (1993) discusses the different views that describe vocabulary knowledge and the role it plays in language learning. The first view, the *instrumentalist view*, sees vocabulary as a factor that serves as a prerequisite for comprehension; its presence, thus, enables comprehension. The *aptitude view* sees vocabulary as a result of having a good brain; an outcome of this good brain is adequate comprehension. A third view, the *knowledge view*, identifies vocabulary as a feature of suitable world knowledge that is translated into fluent access to vocabulary via strong networks and associations. In this view, reading comprehension is possible because the reader brings as much information (through vocabulary knowledge) to the text as he/she extracts from it. Nation (1993) notes that all these views of vocabulary knowledge should be seen as contributing parts of vocabulary knowledge development as they work intricately together in vocabulary growth. Nation argues that as proficiency skills improve, they enable the knowledge of the world to grow. Knowledge of the world, in turn, requires vocabulary knowledge to continue expanding. Vocabulary knowledge, in addition, causes students to broaden their range of expertise while increasing specialized vocabulary-learning strategies that develop independently from background subject matter knowledge. After all, vocabulary is not an isolated factor in language learning and communication. As Nation and Waring (1997) maintain, vocabulary should be seen as a component of language skills and not as a unique requirement for language acquisition and production. This consideration ought to guide teaching and learning practices dealing with vocabulary.

Although native tongue vocabulary appears to develop effortlessly and quickly in beginning stages, second language vocabulary learning requires much more careful attention. Schmitt (2007) insists that learning English vocabulary certainly represents a challenge. He argues: "English probably contains the greatest number of words of any major language, which makes learning a sufficient amount of its vocabulary a formidable task" (p. 745). He goes on to say that while English operates like other languages in the sense that it combines words and adds suffixes to create new words, it also requires the learning of a considerable number of words that are not systematically transparent, and which are thus more demanding.

Research shows that students' vocabulary size is linked directly to their ability to complete different tasks, from basic oral communication to reading novels in the target language. Read (2000, p. 1) sees words as "the basic building blocks of language, the units of meaning from which larger structures such as sentences, paragraphs, and whole texts are formed." These and several other features of vocabulary lend paramount importance to vocabulary studies. Given that lexical

knowledge serves as a cornerstone for L2 acquisition and has a direct impact on students' learning of the second language, it is a key element for language teachers and researchers tracking students' language progress. Meara (2009) insists on the significance of acknowledging the differences between native and non-native speakers concerning vocabulary acquisition, degree of knowledge, and use.

Specialists in vocabulary express different ideas about how to proceed about teaching and learning vocabulary. Schmitt (2007) speaks of the need for instruction both in basic vocabulary and in developing vocabulary-learning strategies, to aid students in strengthening this learning process. Pigada and Schmitt (2006) argue in favor of extensive reading as an effective way to acquire vocabulary. They add that reading can enhance vocabulary acquisition in terms of spelling, meaning and grammar. Nation (2013) ascribes a great deal of importance to high-frequency words, the most commonly used words, and insists that they play such an essential role in learners' language that both instructors and learners alike should invest time in them. For Nation (2013), the frequency, coverage and range of the high frequency words validates dedicating time to them inside and outside of the classroom. Nation (1983, 2006b, 2013) established a division of vocabulary based on word bands that represent frequency of vocabulary use at 2000, 3000, 5000 and 10,000 word levels. In addition, Nation (2006b) mentions, in reference to the different word bands, that whereas the 2000 high-frequency words should be assigned classroom time to be learned, the situation is different for low-frequency words, for which he recommends that instructors spend time teaching strategies such as guessing words from context, using flash cards, memory or a dictionary for students to learn these words outside the classroom.

Moreover, Nation (2006c, p. 448) notes: "[k]nowing a word well involves knowing several aspects covering the form, meaning and use of the word" hence, bringing attention to more in-depth aspects of vocabulary knowledge. Nation (2014, p. 2) provides further rationale on the importance of repetition in the learning process and insists that "learners not only need to gradually meet the most frequent 9,000 word families, but they have to meet them often enough to have a chance of learning them." As for evaluation, Read (2007, p. 121) points to the role of context and states, "[w]e need to complement discrete vocabulary tests with embedded measures of the learner's ability to handle lexical items in context. Traditionally, context has been conceived in linguistic terms as a sentence or larger co-text in which a vocabulary item occurs." Clearly, there are many different coexisting aspects when dealing with vocabulary teaching and learning, and learners may still contribute additional, individual factors to the process.

As with any other process involving learning, multiple factors play a role in second language vocabulary learning. In describing individual differences in vocabulary development, Dóczi & Kormos (2016) comment that the same types of individual differences that affect second language learning also affect vocabulary in second language learning; namely, affective, cognitive and personality-related differences. These authors further maintain that one of the most significant factors affecting efficiency in vocabulary development at the cognitive level is working memory. They found a particularly important connection between phonological short-term memory and vocabulary learning outcomes. Dóczi & Kormos (2016), thus, insist that working memory plays a key role "in committing words to long-term memory, […] it also helps L2 learners harness communicative situations for the learning of previously unknown words, additional word meanings, and other types of word knowledge" (p. 151). In sum, individual aspects are yet another active aspect of vocabulary acquisition.

The above ideas assign much importance to the role of learners, educators and researchers involved in the second language, vocabulary-learning process. Classroom-material designers, instructors and researchers dealing with vocabulary research and teaching must work together to use more efficient methods of vocabulary teaching and learning. The underlying purpose of researchers and educators in examining vocabulary progress should be directed toward finding ways of accompanying students in the process of developing the skills required for using the language effectively. Vocabulary is much more than just a number of concepts that students know. It represents the raw material enabling speakers to build up sentences and eventually allow language use. Researchers should find ways that would eventually facilitate and ensure real communication for students. To achieve that, we must discover what students know in terms of vocabulary, so that the necessary adjustments can be made in both curricula and classrooms. The role of vocabulary is key to benefit or impair students' progress in the SLA process. As a crucial component of language learning, vocabulary should be treated as an active element in second language classrooms.

4.2. Breadth and Depth of Vocabulary

A distinction has traditionally been made between the number of words that learners know, referred to as vocabulary *breadth*, describing vocabulary size; and the quality, that is, how well learners know this vocabulary, denoted as vocabulary

depth (Schmitt, 2008, 2010b, 2014). The distinction in regard to these vocabulary dimensions is associated with a definition provided by Anderson and Freebody (1981, p. 81-82):

> The first [dimension] may be called 'breadth' of knowledge, by which we mean the number of words for which the person knows at least some of the significant aspects of meaning. [... there] is a second dimension of vocabulary knowledge, namely the quality or 'depth' of understanding. We shall assume that, for most purposes, a person has a sufficiently deep understanding of a word if it conveys to him or her all of the distinctions that would be understood by an ordinary adult under normal circumstances. (cited in Read, 2004, p. 210)

Regarding the qualities of vocabulary, Nation (2006c) speaks of the learning burden of vocabulary in reference to several aspects required to know a word, involving form, meaning and use. He maintains that:

> The difficulty these aspects [form, meaning, use] have for any particular word depends on how closely the aspects relate to knowledge the learners already have. This knowledge can come from the learner's first language or other languages they know, and it can come from the knowledge they already have of the second language (p. 448).

He adds that other aspects within the language or in connection to the L1 can help to lessen or increase the learning burden. Some of these aspects are similarity of the written script, sound/form-matching patterns, learning conditions, collocational use, pronunciation features, and cognate vocabulary presence. Considering Cummins's (1979, 1984, 1986, 2000, 2005, 2009) *Interdependence Hypothesis,* we have established that as long as the linguistic aspects are adequately developed in the L1, they will be successfully transferred into the learners' L2. Unfortunately, this is a fact that is most often overlooked in SL settings. Taken together, L1 background knowledge and the developing knowledge in the L2 will serve as a basis for growth and strengthening of L2 vocabulary. Along these lines and addressing the issue of vocabulary depth and breadth, Read (2004) also believes that breadth and depth of vocabulary develop in parallel. For him, as the number of words in one's repertoire increases, so does the learning about the words that one frequently uses.

While vocabulary size is often described as a crucial factor in connection with skills such as reading comprehension (see for example Laufer, 1998; Laufer and Ravenhorst-Kalovski, 2010), there are other factors involved in knowing vocabulary, aside from the number of words for which we can establish a form-meaning connection. Dóczi and Kormos, (2016, p. 63) maintain: "it is important to highlight that the number of exposures needed to remember a particular word along with its semantic characteristics, such as concreteness and imaginability, play an important complementary role in addition to frequency effects." These aspects refer to what was described above as vocabulary depth. Schmitt (2008) highlights the importance of vocabulary depth, especially in connection to productive skills. Schmitt (2008, p. 333), in reference to depth of vocabulary, notes that besides knowing a large number of words, "a learner must also know a great deal about each item in order to use it well" (see also Schmitt, 2010b, p. 15). Nevertheless, according to Schmitt (2014), when it comes to higher frequency words and to learners in possession of limited vocabulary, the difference between breadth and depth measures becomes very limited. This difference amongst the two becomes more evident at upper vocabulary levels and with larger vocabulary sizes.

Contrary to Read (2004), Schmitt (2014) argues that size and depth of vocabulary knowledge do not develop in parallel. He insists that learners could know little information about many different words, or they could know a lot about a limited set of words. Schmitt (2014), however, adds that vocabulary size allows a more direct conceptualization of breadth through counting lexical items, thus representing knowledge of the form-meaning association. On the other hand, depth of knowledge refers to a much more complicated challenge in terms of conceptualization of the notion. Given that situation, researchers may analyze individual (i.e., variation in meaning senses) or more holistic (i.e., connections to other lexical items) features of the items. Depth of knowledge is represented by a broad knowledge of the different aspects that are involved in word knowledge, or to a narrower extent, by a limited knowledge of some of these aspects. It is exactly this variety of elements that conform depth of vocabulary what makes it a difficult phenomenon to measure. For Schmitt (2014), depth can be described in terms of knowing multiple aspects of a word, what he calls its polysemous meaning senses, its derivative forms, its collocation constraints, its fluency in use, or its organization in connections to other words.

This high level of complexity and the fact that it is often difficult to tell breadth apart from depth of vocabulary is what has led some authors to consider depth an inappropriate concept. Read (2004) finds using a single concept a

problematic issue because it refers to at least three different ways of conceptualizing what knowledge of a word represents. He claims that depth should be analyzed in reference to at least three different positions; namely, precision of meaning, comprehension of word knowledge and network/link knowledge. Read (2004, p. 224) suggests dispensing of the term and instead argues: "the dimension of knowledge that [measures of vocabulary] are designed to measure should be carefully defined and not simply labeled with a catch-all term like depth […]" Read (2007) recognizes the challenge that measuring depth, that is, measuring how words become functional units in the learner's language, represents. He also acknowledges the evident lack of an appropriate depth measure (partly justified by the complexity of what depth entails) and the pressing need that there is for it. For Read, how depth studies lag behind becomes all the more evident when comparisons to existing tests such as the Vocabulary Levels Test (used in measuring vocabulary breadth) are made.

Schmitt (2014), agrees with Read (2004) (see also Dóczi and Kormos, 2016) when it comes to the difficulty of pinning down appropriate measures of depth-related aspects. Schmitt (2014) maintains that the notion is problematic to the extent that size, in itself, already involves mastery of depth levels. In addition, he notes that there is clear interrelation between the different aspects of vocabulary and that the distinction between size and depth can be hazy. He concludes that all of these aspects contribute to the difficulty in conceptualizing depth.

The fact that words are not seen as empty, isolated entities and are instead analyzed in connection to the other elements of language is recognized as a step away from theories that once viewed words as simple elements that completed empty slots in sentences. On this matter, Dóczi and Kormos, (2016, p. 11) note:

> As the boundaries between lexis and grammar are becoming increasingly blurred, the information the mental lexicon needs to hold has also been extended; for example, in Levelt's (1989) model of speech production, an essential component of the lexicon is syntactic information related to a particular lexical entry.
>
> These developments have important consequences for the conceptualization of the mental lexicon, as it has become difficult to separate syntactic regularities of language, traditionally seen as "rules," from linguistic construction units (e.g., words, phrases, formulaic expressions, and chunks). This has resulted in the expansion of the different types of knowledge the mental lexicon needs to store […]

Dóczi and Kormos' ideas reinforce the notion of the intertwined role that vocabulary maintains with other aspects of language. As such, the function of vocabulary in connection with productive and receptive skills is fundamental. It is this central role of vocabulary what lends support to his study. While there is controversy as to the appropriateness of using the terms "depth" and "breadth," the distinction remains in use in the research arena. The following sections describe the concepts of breadth and depth of vocabulary to a greater extent, thus still acknowledging this distinction. The current study will concentrate on aspects of breadth of vocabulary and its impact on reading and writing. However, as other aspects of word knowledge inevitably intervene, a theoretical description of depth and its implications is also included in this section.

Although the term vocabulary has been used above as a broad term encompassing passive and active vocabulary regardless of its depth or breadth, an important distinction is made in the literature regarding the difference between passive and active vocabulary. In that respect, this study will embrace the notion presented by Heaton (1990, p. 79) regarding passive or receptive vocabulary as that which "you expect your students to recognize" and that of active or productive vocabulary as that which "you expect your students to use." These concepts will be used with these meanings throughout the study.

4.2.1. Size of vocabulary (breadth) and word lists

Research on vocabulary size and its application for language functions is based on vocabulary levels. Information about frequency and range in the use of words facilitates information about the existing levels of vocabulary (Nation, 1983, 2003, 2013). The relatively easy access to vocabulary size measures has caused breadth of vocabulary to be studied much more than depth. The grouping of words according to different vocabulary levels, in turn, allows learners to have access to groups of words that follow a sequence and whose progressive learning brings greater benefits to students. Once the levels of words have been identified the analysis of their frequency becomes more accessible. Nation (2013) proposes different ways of counting words so that they can be organized into levels. These are tokens, types, lemmas and word families.

According to Nation (2013), one basic way of counting words is using *tokens* or *running words* in a text. The unit of *token* refers to the practice of counting

every single word as it appears in a text, even if it appears more than once. Counting *types*, on the other hand, refers to counting words in a sentence or text without including repeated words in the final word count. That is, if the same word appears more than once in the sentence or text it would be counted only once, as one *type*. As a result of this differentiation between tokens and types, it is not uncommon for a sentence to contain more tokens than types, as a single word (such as the pronoun *it*, for example) may frequently appear more than once in a sentence and even more so in a text. Another possibility of word count is to count words grouping them as *lemmas*. For Nation (2013, p. 10), "instead of counting different types as different words, closely related words could be counted as members of the same word or **lemma**." It follows that all members usually belong to the same part of speech. Bauer and Nation (1993) include in this group the plural forms, the third person singular, present tense, past tense, past participle, *-ing* verb inflections, the comparative and superlative forms and the possessive marker. For Nation counting units as *lemmas* results in great reduction of numbers in a corpus.

The last possibility of word count units uses affixes beyond the point that *lemmas* do; that is, grouping them in *word families*. Word families consider other affixes (both prefixes and suffixes) that behave consistently and thus make it easier for students to learn the words that contain them once they know the base form of the words. For Nation (2013, p. 11), "[a] **word family** consists of a headword, its inflected forms, and its closely related derived forms" (emphasis in original). Bauer and Nation (1993) provide a very detailed categorization of affixes divided into seven different levels. The most basic level describes the inability to recognize basic affixes (such as the plural form *-s*) at Level 1, to the ability to identify inflectional suffixes in Level 2, to frequent derivational forms in Level 3, and so on, incrementing the level of complexity in affix forms as the level increases.

The notion behind Bauer and Nation's (1993, p. 253) classification was that "once the base word or even a derived word is known, the recognition of other members of the family requires little or no extra effort." This allows for vocabulary growth as affixation knowledge develops. The notion of word family applies as long as the meaning of the base and its derived forms are similar when these are standing alone; inability to abide by this principle would mean that the elements belong to different word families. Read (2007) describes the word family as the preferred lexical unit in vocabulary size research. Read (2007, p. 108) defines the concept saying that a "word family […] consists of a base word and its inflected forms, together with derived forms which share a common meaning with the base

form." The presentation of the results for the writing task in this study will consider word types and word families as the word units in its analysis.

4.2.2. Vocabulary levels based on frequency

Based on frequency and range of use across a variety of texts and disciplines, words have been grouped into different lists to represent vocabulary at different levels of frequency. Earlier accounts (Nation, 1983, 2001, 2006b) of these word lists offered a distinction of four kinds of words: high frequency words, academic words, and technical and low frequency words. Nation (2013) offers a slightly different division of word lists, while still maintaining the core principles on which he has based many years of research. The current subdivision contains high-frequency words, mid-frequency words, low-frequency words, and technical vocabulary containing both academic words and technical words. It is important to keep in mind, that as Dóczi and Kormos (2016, p. 63) say, "the breadth of vocabulary knowledge follows the power law of learning, which results in large and rapid initial learning gains which then level off with the development of language competence." This justifies the idea of learning very frequent words first and then moving on to less frequent words later. Nation and Chung (2009) insist on the importance of learning vocabulary by following a sensible order. The following vocabulary levels are based on this 'sensible order' and were created by Nation (1983, 2001, 2006b, 2013) and Nation and Chung (2009) to be used as a guide in vocabulary learning.

The first list contains *high frequency words*, and, Nation (2013) says, is based on "A General Service List of English Words," developed by West (1953). Nation (2013) comments that although some consider this list outdated, it still represents the best base for the level. The high-frequency vocabulary list contains around 2000 word families, both content and function words, and accounts for almost 80% of the running words in an academic text and about 90% of vocabulary in conversations and novels (Chung and Nation, 2003). Nation and Waring (1997), however, note that the percentage of high frequency words may be reduced slightly as low frequency words increase in more diverse types of texts such as specialized texts. In general terms, according to Nation (1983) we can expect to find high frequency words in almost any type of text. This small number or words is key in the second language learning process for two reasons. It is fundamental because of

its role in the transmission of the message in the text and because it represents a large proportion of words in texts.

Nation (1993) insists that failure to learn high frequency words results in inability to achieve fundamental skills in language use. Nation (1983, 2001) also insists that given the importance of these words, time should be spent on them to ensure their learning. Nation (2003) mentions that high-frequency vocabulary may serve as a strong basis to set the context for unknown low-frequency words, thus allowing the inference of meaning of the unknown words. In addition, Nation (2011, p. 531) maintains: "[b]ecause each word in this group is frequent, the learners will get a very good return for learning them," and this justifies investing time, in class and outside of class, in learning this list. Moreover, Nation (2013), comments that the fact that this list is small facilitates the possibility of dedicating time to these words across a language program. Nation (2001) warns that the 2000 boundary can be interpreted as arbitrary, as there are many words in the low-frequency list (now mid-frequency list) that are used moderately and thus for some users these could represent high-frequency items.

On this issue, Schmitt and Schmitt (2014) strongly contest that high frequency words should include the most frequent 3000 (not only 2000) word families. They find this number more fitting and a more "pedagogically-useful criterion" (p. 13). They acknowledge this number as an "important milestone in language development" (p. 13) as well. Interestingly, throughout his studies, Nation has also hinted at the arbitrariness and need for some flexibility in this list. Ideas such as "[t]he 2,000 and 3,000 word levels contain high frequency words" (Nation, 1983, p. 14), and "the learner needs to know the 3,000 or so high-frequency words of the language" (Nation and Waring, 1997, p. 11), exemplify that a certain degree of flexibility is not only feasible but also called for. Regardless of this flexibility, Nation (2013) maintains that 2000 word families are still a suitable representation of high-frequency words, although he also manifests an awareness of the debate over this number. Nation (2013) asserts that learning this set of words should take place through incidental learning as well as direct learning and teaching that allow for repeated exposures to the concepts. He further justifies deliberate attention to this word band on three bases. First, these words permit access to written and spoken discourse containing a substantial proportion of these words. Second, if students do not handle this word level fluently, text comprehension is affected. Finally, knowledge of high-frequency words is necessary for production of written and spoken discourse. High-frequency vocabulary, whether on the basis of 2000 or

3000 families, has proven to be key in vocabulary learning and language development. The present study concentrates on this frequency band.

The second group of words contains the 7000 word families that Nation (2013) calls the *mid-frequency words*. These are the words that range between the 3000 and 9000 word families. In addition, they exhibit a moderately frequent use, for which many, says Nation, were close to being included in the high frequency list; hence, the ongoing debate as to where to draw the line between the different word levels. For Nation (2013), this list of words, together with the high frequency words, are necessary for an independent use of English. Nation (2013, p. 25) argues that these words occur "frequently enough to be a sensible learning goal after the high-frequency and specialized vocabulary is known." This frequency band, although clearly identified in previous accounts, did not receive a label in former descriptions of word lists by Nation (1983, 2006b).

Schmitt and Schmitt (2014, p. 18) deem this label important for the word band to be recognized "as a discrete phenomenon, with its own unique benefits for users, and pedagogical challenges for language practitioners […]." These authors also insist that establishing this word list as a learning goal would yield palpable rewards for students, giving learners the chance to function in various topics and situations. Nation (2013) justifies the creation of this category based on several ideas. The indication of a division (at the 9000-word level) between more and less frequent words is possibly the point at which native speaker vocabularies become specialized according to their interests. The coverage required for reading comprehension, the familiarity of native speakers and highly proficient non-native speakers with these words, and the need of this vocabulary to reach certain learning goals all lend support to the creation of the label.

In terms of the teaching learning process, Nation (2013) argues that rather than investing too much time in teaching these words, instructors should spend time teaching strategies (guessing techniques, flashcards and dictionary use) for learners to deal with the words. Learners, on the other hand, who should have begun training with these strategies while learning high-frequency words, should simply continue the process of building their vocabulary. Laufer and Ravenhorst-Kalovski (2010), on the other hand, disagree with the idea that class time should not be devoted to words outside the high-frequency band. They find that the slight increase in reading scores seen when learners know vocabulary in higher word bands justifies spending time learning these words. For them, given that knowledge of additional infrequent words contributes to improvement in text coverage, investing time in learning these words is worthwhile. To make a decision on this matter, we would have to consider

the particular characteristics and interests of the students and the program to decide what possibility would suit the given population better.

The third word band described in Nation (2013) is that containing *low-frequency words*. These are the words that are grouped here in the 9000-word level family and above; the boundary used to be 10,000 words (Nation, 1983, 1993). He claims that these words are part of an extensive list containing several thousands of them, and they represent around 3% of the words in texts (around 5%, for Chung and Nation, 2003). For Nation (2013), while this is the most extensive group of words there is in the language, it represents only a small proportion of the running words in texts. In addition, Nation notes that these words are rarely used in the language and may be seen as technical terms in various fields. Nation (1983, 1993, 2003) maintains that these words usually play a key role in conveying information in the text, but that due to their low frequency and limited proportion in running text, time should not be invested in teaching or learning them as should be done for high-frequency words. Instead, teachers should train their students in developing strategies such as inferencing, drawing meaning from context and dictionary use to deal with, and eventually learn these words through multiple encounters.

Along these lines, Nation (1993, p. 129) maintains that "low frequency words of a language must grow very slowly and must require substantial amounts of reading or listening to language that contains more low frequency words than colloquial language does." Nation (2003) (see also Nation and Meara, 2010) also argues that while learners need to know low-frequency words, they should best be learned after the high-frequency words, unless a special situation requires them to be learned earlier. He recommends using word part knowledge; that is, being able to recognize and know the meaning for affixes, for example, as a way of learning these words. Nation (2013), on the other hand, suggests incidental learning (reading and listening) as a viable option.

The fourth word band in Nation (2013) describes *specialized vocabulary* and is subdivided into *academic words* and *technical words*. Nation (1983) also viewed academic words (then called University Word List) as part of specialized vocabulary. For Nation (1983, p. 17), "[s]pecialized vocabulary can be treated in much the same way as high frequency vocabulary because it is frequent within a specialized area. Its frequency justifies attention to individual words [...]" Academic words are commonly found across different fields in academic texts of various disciplines (Nation, 2013; Nation and Chung, 2009). Typically, they represent 9% of words in a text. The list contains 570 headwords and the most common Academic Word List is the one created by Coxhead in 2000 (Nation,

2013). Nation (1993) argues that knowledge of the high-frequency words and the Academic Word List renders text coverage of 95%, which offers great advantages for students' comprehension of texts. Chung and Nation (2003, p. 104) describe the Academic Word list as a "specialized extension of the high-frequency words" and Nation (2013) (also Nation and Chung, 2009) insists on the importance of knowing these words for students with academic aspirations.

According to Nation (2013), *technical* vocabulary is intrinsically related to the topic and field of expertise of the text. These words may be common in the field, but they are not as common elsewhere. Nation (2013) proposes certain features to identify these words. Namely, they help establish a connection between the text and the topic, they cover a large proportion of the running text, they differ from area to area, and they can make up for 20% to over 30% of the words in the text. Nation (2013) also comments that high-, mid-, and low- frequency words for other fields may be part of the technical lists of a given area. As occurs with academic words, learning technical words can be beneficial for learners interested in pursuing university studies, because it represents an efficient way of learning vocabulary for specific purposes.

Nation (1993) (also Nation, 2003) insists that technical vocabulary has to be learned while dealing with the respective discipline and should be studied after academic vocabulary is already mastered. He further insists that this vocabulary "must develop as knowledge in the field develops […] Knowledge of vocabulary is a result of mastery of the knowledge of the field" (p. 121). Along those lines, Nation (2001) views specialized vocabulary as a way of extending high-frequency words for a specific purpose and he comments that these vocabularies come into existence as a result of a systematic restriction of the topics of a given discipline. In addition, Nation (2001) argues that that specialized vocabulary may sometimes make sense only in the context where it applies to the discipline under study.

Chung and Nation (2003) studied technical vocabulary in two areas of specialization (Anatomy and Applied Linguistics) and determined that the percentage of technical vocabulary varies depending on the discipline. For some, technical vocabulary can be a burden while for others it is not (i.e., 37.6 % of the vocabulary is technical in the Anatomy book analyzed while for the Applied Linguistics book it represents 16.3%). In addition, they concluded that technical vocabulary could be more restricted to the field in some disciplines than in others (i.e., 64.4% of vocabulary in Anatomy, and 11.6% in Applied linguistics is restricted to the area). Chung and Nation (2003) provide a list of clues that can help identify technical vocabulary. The terms are typically based on Greek or Latin

forms; they tend to be followed by explanatory definitions; definitions may also contain a typographical marking such as bold or italic font; and they may be repeated to highlight importance.

Because technical vocabulary can be large and can account for a large part of specialized texts, Chung and Nation (2003) provide a few suggestions to learn this vocabulary. First, learners (and teachers) should identify unconventional uses of a word and how these relate to regular uses of the term. Second, when the term represents an extension of a word that is generally used, establishing a connection between the technical sense and the core sense of a word is useful. Finally, learners should identify whether the technical use of the word required specific collocations or grammatical constructions that differ from other uses. This vocabulary band is best learned through productive practice of the terms. Chung and Nation (2004) followed up on the previous study and confirmed its findings. They concluded that technical vocabularies can, indeed, be very large (more for some disciplines than for others); technical vocabularies account for large proportions of the running text; and technical vocabularies differ in the kinds of words that they contain, some being more peculiar to the particular field than others. Specific learning of this list, again, is linked to the needs that the given population has.

4.3. VOCABULARY KNOWLEDGE

Why are frequencies of vocabulary so important? What information can frequency bands give us? The following paragraphs show why word type and family frequencies are essential for understanding how language develops. According to Nation (2013 p. 29), "[a]dult native speakers have receptive vocabulary sizes of around 20,000 word families and learners who already know the mid frequency words may want to see native speaker vocabulary size as a learning goal." Although ideal, the figure does seem *daunting* as Schmitt (2008) describes it. Nation and Waring (1997) also note that this 20,000 figure, usually assigned to native speakers, grows in a rough estimate of 1000 families per year. This makes the number even more difficult for non-native speakers to attain. Frequency counts, however, may facilitate the goal a little, as according to Nation and Waring (1997), 2000-3000 word families may result in adequate productive use, and 3000-5000 word families provide a good basis for comprehension. Therefore:

> a small number of the words of English occur very frequently and if a learner knows these words, that learner will know a very large proportion of the running words in a written or spoken text. Most of these words are content words and knowing enough of them allows a good degree of comprehension of a text. Nation and Waring, (1997, p. 9)

Seminal work by Nation (1983, 1990) has served as a guide for vocabulary studies. Of particular significance is his determination in establishing vocabulary measures for teaching practices and program designs and his contribution to the development of the Vocabulary Levels Test (VLT) (Nation, 1983, 1990; Schmitt et al., 2001), and the subsequent Productive Vocabulary Levels Test (PVLT) (Laufer & Nation, 1999). Both versions include a set of tests divided into word bands according to the frequency of occurrence of words in the language: 2000, 3000, 5000 Academic Word List (AWL), and 10, 000 word bands as described in the previous section. The first 2000-word list is made up of "high frequency vocabulary" (Nation & Hwang, 1995), and these are the words that students should aim to learn first. As mentioned above, Nation and Chung (2009) (see also Laufer & Nation, 1999) discuss the importance of learning vocabulary according to its frequency of occurrence and they insist on giving special attention to the first 2000 words, not only because these represent 80 to 90% of running words in texts but also because knowing them allows learners to reap greater benefits. Curiously enough, the 2000-word band also has been identified as the dividing line between high and low frequency vocabulary (Laufer & Nation, 1999).

Nation (2006b, p. 494) argues that "[t]he most frequent 2000 words of English, which include most of the function words, are the essential widely used words of a language." Read (1988) also finds that the most frequent words in a language are the first ones learned and that any knowledge of vocabulary pertaining to upper levels will mean, by default, that students know that of the preceding levels. For Read (see also Schmitt et al., 2001), vocabulary knowledge grows progressively from the simple, most common words toward the least frequent, more specialized words. Concerning vocabulary measures, Nation & Webb (2011, p. 245) point to how "[m]easures of lexical richness should allow us to distinguish between the language of more and less proficient learners." Having access to this information provides educators with the tools required to promote practices that can aid learners in the language-learning process and help them advance in proficiency. In that vein, this investigation aims to determine how many of those 2000 key words are known, at the receptive and productive levels, by the participating students.

Researchers have provided figures on the number of words required for certain tasks. For speaking, Adolphs and Schmitt (2003) claim that a vocabulary of around 2000 words is necessary for a basic conversation, but according to other vocabulary specialists, if students want to obtain an appropriate comprehension of a text, they need a passive vocabulary of at least 3000 word families to understand 95% of a text (Laufer 1989, 1992); or from 8000 to 9000 word families for a 98% text coverage (Hu & Nation, 2000; Nation, 2006a). For listening, Nation (2006a) claims that lexis of 6000 to 7000 word families is required to comprehend oral texts successfully. What it comes down to is, as Schmitt (2010b, p. 4) contends, "vocabulary has strong relationships with the language skills" and depending on the task we perform we may require different numbers of words. To illustrate how word counts can help us visualize the progress of vocabulary development and to understand the contribution of vocabulary frequency lists in general, Table 5 summarizes some of the results of selected studies and includes the approximate number of words that they claim to be required for different tasks.

Table 5. Research findings for vocabulary coverage and word families

Study	**Percentage of text's vocabulary coverage**	**Word Size**	**Ability**
Laufer (1989)	95%	5000 word tokens	Satisfactory reading comprehension
Hu and Nation (2000)	98%	Not mentioned	Unassisted comprehension of a fiction text
Nation (2006a)	98%	8,000-9,000 families	Unassisted reading comprehension (newspapers, novels)
Nation (2006a)	98%	6000-7000 families	Comprehension of spoken text (children's movies, unscripted spoken English)
Laufer and Ravenhorst-Kalovski (2010)	98%	6000-8000 words	Independent reading comprehension
Laufer and Ravenhorst-Kalovski (2010)	90.56%	3000 words	Adequate independent reading
Schmitt, Jiang and Grabe (2011)	98%-99%	8,000-9,000 families	Independent reading for comprehension

Hu and Nation (2000) emphasize that a 98% vocabulary coverage should not be seen as an isolated requirement for reading comprehension but as one of several other skills and knowledge necessary for adequate comprehension. Also, for Hu and Nation (2000), L1 reading ability, knowledge of grammar, L2 reading

experience, and background knowledge are all seen as contributing elements in the process. Laufer and Ravenhorst-Kalovski (2010) argue that the percentage of text coverage required varies depending on what the researcher considers "adequate," or whether it describes independent or aided reading ability. This may justify number differences in the table for activities that appear to be similar. As shown in Table 5, depending on the purpose of the learner, different word sizes are needed to perform different tasks. For these studies the minimum, however, is 3000 word families. Better understanding and comprehension clearly demand higher levels of vocabulary size. A more detailed description referring to vocabulary in connection to reading and writing will be presented in the analysis section of this study, where vocabulary measures in connection to reading and writing will be discussed.

Further rationale for the connection of vocabulary size and language skills is brought forward by Stæhr (2008), who has conducted research on the possible correlation of receptive vocabulary size and the skills of reading, listening and writing. He found vocabulary size to be highly correlated with reading (0.83) and writing (0.73), and moderately correlated with listening (0.69). Stæhr (2008) also extracted other interesting findings from his study, particularly in regard to writing. He found that learners with larger receptive vocabulary used more words from the upper vocabulary levels, both low frequency and academic vocabulary, than from the high-frequency band. This lent more sophistication to the writing of these students. Stæhr (2008) concludes that there is a strong connection between vocabulary size and the quality of writing. For the receptive skills, he found a stronger association between vocabulary size and reading than between vocabulary size and listening. Stæhr (2008) also determined the 2000 word level to be a *probabilistic threshold* as students reaching this vocabulary size performed above average in listening, reading and writing tests. Attention is again drawn toward the importance of the 2000 word band for the development of language skills.

These studies show how important it is to know the students' vocabulary size, either to provide opportunities to learn a word band, or to determine the next list of vocabulary to focus on according to the needs of the population.

4.3.1. Word knowledge (depth)

For Nation (2013), the concepts *receptive* and *productive* apply to several types of language knowledge and language use (see also Nation 1990, 2001).

Nation (2013, p. 50, emphasis in original) notes that "[i]n general it seems that receptive learning and use is easier than productive learning and use, but it is not clear *why* receptive use should be less difficult than productive." In his discussion he provides a detailed list of questions to determine the kind of knowledge that one can have of a word. For Nation (2013), "[a]t the most general level, knowing a word involves form, meaning and use" (p. 48). Dóczi and Kormos (2016, p. 9) argue that the definition of word knowledge can be studied from two different perspectives: "[o]n the one hand, one can consider how well students know a particular word, which is a "word-centered" conceptualization; on the other, we can perceive depth of word knowledge as knowledge about how words relate to each other in the lexical system." Regardless of the level of knowledge assigned to words, knowing a word certainly means knowing many different aspects. Language learners, like native speakers, do not have a command of all these forms of knowledge categories for many of the words they know, or they may have certain knowledge only at the receptive level (Schmitt and Meara, 1997).

Schmitt (2007) acknowledges the fact that truly knowing a word involves knowing various aspects of words. He insists that an awareness of more aspects of word knowledge enables learners to use the word in the right context and in the right form. Schmitt (2007) further argues for incremental knowledge characteristics in the process of learning a word. According to this feature, for Schmitt (2007), different types of word knowledge are mastered at different levels, "[v]ocabulary acquisition is not only incremental, but it is incremental in a variety of ways" (p. 749). This means that not all types of word knowledge are learned simultaneously, that word learning is incremental at individual word level aspects, and that word knowledge progresses at the receptive and productive level independently of the different levels of knowledge. Some types of word knowledge may be acquired at one point in time while others may not yet have been acquired or will never be.

From Schmitt's (2007) description, it can be understood that learning a word is a complicated process where learning progresses at different rates in different levels of knowledge. Schmitt (2007) describes it as a complicated, gradual process. He concedes that certain aspects such as meaning and grammatical functions could be taught explicitly, but he also recognizes that features such as collocation, register or degree in frequency of use can only be learned through contextual information provided via different encounters with the words. Dóczi and Kormos (2016), along these same lines, argue that while vocabulary develops incrementally, this happens in a non-linear fashion characterized by language development periods in which vocabulary knowledge fluctuates and shows features of attrition as well as growth.

Along these lines, Schmitt (2010a, 2010b) also admits that knowing a word, that is, knowing what a word means and knowing how to spell it, probably represents very basic, initial knowledge of a word. This form-meaning connection is recognized, although crucial, as the minimal specification of word knowledge and thus the first step in vocabulary learning. The rest of the aspects, as listed by Nation (2013), would require greater effort and contextualization. Schmitt (2010b) refers to a number of aspects of meaning that deserve additional consideration, the first of which are "imageability" and "concreteness"; that is, whether learners can easily imagine what the concept represents and whether learners can experience the concepts through their senses as something that is concrete. This distinction is important as Schmitt argues that imaginable, concrete words are more easily learned.

Second, for the distinction of "literal" vs. "idiomatic meanings"; literal meanings are easier to learn than idiomatic meanings. Third, regarding "multiple meaning senses"; for Schmitt, many English words convey more than one sense, and tend to require more attention than those conveying a single sense. Fourth, "content vs. function words"; the grammatical or functional vocabulary represents a greater challenge for learners, and it is also more difficult for teachers or researchers to test. Fifth, "intrinsically difficult words"; words whose length or grammatical class varies in terms of language system, irregular words, words that behave differently in connection to other words can all complicate learning, says Schmitt. Sixth, "network connection (association)"; that is, the different types of links created amongst concepts and word senses in the learners' mental lexicon, and which are supposed to enable vocabulary growth can also facilitate or interfere with vocabulary learning. According to Schmitt, if the connections are not strong enough, or if these connections are not created, the development of the form-meaning link and its contribution to vocabulary growth would be hindered.

Now, returning to the distinction proposed by Nation (2013) regarding types of word knowledge, these are grouped under the notions of form, meaning and use, for both receptive and productive vocabulary. Tables 6 and 7 have been adapted from Nation's Table 2.1 (2013, p. 49): "What is involved in knowing a word."

These tables include the diverse aspects of word knowledge proposed by Nation (1990, 2001, 2013), but instead of providing only those questions which should be asked to recognize what one knows about a word (as he does), we have combined here both the questions and the answers to these questions, as they are shown subsequently in the description by Nation (2013, see lists, p. 48, 50). For this purpose, slight adjustments have been made in wording and in format.

Table 6. Receptive vocabulary knowledge. Adapted from Nation (2013, p. 48–50)

VOCABULARY: THE RECEPTIVE DIMENSION		
Aspects		***What is involved in knowing a word?***
Form	**Spoken**	**What does the word sound like?**
		Being able to recognize the oral form of the word when it is heard.
	Written	**What does the word look like?**
		Being familiar with its written form so that it is recognized when it is encountered in reading.
	Word parts	**What parts are recognizable in this word?**
		Recognizing that it is made up of the parts *under-*, *-develop-*, and *–ed* and being able to relate these parts to its meaning.
Meaning	**Form and meaning**	**What meaning does this word form signal?**
		Knowing that *underdeveloped* signals a particular meaning.
	Concept and reference	**What is included in the concept?**
		Knowing what the word means in the particular context in which it has occurred; and knowing the concept behind the word which will allow understanding in a variety of contexts.
	Associations	**What other words does this make us think of?**
		Knowing that there are related words like *overdeveloped*, *backward* and *challenged*.
Use	**Grammatical functions**	**In what patterns does the word occur?**
		Being able to recognize that *underdeveloped* has been used correctly in the sentence in which it occurs.
	Collocations	**What words or types of words occur with this one?**
		Being able to recognize that words such as *territories* and *areas* are typical collocations.
	Constraints on use (register, frequency)	**Where, when and how often would we expect to meet this word?**
		Knowing that *underdeveloped* is not an uncommon word, and that it is not a pejorative word.

In addition, for further clarity, the original table has been divided into two: The receptive dimension is presented in Table 6, and the productive dimension in Table 7. As in Nation (2013), the word “underdeveloped” is used to illustrate his explanations.

In sum, these tables illustrate how complex knowledge of a word really is. The different dimensions that knowledge has demonstrate the importance of offering learners a wide variety of encounters with vocabulary that provide them with the necessary input to acquire the form, meaning and use senses of words.

Table 7. Productive vocabulary knowledge. Adapted from Nation (2013, p. 49–50)

VOCABULARY: THE PRODUCTIVE DIMENSION		
Aspects		***What is involved in knowing a word?***
Form	**Spoken**	**How is the word pronounced?**
		Being able to say it with the correct pronunciation including stress.
	Written	**How is the word written and spelled?**
		Being able to write it with the correct spelling.
	Word parts	**What word parts are needed to express the meaning?**
		Being able to construct it using the right word parts in their appropriate forms.
Meaning	**Form and meaning**	**What word form can be used to express this meaning?**
		Being able to produce the word to express the meaning *underdeveloped.*
	Concept and reference	**What items can the concept refer to?**
		Being able to produce the word in different contexts to express the range of meanings of *underdeveloped.*
	Associations	**What other words could we use instead of this one?**
		Being able to produce synonyms and opposites for *underdeveloped*.
Use	**Grammatical functions**	**In what patterns must we use this word?**
		Being able to use the word correctly in an original sentence.
	Collocations	**What words or types of words must we use with this one?**
		Being able to produce words that commonly occur with it.
	Constraints on use (register, frequency)	**When, where and how often can we use this word?**
		Being able to decide to use or not use the word to suit the degree of formality of the situation…

Along the lines of analyzing word knowledge, Laufer and Goldstein (2004) propose a hierarchy with four degrees of knowledge that are established based on their strength. Laufer and Goldstein (2004) studied passive word knowledge, typically associated with listening and reading, and active word knowledge, commonly associated with speaking and writing. As a result, the authors claim that some degrees of knowledge are more advanced than others and thus entail the presence of the less advanced degrees of knowledge. They analyzed both passive and active knowledge in terms of passive and active recall and recognition. That is, their analysis moved beyond the basic link that is established between meaning and form, all the while acknowledging knowledge of word meaning as the central element of word knowledge. In connection to vocabulary, Laufer and Goldstein (2004) defined passive word knowledge as the ability "to perceive the form of the

word and retrieve its meaning or meanings"; and active word knowledge as the ability "to retrieve the appropriate spoken or written word form of the meaning we want to express" (p. 404). Laufer and Goldstein (2004) further defined active recall as "the ability to recall a word's form" (p. 422) and passive recall as the ability to retrieve its meaning, which, they insist may be acquired before one can recall the word form.

Furthermore, following the assumption that learners' passive vocabulary is usually larger than their active vocabulary, Laufer and Goldstein (2004) conclude that active vocabulary represents a more advanced type of knowledge. In addition, their analysis shows that recall of word form or meaning embodies a more advanced degree of knowledge than the recognition of a word form or meaning. They determined that active recall is thus more advanced than, say passive recall and even more so than passive recognition. Laufer and Goldstein (2004, p. 421), thus, conclude that this hierarchy dictates that "the main component of vocabulary knowledge, which is the knowledge of form-meaning relationship, can be construed as a hierarchy of four degrees of strength: active recall [hardest], passive recall, active recognition and passive recognition [easiest]." This study supports the notion of vocabulary acquisition being incremental. It also shows that vocabulary knowledge is incremental, not only in the sense of vocabulary as a whole, but for different degrees of knowledge in relation to the words themselves. Finally, another crucial finding of this study is that passive recall, the capacity of recovering the meanings of words from memory, came up as a determining factor for comprehension and thus for students' course grades. This means that passive recall plays a fundamental role not only for lexical knowledge, but for language performance, particularly at the receptive level, as well.

Schmitt and Meara (1997) insisted that the different types of knowledge of words must be related to one another. They studied the interrelationship between grammatical knowledge (use) and word association knowledge (meaning). While they acknowledge that grammatical knowledge is a broad term encompassing various aspects, they focused on morphological aspects, particularly because of its importance in word family knowledge and hence vocabulary size. Word association was deemed important because of the importance it plays in bringing other words to mind. Schmitt and Meara (1997) assumed a connection between greater knowledge of suffix reflected in larger vocabulary (resulting from access to more word family members) and associations that derive from the connections created by the associated words.

These authors conclude that vocabulary does not necessarily exhibit steady growth but seems to present a "state of flux" instead. Schmitt and Meara (1997) found productive growth of *-ion, -ly* and *-able* suffixes and receptive growth in *-s, -age,* and *-ive* suffixes. The growth of this series of affixes happened independently in productive ability and receptive ability. Regarding associations, the study shows that learners vary in terms of how they acquire *native-like associations*. The study confirms that the participants did not have high levels of associative knowledge in general, not even for the words they claimed to know. Schmitt and Meara (1997) also determined that productive association knowledge is related to vocabulary size and general language proficiency. They also found that receptive, derivative suffix knowledge is also related to vocabulary size. In general, Schmitt and Meara (1997) claim that use of grammatical knowledge (in the form of suffixes) and meaning, as association knowledge, are related to each other as well as to vocabulary size and second language ability. They conclude as well that the limited knowledge of these aspects by the participants in the study is an indication of a dire need to give more attention to word association knowledge (meaning) and grammatical functions (use) knowledge of words. Finally, Schmitt and Meara (1997) point, once more, to the incremental nature of word knowledge and how knowledge of meaning of a word does not necessarily guarantees other types of word knowledge.

Further evidence regarding depth of knowledge is brought forward by Laufer and Nation (2001), who investigated the connection between passive vocabulary size and fluency (reflected in speed of access). They believe that word knowledge must be readily available if fluent conversation or reading as expected. Laufer and Nation (2001) found that non-native speakers' reaction response was slower and more varied than that of native speakers, and it was much slower with less frequent words. This response speed, in addition, consistently varied as the level of word frequency changed, establishing a connection between speed and strength of knowledge. Non-native speakers' fluency was also non-linear as it fluctuated according to frequency levels. The native speakers' response, on the other hand, was faster and was held more constant throughout the different vocabulary levels. In general, subjects with vocabulary sizes that were larger than the frequency band being tested responded more quickly as they came up with word meanings faster than subjects with smaller vocabularies. It follows that to improve fluency at a given level, requires exposure, practice, and acquisition of vocabulary beyond this level. They concluded that the more fluent groups were in the top range of vocabulary sizes; that means that an increase in automaticity lags behind an increase in size.

In this same vein, Schmitt (1998) analyzed the progress in vocabulary knowledge depth at the levels of spelling, associations, grammatical function and meaning, over a year's time, in advanced, adult learners, for a total of 11 words. Schmitt (1998) determined that once productive meaning knowledge is acquired, words are not likely to be forgotten. He also found that other forms of knowledge could be present even in the absence of meaning knowledge.

Spelling, he maintains, is a type of knowledge that participants typically demonstrated, and its absence does not limit participants from other kinds of word knowledge. Regarding association, Schmitt (1998) found that students' knowledge improved progressively, but when fluctuations did take place, they did not follow native-like patterns. Schmitt (2010a) also agrees with the idea of non-linear vocabulary growth, and comments that teachers attest to students' commonly forgetting material.

Also, for Schmitt (1998), knowing at least one of the senses of a word was necessary for participants to provide native-like associations, which were found to be related to grammar knowledge and meaning knowledge, although associations and grammar knowledge are not strongly connected. This occurs, Schmitt says, due perhaps to major advancement of grammar knowledge to the level of features such as word class categorization and derivational forms.

These features were present, however, to a lesser extent when adjectival and adverbial forms were analyzed. Along these lines, Dóczi and Kormos (2016) follow a similar idea as they also argue that word knowledge behaves differently across different types of word classes (i.e., adjectives lag behind nouns and verbs). Schmitt (1998) concluded in his study that there was no evidence for hierarchy of word knowledge types.

A contrastive position is presented by Dóczi and Kormos (2016, p. 87), who claim that, based on the results of their study, they have established "a potential order for the development of word knowledge types." By using a continuum, in Figure 3. we will illustrate the hierarchy of word knowledge types that they proposed. A plus symbol (+) indicates that the knowledge progressed toward the following type of word knowledge. An equals sign (=) indicates that two types of word knowledge may appear together, and that they improved at similar rates during their experiment.

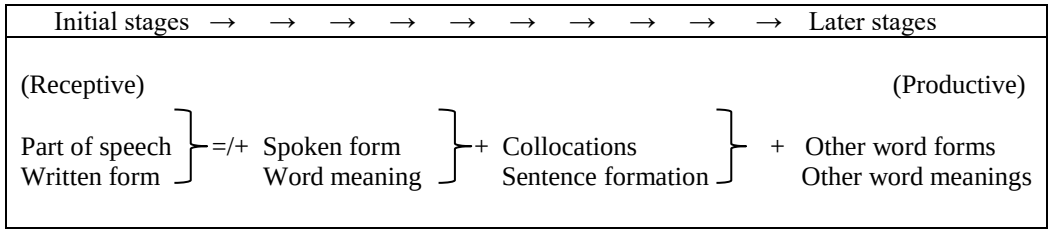


Figure 3. Hierarchy of Word Knowledge Types, based on Dóczi and Kormos (2016)

The types of word knowledge seem to evolve from receptive aspects toward more productive aspects of the language. Part of speech and written form knowledge developed at similar degrees. These are followed by knowledge of spoken form and word meaning. More productive features of vocabulary knowledge; namely, collocational use and sentence formation knowledge follow. The next element is knowledge about *other word forms* and *other word meanings*, for which Dóczi and Kormos (2016) indicate, no evidence of development was shown in their study. Figure 3 should be used for illustrative purposes only, as Dóczi and Kormos (2016, p. 88) point out that "this hierarchical order should not be seen as consisting of entirely distinct and separate stages of development but as phases whose boundaries might be fuzzy and that might potentially feed back to stages lower in the hierarchy." Their discussion is represented here through a figure to illustrate the process they propose.

While the study of words' depth is sometimes seen as separate from that of words' breadth, it is argued that they contribute to each other and are barely separate from one another (Schmitt, 2008). Although knowledge of word depth is not part of this investigation, it should be seen as a contributing factor to vocabulary size, the main focus of the present study.

4.3.2. Explicit (intentional) vs. implicit (incidental) learning

There has been some debate as to whether vocabulary learning should take place incidentally or intentionally. This debate has been mostly sparked by communicative practices that see concentration on form as inappropriate as for them it represents a non-communicative practice. The following findings from different studies attempt to illustrate what different studies have to say about the nature of explicit or implicit vocabulary learning. Pigada and Schmitt (2006)

studied the effect of number of encounters with a word through extensive reading and word knowledge at the levels of spelling, meaning and grammatical functions. They conclude that there is evidence that extensive reading leads to enhancement of spelling, meaning and grammatical behavior, although, at different levels for the different types of word knowledge. The strongest effect is seen in connection to spelling; spelling improved to a higher degree in regard to nouns than to verbs forms.

Schmitt (2007) proposes a cline for spelling knowledge. This cline follows four steps: First, the learner cannot spell the word, then he knows some letters of the word; then he knows the word phonologically, and finally he can produce the correct spelling. For Pigada and Schmitt (2006), meaning also improved although not to the same degree that spelling did. For meaning, again, improvement was noticed to a higher degree for nouns than for verb forms. The same pattern was observed at the grammatical mastery level for nouns, as well. Pigada and Schmitt (2006, p.12) conclude that "substantial learning of the target words occurred during the extensive reading treatment" and that this learning reaches much higher levels than were thought of before. Pigada and Schmitt claim that around 20 encounters seem necessary for all three types of word knowledge to improve. In sum, this study shows that incidental learning through reading is possible.

Schmitt (2007) offers some possibilities to facilitate oral incidental learning through maximizing the learner's exposure to the target language. These include, maximizing use of the target language in the classroom, working in groups where students can learn from one another, and spending time immersed in the target language culture. As for incidental learning through reading, Schmitt (2007) mentions the possibility of *narrow reading*; that is, reading a large number of texts on a single topic. In addition, in light of the extensive use of methods that would implement incidental learning, Schmitt (2008) highlights features of incidental vocabulary learning in this realm. Namely, the effectiveness of extensive reading in incidental vocabulary learning; the fact that incidental learning helps to enhance the knowledge of words, especially those which have been met before; the possibility of acquiring vocabulary via incidental exposure through listening; the possibility of using techniques such as inferring meaning from context, glossing, and other activities to reinforce incidental learning. Schmitt (2008) concludes by calling for support to explicit vocabulary learning through implicit encounters of words in reading and listening while strengthening incidental learning through using (incidental) vocabulary in explicit productive activities that require students to implement this vocabulary.

On a different vein, according to Schmitt (2008), explicit learning has been found to be a very effective approach that appears to be faster, offering greater gains and better chances of retention that allow use of vocabulary at productive levels. This intentional kind of learning can be promoted through different events such as activities that augment lexical engagement with target items, optimization of repeated encounters with these items, and careful selection of the aspects of lexical knowledge to be emphasized in class. However, Schmitt (2008) notes that incidental learning is favored by the popular meaning-based methods, which promote proficiency in skills and cumulative vocabulary gains after repeated exposures. For this view, alleged proficiency, along with the fact that it is simply impossible to explicitly teach all the vocabulary that students need to know, lend more importance to implicit learning.

Schmitt (2008, p. 329) insists, "[v]ocabulary learning programs need to include both the explicit, intentional learning component and a component based around maximizing exposure and incidental learning"; ideally through a strong component of a reading program. He adds that more than simply complementing one another, intentional and incidental learning require each other to benefit learners fully. Schmitt finds support for this in what he perceives as failure in many learners' achieving successful levels of vocabulary learning. Schmitt (2008) maintains that this failure should be interpreted as a sign that more proactive measures are needed for vocabulary learning rather than leaving the process of vocabulary learning to be acquired incidentally (see Schmitt, 2010a). Schmitt (2008, 2010a) further insists that learners, teachers, material designers and researchers ought to work together to achieve the purpose of vocabulary learning.

Schmitt (2008) notes that, at the initial stages of second language learning, explicit approaches directed to strengthening the form-meaning link are most effective whereas enhancement of contextual knowledge can be more fruitful at later stages. He argues for pedagogical involvement (a type of intentional learning) at two levels: teaching form and engaging students with vocabulary learning. For him, while learning meaning is certainly important, form should not be disregarded, as it plays a crucial role in the learning process (see Schmitt, 2010a). For Schmitt (2008), constantly involving students with new words, regardless of the type of engagement, would result in greater likelihood of learning vocabulary. Schmitt (2010a) claims that even the process of being tested on specific vocabulary appears to serve as a facilitator of better retention, as it engages students with the vocabulary being tested, thus resulting in its learning.

Dóczi and Kormos (2016, p. 144) argue that "vocabulary learning seems to be more effective under intentional than incidental conditions, and for incidental learning to take place, a great amount of input and repeated encounters with new words are necessary." Further support is found in Nation (2006b), who insists that incidental and deliberate learning should be seen as mutually complementary. This idea makes it essential that both forms of learning, rather than just one be present in SLA. Nation (2006b) maintains that it is not easy to determine whether one or the other makes vocabulary learning easier:

> The deliberate learning of vocabulary and negotiation of vocabulary involve more focused effort. Incidental learning, in the short term at least, is less effective but also involves much less focused effort. If difficulty is measured by results then deliberate learning is easier. If it is measured by the amount of focused effort, then incidental learning may be easier. (p. 450)

Nation and Meara (2010) show support to explicit teaching and learning of vocabulary. For these authors, explicit teaching provides a model that students learn in class and can follow when they practice on their own outside of class. They claim that deliberate vocabulary teaching encourages deliberate vocabulary learning and can be seen to have three main goals:

> First [it] can aim to result in well-established vocabulary learning […] Second, deliberate vocabulary teaching can have the aim of simply raising learner's consciousness of particular words so that they are noticed when they are met again […] Third, deliberate vocabulary teaching can have the aim of helping learners gain the knowledge of strategies and of systematic features of the language that will be of use in learning a large number of words. (p. 41)

This idea draws attention to the effect that teaching practices can have in students. It is the role of the teacher to demonstrate the essential role that vocabulary plays. Only by having learners experience vocabulary learning and its importance in class can teachers instill in students adequate, autonomous practices for vocabulary learning and retention.

While explicit vocabulary learning seems to receive a lot of support from specialists in the field, it is clear that implicit practices also contribute to vocabulary

learning. As Nation (2006b) and Schmitt (2008) maintain, both forms of learning should continue to coexist in teaching and learning practices as they complement and strengthen each other. Explicit vocabulary learning, a form of learning that has been undermined by many for a long time now, continues to be a very valid type of learning when implemented correctly. Teachers and students should find ways of using implicit and explicit vocabulary techniques that lead learners to reap the best benefits of vocabulary learning.

4.4. VOCABULARY AND TEACHING

Due to the increasingly prominent role that vocabulary teaching and learning has, several authors have presented their ideas regarding how do deal with this issue. Nation and Meara (2010, p. 35) maintain that to determine what vocabulary learners should focus on, one needs to consider "the needs of the learners and the usefulness of the vocabulary items." While this is stated quite simply, many elements should be considered for adequate teaching and learning of vocabulary. The following paragraphs present some ideas brought forward by some of the most influential researchers in the field.

4.4.1. The role of noticing, involvement, engagement and experience

According to Schmidt (2001, 2010) noticing—that is, paying attention to the relevant details and features of the items that learners encounter—is what really leads learners to acquire the linguistic aspects of a second language. In Schmidt's noticing theory, this noticing is more important than rules or other formal aspects of language learning. Elaborating on this idea, Laufer and Hulstijn (2001) propose the construct of task-induced involvement according to which:

> the more attention that is paid to the formal and semantic aspects of words and the richer the associations that are made with existing knowledge (e.g. in the form of establishing similarities and contrasts between old and new information), the higher are the chances that the new information will be retained. (p. 1)

Laufer and Hulstijn (2001) suggest that learners' involvement be established on the basis of *need, search and evaluation* of vocabulary. They argue that these dimensions include cognitive and motivational aspects that contribute to making the process more meaningful for students. Involvement is seen as a key factor for vocabulary learning, retention and recall, and is said to be perceived as driven by a combination of motivational and cognitive factors. In Laufer and Hulstijn's *need, search, and evaluation* notion, need is identified as the motivational factor triggered by the necessity to fulfill task requirements or communication demands. Search and evaluation, on the other hand, are identified as cognitive factors. Search represents the attempt, on the part of the learner, to discover the meaning of an unknown item and to find the term with the meaning he/she wants to convey. Evaluation, on the other hand, describes the process by which learners compare different words, meanings, or word combinations to select the word that they think is most appropriate. Laufer and Hulstijn (2001) describe evaluation as having two different levels. A *moderate* level of evaluation looks at differences between words or differences between various senses of a word. A *strong* evaluation would require making decisions that involve other words in the sentence and how the new item combines with these.

Laufer and Hulstijn (2001) argue that different tasks require more or less involvement. The more elements of the *need, search, and evaluation* construct are involved in the process, the higher involvement there would be, guaranteeing a higher retention of vocabulary. According to these authors, it is the mission of the instructors, then, to develop input or output tasks that require need, search, and evaluation to ensure vocabulary learning and retention. Schmitt (2007) put forward a similar idea in relation to engagement. Schmitt (2007, p. 755) argues that "activities which require more engagement with and manipulation of the information to be learned (*deeper processing*) generally lead to better retention [and] it seems that instructing learners in deeper processing strategies could lead to more efficient learning" (emphasis in original). These ideas reinforce the necessity for greater, more conscious use of vocabulary learning strategies and activities in the SLA process.

In a different vein, Nation (2007b, p. 33) presents *experience tasks* as "the most important kind of language learning tasks because they are essential for fluency development across the four skills […] and they are the most common means of learning from meaning-focused input and meaning-focused output." Experience tasks ensure that most of the ability and knowledge that is required for learners to complete the task comes from the experience that learners have

accumulated. Experience tasks can be implemented in two different ways. In the first, the task is designed to adjust to the current knowledge, ideas, language, and structure that the learners have acquired. For the second, the learners are provided with the knowledge and experience before the actual task takes place, either through pre-teaching activities or through reminding the learners of relevant cues and helping them organize concepts in a meaningful way. Both of these applications ensure that the learner has enough previous experience to deal with the task in the most effective and beneficial manner. Nation (2007b, p. 42) insists that "vocabulary knowledge is clearly a dominating factor in determining whether a task will be an experience task or not. Background knowledge cannot substitute for lack of vocabulary knowledge […]" Designing materials that offer fitting experience tasks, says Nation, is key to develop language proficiency at the receptive and productive levels.

4.4.2. Key issues in teaching and learning

Schmitt (2010a) identifies six issues that are crucial to the teaching and learning of vocabulary, and which, he insists, summarize research findings. The first has to do with the large number of words required to perform different functions in a language. Research has provided different figures depending on the task the vocabulary is needed for. More detail was provided on this in the section on breadth of vocabulary (Section 4.2). As described above, a minimum of 2000 to 3000 words is necessary for basic activities. The second issue has to do with the idea that there are different types of word knowledge that are required to use a word correctly. Knowledge of a word form and its corresponding meaning does not entail full command of that word. Third, vocabulary knowledge is incremental, both at the level of general vocabulary and at the level of individual words (see Schmitt, 2010b).

Fourth, vocabulary learning requires consolidation. That is, learners need to meet a word several times, at regular intervals to remember the word. He insists that while recycling vocabulary is important, consolidation must take place to prevent forgetting. Multiple encounters with a word result in learning more about that word, this then results in better consolidation of the concept. Fifth, the learning of vocabulary requires enhancing previous, partial knowledge. This means that repeated exposures to a word are needed to develop contextual information of a word so that the knowledge of that word does not stay at the form-meaning level

but evolves to other aspects of knowledge, as well. Finally, Schmitt (2010a) points to the teaching of vocabulary as a way of providing students with the number of words they need and the number of encounters required for learning. The teaching method could vary depending on the situation and it could implement several approaches, either through intentional or incidental learning.

File and Adams (2010) studied the effects of two types of instruction: what they call *integrated vocabulary learning*, where the focus on form occurs in communicative contexts (planned or incidental), and an *isolated approach to vocabulary instruction*, seen as a focus on the features of vocabulary. They studied these approaches in connection to reading. They also included a number of words to be tested, some of which could only be learned through incidental encounters in reading. They depart from the premise that incidental vocabulary learning depends on extensive reading and that it requires multiple encounters with the same item to result in vocabulary learning. They found that any kind of instruction either isolated or integrated resulted in better retention and learning of vocabulary, although this difference is not statistically significant in relation to incidental learning alone. They attribute better retention of vocabulary to the instruction provided, and report on the students' *noticing* the words better when reading, as they had been brought to their attention before. As the words were presented in isolation, students could focus on them and hence better processing was possible. In addition, instruction could have been perceived as clearer when words were presented in isolated instructional treatment. Isolated instruction offered another encounter with the word before it was met in the reading context. For the authors, this may have caused focusing on the form once it was encountered in context, following a subsequent recall of earlier instruction. Based on this idea, File and Adams (2010) point to the importance of isolated instruction of vocabulary before reading as a potential solution to prevent reading difficulties.

Schmitt, Jiang & Grabe (2011) question the existence of a threshold that would determine a marked difference in terms of understanding a text. Instead, they argue for a linear relationship between vocabulary coverage and reading comprehension, according to which as vocabulary increases, reading comprehension also increases. Put differently, since comprehension improves with an increased vocabulary, a higher degree or demand of comprehension would require greater, larger vocabulary knowledge. Schmitt, Jiang & Grabe (2011) insist that comprehension is strengthened or hindered by vocabulary knowledge of the text; however, they insist that while vocabulary is certainly a crucial element in reading comprehension, it is not the only one. They note that vocabulary

undoubtedly interacts with other factors in the reading comprehension process. While learners with perfect vocabulary knowledge did not understand the text completely, others with less vocabulary knowledge did achieve extensive comprehension of the texts, that is, there are clearly other aspects at play.

Macedonia (2015), in an analysis of learning styles and vocabulary acquisition, disproves the theory of learning styles on the basis of lack of evidence and relying on the fact that the brain processes and stores incoming information in regions that have that function and for which more than one sensory function is necessary. She does not find it feasible for the brain to rely on only *acoustic* or *visual* clues, for example, to carry out the task of vocabulary learning, or any kind of learning for that matter. On the contrary, Macedonia insists that the brain implements multiple sensorimotor activities to carry out learning, instead of assigning the processing of information to one region alone. She maintains that targeting a style-specific type of learning hinders learning rather than promoting it. For Macedonia (2015), because large network connections (for example, cognitive, semantic, memory, and multisensory processing) are used to store and process input, students should be provided with multiple sensorimotor experiences when learning vocabulary so that they can learn vocabulary in the most efficient way possible. A further discussion on brain and its vocabulary networks is presented in section 4.5.2.

4.4.3. Spacing techniques

In a different vein, Schuetze (2015) experimented with spacing techniques for second language vocabulary acquisition. He used two experiments testing the same number of words. In one, participants used a *one plus three* repetition rate; and in another one, subjects experienced a *one plus four* repetition pattern. Some students were subject to uniform spacing for which the break within repetitions was regular; others were part of an expanded schedule for which spacing intervals increased. Schuetze (2015) concludes that a higher number of repetitions led to greater recall rates in both the one plus three and one plus four groups. Students using expanded intervals were more successful in short-term gains, while in long term retention uniform intervals produced higher scores. The uniform group showed more consistency in their scores; although grades were low at the beginning, subjects proved to recall more words in the long run. The expanded group, whose grades were higher in the first test, produced lower scores on

subsequent tests. The results show higher rates of recall for content than for function words and greater difficulty in recalling more complex phonological items. This study highlights the importance of spacing techniques and contributes to the idea that multiple encounters are key for vocabulary learning.

Nakata and Webb (2016) implemented a study on *part* learning (division of material into smaller chunks) and *whole* learning (presentation of material in a single, large block) in vocabulary acquisition and its connection with spacing techniques. As in the previous study, spacing refers to the time intervals that are created between different encounters with the material, i.e., vocabulary in this case. The study measured vocabulary learning at productive and receptive levels. Nakata and Webb (2016) found that part learning produced more correct retrieval of words at the beginning although this difference was very little in post-test results. These authors warn us of the possibility of part learning creating a false illusion of increased learning due to good results in performance at beginning stages. If students feel overly motivated at that point, they may stop studying and may interrupt the learning process. Furthermore, Nakata and Webb (2016) found that as long as spacing was the same for both study groups the part/whole dichotomy had little effect on learning, so that lends a great deal more of importance to spacing. Nakata and Webb (2016) emphasize the need for a greater awareness of the value of this technique. They found that the treatment groups that had longer spacing had a notoriously better result in delayed posttests. Nakata and Webb (2016, p. 547) maintain: "although spacing has a large effect on vocabulary learning, its value has not been exploited fully in typical instructional settings […]."

4.4.4. The Four Strands of Language Learning

Relevant information about teaching and learning techniques can be added from Nation's *Four Strands of Language Learning* for language courses, introduced by Nation (1996, 2003, 2007a), Nation and Chung (2009), and Nation and Yamamoto (2012). It refers to a way to create a well-balanced language classroom where the vocabulary component would have a prominent role across the different activities and different skills. The main idea is that all four strands should roughly be distributed equally over a period of time. For Nation (1996, p. 7), the strands aim at students' achievement of "fluent control of the sounds, spelling, vocabulary, grammar and discourse features of the language, so that they can be used to communicate effectively." For Nation (2003), establishing and

implementing the four strands guarantees both deliberate and incidental learning in the classroom. They apply, he insists, to all the aspects of language learning in a course and target a diversity of forms of learning. Nation (2007a, p. 1) notes that the term *strand* illustrates the nature of "long, continuous sets of learning conditions that run through the whole language course." The four language learning strands are discussed in more detail below. These include: meaning-focused input, language-focus learning, meaning-focused output, and fluency development.

Meaning-Focused Input: According to this strand, learning takes place through reading and listening activities where the learner concentrates on the message that is being received. Nation (2003) points to the importance of the presence of this strand since beginning stages of the learning process (see Nation and Meara, 2010). Nation (2007a) describes a number of requirements for meaning-focused input to take place. Among these, 1) the information that is read or heard is familiar to the learner; 2) the students are interested in the information; 3) only a small percentage of the information is unknown to the learners (5 to 2 percent); 4) context and background knowledge facilitate the understanding of the unknown forms; 5) and there has to be a lot of exposure and access to input. Nation and Meara (2010) suggest a strong reading program as a good base to implement this strand. They point to the cumulative feature of incidental learning and call for enough exposure through multiple encounters to ensure strong knowledge of words. Nation (1996) suggests listening to stories, participating in conversations, following instructions and watching television, as activities to promote this strand of learning. Nation (2007a) adds activities such as extensive reading, shared reading and being a listener in a conversation to strengthen this strand.

Language-Focused Learning: This strand requires the learner's attention to be focused on language features that take place while direct teaching of pronunciation, spelling, vocabulary, grammar, and discourse related aspects are emphasized. Nation (1996) argues that there is sufficient evidence that points to the contribution of language-focused instruction in SLA. As occurs with the other strands, Nation (2007a) suggests five conditions to facilitate learning through the focus on form strand: 1) Learners should deliberately attend to language features; 2) language features should be processed in deep, thoughtful ways; 3) spaced, repeated attention to the same features should be enforced; 4) the features that learners are concentrating on should be simple and should not rely on developmental knowledge that students have not acquired yet; and 5) the features under study should also be encountered in the other three strands. As for activities, Nation (1996) points to overt explanation of word meaning, pattern use, and exercises

based on rule application activities to foster this strand of learning. Nation (2003) notes that the most obvious form of deliberate vocabulary learning is one involving memorization of vocabulary through translation of items into the students' L1. Nation (2006c) claims that translation is one of the simplest and clearest ways to provide meaning for a word, considering that vocabulary meaning is facilitated when it is conveyed in short, simple ways, translations should definitely be an option to consider.

Nation (2007a) insists that traditional activities, whose benefits have been undermined on behalf of so-called more communicative practices, should be used to promote this learning strand. Emphasis on pronunciation and grammar drills, substitution tables, vocabulary cards, intensive reading and translation, memorization, feedback on writing, guessing from context and dictionary use are all highly beneficial if done correctly, says Nation. Nation and Meara (2010) note that deliberate vocabulary learning is more effective because it is more focused and goal-oriented and as such represents an excellent option for vocabulary learning that results in quick and more permanent vocabulary growth. Nation (2011) notes that opposition to rote-like learning strategies represents one of the various research findings that have not been well applied or well transferred from the field of applied linguistics into the teaching of languages. Followers of communicative approaches do not support rote, decontextualized learning practices. As a result, these practices have fallen into disuse in most second language teaching classrooms and thus students do not enjoy their multiple benefits. Nation (2011) insists that if done correctly, rote learning can certainly turn into effective learning.

Meaning-Focused Output: In this strand, attention is concentrated on learning that takes place when learners convey their ideas to others through production of writing and speaking. Nation (1996) acknowledges the need to *push* students so that production knowledge develops. This knowledge develops when they are faced with the need to produce language in areas unfamiliar to them, or with language forms in ways that they have never used before. Nation (2007a) argues that the same type of conditions applied for meaning-focused input can be applied to the meaning-focused output strand; namely, 1) familiarity with topics used for writing or speaking; 2) an emphasis on the message communicated; 3) that only a small proportion of the items needed for speaking or writing are unknown; 4) that communication strategies, dictionaries, or previous input can aid communication; 5) and that students have multiple opportunities to read and write.

Nation (1996) lists some possible activities to develop this strand. These include role playing, writing activities and split information tasks. Nation (2007a)

adds talking in conversations, giving a speech or lecture, writing a letter or note, keeping a diary, telling a story or instructing someone on how to do something. In addition, Nation and Meara (2010) include annotated pictures or definitions and group work to this list.

Fluency Development: This strand deals with the idea of students' using what they know by actively practicing listening, reading, speaking, and writing activities. For Nation (1996), fluency is only demonstrated when, given plentiful opportunities, the speed and ease in the use of language appropriately and effortlessly fits into the flow of a conversation. During this time, learners concentrate on putting into practice what they already know, while taking a break from learning new items. Nation (2003) argues for development of fluency, independently, in each one of the language skills. Nation (1996) claims that certain conditions need to be met for fluency to develop. First, items required for the activities are within learners' previous knowledge. Second, the activity focuses on communicating a message. Third, in Nation (2007a) the learners are supported (or pressured) and encouraged to perform at higher levels or at a faster rate. Nation (2007a) adds that a great load of input and output is required for fluency to develop.

Some possible activities recommended by Nation (1996) include 4/3/2 technique (students take turns retelling a story in 4/3/2 minutes) and other activities where they are given extra time to prepare for production. In addition, Nation (2007a) adds speed reading, skimming and scanning activities, repeated reading/retelling, ten-minute writing, and listening to easy stories. Furthermore, Nation and Chung (2009) include, pyramid-ranking activities and writing about familiar topics. Nation and Meara (2010) maintain that there are two common approaches to the development of fluency in vocabulary learning. The first, *the well-beaten path,* relies primarily on repetition of forms to ensure learning. The second, *the richness approach*, relies more on the establishment of connections and associations of new items with already existing forms to ensure learning and retention of vocabulary.

The four strands of language is becoming an influential theory in SLA and in vocabulary practices in particular. It represents a well-structured goal for language planning and material designers as well.

4.4.5. The instructor's role

Regarding instructors, Nation (1993) identifies important roles for teachers to ensure first and second language vocabulary development:

> The first and most important role is the informed direction of teaching effort. The other three roles are particular applications of this, namely direct teaching of high frequency vocabulary, training in vocabulary learning strategies, and the encouragement of activities that lead to indirect vocabulary learning. (p. 127)

Teachers have to ensure multiple encounters with the items they want learners to acquire, as a single encounter does not suffice for such purposes and may simply serve as a portion of learning of that concept that will add to the next encounter with the word (Nation, 1993). Nation (2003) also sees the need for teachers to identify *core meanings*, that is, the meaning that different senses of a word have in common. For him, core meanings are useful because they reduce the number of items that students have to deal with when learning one particular item, and because they allow students to access different uses and contrasts that may not be familiar in the L1.

Nation (2003) offers a list of four principles for teaching vocabulary that can be used by teachers in the classroom or can be applied by autonomous students. Learners should focus on learning the most useful vocabulary first with the purpose of obtaining the best return of their learning. They should start with high frequency words (first, second and third 1000 word families). Then, they should analyze their goals and continue with academic vocabulary, specialized vocabulary (while studying a specific major) or low frequency words. Second, learners should focus on vocabulary learning appropriately, starting by using strategies such as word parts, guessing from context, word cards and dictionary use. In this same vein, Nation and Meara, (2010) consider these to be the four major strategies for learning and retention of words. The learner has to consider that deliberate learning and incidental learning complement each other and, thus, both should be practiced.

Moreover, in the third principle, attention should be given to high frequency words across all four strands of learning both in receiving and communicating ideas that implement a good fluency level. Fourth, learners should be encouraged to reflect and take responsibility for learning. For Nation (2003), this presupposes

knowing what to learn and knowing the different possibilities to learn it, having the skill to choose the most appropriate options, and having the ability to monitor and evaluate vocabulary growth through these options. Finally, Nation (2006b) divides what is presented above as the fourth feature to emphasize, even more, the value of students' responsibility. The new fifth feature states, thus, that learners should be guided to take responsibility for their own learning. Teachers can guide students, either through explanations or strategy training, to reach the goals of vocabulary growth.

Further rationale regarding appropriate teaching and learning practices regarding vocabulary learning is provided by Dóczi and Kormos (2016). Dóczi and Kormos (2016) offer what they call the "ten commandments for the 21st century vocabulary teaching." They can be used as a guide to help teachers find the right balance for vocabulary inclusion in the classroom or language program. These guidelines are summarized below:

1. Teachers should identify the features that make vocabulary easy or difficult so that they can provide the aid that students need when dealing with these forms.

2. Context and selection of appropriate vocabulary are crucial to ensure vocabulary growth; a clear aim at teaching the first 3000 word families is a beginning step.

3. Teaching and revision of types of vocabulary knowledge such as collocation and word formation knowledge should be taught along word form-meaning information.

4. Teaching-material writers and teachers need to directly address the issue of lexical fluency.

5. Incidental vocabulary learning should be strengthened through extensive and narrow reading as well as varied input.

6. Teachers should aim at providing frequent, focused meetings and enriched contexts for vocabulary to ensure explicit vocabulary learning; student engagement is key.

7. Supporting students' autonomous learning is recommended.

8. Study of lexical phrases is important as it contributes to pragmatic knowledge, fluency and production. Exposure to and practice with these forms is necessary.

9. Teaching of lexico-grammatical patterns is recommended as they foster comprehension, accuracy, fluency, flexibility and depth of knowledge.

10. Use of technology (i.e., mobile platforms) can promote incidental and intentional vocabulary learning through active, rich engagement that is motivating for students.

Taking into account what other authors have said about vocabulary teaching and learning, to those "ten commandments" we could consider adding a few more elements, such as the following.

11. Relying upon the students' L1 for support could be very beneficial, for instance. As Nation (2006c) suggests, use of translation, for example, is a simple, clear way of providing meaning in a straightforward manner, which in turn facilitates vocabulary learning. Use of L1 is, let's not forget, a practice well supported by theory (Cummins, 1979, 2000, 2007, 2009, 2014a, 2014b) that establishes the interdependence that exists between the knowledge of the speakers' different languages. L1, being a readily available resource, should be used, though with care, to its fullest potential.

12. Use of spacing techniques has proven to exert numerous benefits in the process of second language acquisition (see Castro García, 2014 for reference). Schuetze (2015) and Nakata and Webb (2016) highlight the importance of using spacing techniques to consolidate vocabulary knowledge. As the latter mention, despite the great value of spacing techniques, these are not used to the fullest in the classroom setting.

13. Application of Nation's (1996, 2003, 2007a) four strands of learning is necessary as it intends to create a well-balanced presence of the vocabulary component along the different elements of the curriculum in the classroom.

14. In general, raising awareness of the important role of vocabulary, between instructors and learners, is necessary so that notions such as noticing, involvement and engagement with vocabulary play an active part in the SLA process. These, by themselves, would have tremendous impact on vocabulary learning and efficient second language acquisition.

4.5. MENTAL LEXICON AND SEMANTIC NETWORKS

Vocabulary, being a quintessential feature of language, has also received scientific attention from areas more oriented toward neurolinguistics through studies that examine features relating to the mental lexicon and the semantic networks that, to a great extent, determine its functioning. Jean Aitchison, in her book *Words in the Mind: An Introduction to the Mental Lexicon* (2012), takes us through the analysis of several key aspects of the mental lexicon, which, although mostly departing from a first language perspective, can easily be extended to second languages as well, she says. In her concluding chapters, Aitchison (2012) acknowledges the undeniable growth of studies and the evident need to further knowledge in a field that still has a long road ahead in answering questions about the functioning and behavior of the mental lexicon and of the semantic fields and networks that make it up. Given the nature of these studies, and the challenges faced just by working with human brains alone, we cannot expect absolute, conclusive findings. Most authors in the field (Aitchison, 2012; Baddeley, 2015; Sailor, 2013) mention the need for further studies. In what follows, the brief account describing the mental lexicon and its semantic networks should be seen as mere reference to these areas of studies, justified by their key role in vocabulary formation, and should not be taken, by any means, as an exhaustive review of the topic. We will begin our exploration by referring to how we first process information that may later result in strengthening our mental lexicon.

4.5.1. Information in working memory and long-term memory

We will first take a quick look at how information, and thus vocabulary, is incorporated into our brain. According to Baddeley (2015), information comes to our brains through the phonological and visuospatial subsystems (see a reproduction of his model in Figure 4). Both of these subsystems contribute information from a large number of sources. The visuospatial subsystem collects information from the visual (color and shape) and spatial fields, as well as from the rich fields of touch and kinesthesis. For Baddleley (2015), the phonological loop, which facilitates the acquisition of new words, brings information from speech and other language related sources such as lip reading, signing and music as well as other non-verbal cues. He notes that this subsystem has a vocal and sub-vocal

articulation function that allows for effective rehearsal. In this rehearsal, focusing attention on a given item within the episodic buffer results in maintenance of that item. The episodic buffer, in turn, is a passive store that holds bound elements coming from language and visual information. Information travels back and forth between the buffer and the central executive system. The central executive controls the episodic buffer and is in charge of attentional processing.

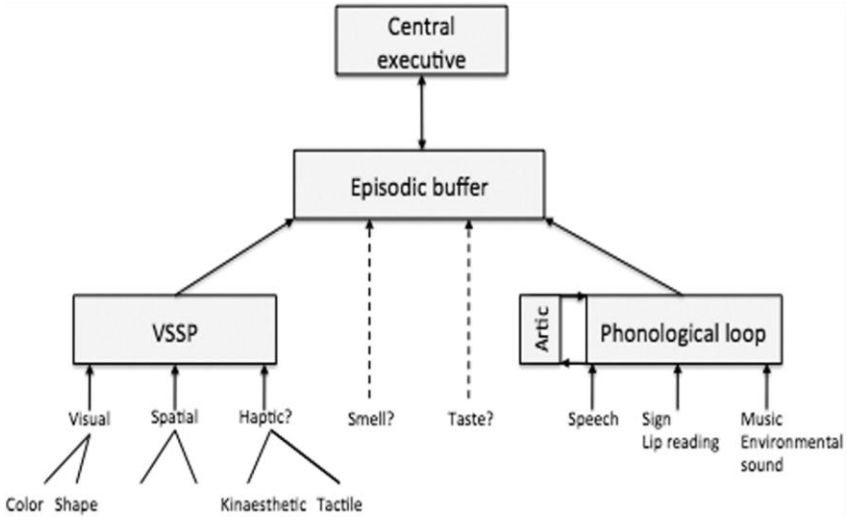


Figure 4. A speculative view of the flow of information from perception to Working Memory; taken from Baddeley (2015, p. 22).

All these elements take information into what is known as *working memory*, but information does not simply stay there. For Baddeley (2015), working memory serves as an interface carrying information to and from cognition (long-term memory) and action. Cognition hosts information from many different sources, long-term memory, semantics and perception as well as that information that comes from working memory, experience, vision and so on. Baddeley (2015) notes that the permanent activity taking place between "cognition, working memory and action" substantiates our interaction with the world. This interaction occurs as we receive information, respond, and act upon it. This is how Baddeley (2015) perceives the intricate interaction between working memory and long-term memory. As we receive, act upon, and store information, particularly in the form of words (our main interest here), the mental lexicon and the semantic networks

conforming it become stronger. We move now to review how information is stored and interacts in our mental lexicon.

4.5.2. Semantic networks in the mental lexicon

Wojtinnek and Pulman (2011) see semantic networks as structured representations of the concepts that conform the mental lexicon. Along these lines, for Aitchison (2012), they represent a system of interconnected elements. Aitchison (2012) further argues that elements conforming a semantic network are organized into semantic fields. She argues that the existence of semantic fields is supported by evidence that comes from experiments with word associations (clearly pointing at the strong association of coordinates), from slips of the tongue, and word search tasks. In all of these cases, subjects choose words that are located in the same semantic area and which are expressed either through a coordinating element, an element of the same topic, an opposite word to the triggering element, or an element of the same word class. It follows that there are features of shared similarities that help establish connections among network elements. For Aitchison (2012), "firm connections, such as those between coordinates and those between common hyponyms and superordinates, are used in conjunction with our reasoning ability to make other temporary links as we need them" (p. 112). This means that different aspects contribute to the set-up and shape of the structure of the network.

Aitchison (2012) advances further rationale as to how the mental lexicon is organized:

> The mental lexicon, therefore, seems to be a mixed system which has found a workable compromise between the requirements of production and those of comprehension. The component that is required first in each case has imposed its demands on the organization. Production begins with the semantics and syntax, so these are arranged to suit production. Recognition begins with sounds, so these are organized to suit recognition. (p. 263)

Aitchison (2012) explains this organization along the following lines. For the organization of production (through lemmas), she maintains that children acquire the meaning of words following the tasks of *labeling, packaging* and *network building*. *Labeling* involves the ability of assigning a symbol to each given concept;

it emerges progressively between ages 1 and 2. *Packaging* involves classification of a number of objects under specific labels, thus, allowing for the presence of clusters. Finally, *network building* occurs increasingly and continues throughout life. While words become part of a network, links are established for coordinates, seemingly as a result of syntax development and possibly to allow for later, rapid word tracking as vocabulary grows. Aitchison (2012) maintains that children and adults follow the same principles.

Hosoda, Tanaka, Nariai, Honda and Hanakawa (2013), on the other hand, argue that L2 vocabulary learning triggers reorganization of structures outside cerebral areas that are typically associated to language networks. In this vein, Hosoda et al. (2013) argue that L2 vocabulary ability in L2 learners presents greater activity in the right hemisphere, uncommon in monolinguals. Hosoda et al. (2013) associate this activity in the right hemisphere with vocabulary and phonological aspects of L2 processing activities in their subject. They state that specific functions of these activities can only be speculative.

With the purpose of assisting speech planning, Aitchison (2012) sees the grouping of words into word classes and coordinating groups as a way to aid memory in dealing with large numbers of words. As for the recognition organization (through sounds), she claims that we group sounds together in *natural classes* that either behave similarly, share phonetic features or belong to words that have similar beginnings, endings or rhythms and which although not necessarily spatially close, maintain strong ties. She insists that words are originally stored as auditory images. Aitchison further compares the role of this mental grouping with that played by the alphabetical system of the printed lexicon. In sum, the mental lexicon is seen by Aitchison (2012) as a very active, flexible system where needs are met for proper production, perception and memory functioning.

Aitchison (2012) insists that the mental lexicon is mainly concerned with links and their interactions. She argues that key features can be assigned to the mental lexicon, and she has summarized them as follows: First, she maintains that the lexicon is made up of full words. Second, these words are organized based on their features for meaning and word class and their features for sounds. Third, the meaning and class of the word aids production while the sound features of the words aid perception. Fourth, the mental lexicon constitutes a quite complex network, which although containing fairly durable links, is still a flexible and malleable unit where new links can be created, and old links can be modified. Fifth, multiple activation of links takes place while the network is at work; some of the activated links will not be needed once the correct concepts are identified. Sixth,

much interaction takes place between the different nodes that head the links conforming the network. Seventh, *excitation and inhibition* of the different activated links takes place during the process of network use until the suitable element is found. Several words are checked, evaluated and dismissed during the selection process. Eighth, the selection process begins with a broad search for a concept. After activation of large numbers of elements, a narrowing down process filters the words until the right word is identified. Ninth, Aitchison (2012) talks of a *lexical tool-kit* that has the function of creating words that become part of the lexicon. Tenth, Aitchison (2012), finally argues that all words in the mental lexicon are constantly being re-analyzed and their links to other words are constantly re-assessed.

As seen in Aitchinson (2012), the functioning of the mental lexicon is complex, involving different elements with detailed functions and purposes. In the following paragraphs, we will provide an overview of certain results of studies that analyze different aspects concerning the mental lexicon and its networks.

4.5.3. Other scientific evidence

Further rationale on how network growth takes place comes from Sailor (2013), who describes this growth through the perspective of Barabási and Albert's (1999) "preferential attachment principle" (cited in Sailor, 2013). Sailor (2013) states:

> The preferential attachment account claims that new words are added to a network by establishing connections with an existing node and some of the nodes to which it is connected. The probability of such a connection to an existing node is an increasing function of how many connections it already has to other nodes in the network. (p. 1657)

It follows that association of a new word, a new learned element to a known word that contains many associations to other known words would result in this new word having a much better chance of establishing a strong network that warrants its growth and value. Sailor notes, however, that the number of connections created by a new item tends to decrease as age increases. In his study, Sailor found that early-acquired elements tend to exhibit a much larger number of

connections than later acquired words. The importance of this information lies in the idea that words with greater connections allow for better and stronger connections of newly incorporated items.

In a different vein, Henderson, Weighall and Gaskell (2013) discuss offline consolidation processes (where sleep and time can play a key role) for the establishment of new lexical items and the integration of these forms in existing lexical knowledge. New information is initially processed separately from existing knowledge and subsequently integrated into long-term memory. They note that new phonological information and semantic information are crucial for vocabulary acquisition and its stable lexical representation. This notion, we can see, supports Aitchison's (2012) idea of the significance of the role of phonological and semantic information in lexical strengthening. Takashima, Bakker, van Hell, Janzen and McQueen (2014) also describe the importance of consolidation periods and insist that integration of the new items into the mental lexicon relies heavily on the information available during encoding of new words.

Bakker, Takashima, van Hell, Janzen and MacQueen (2014) elaborate on the idea of consolidation periods. They note that integration of novel words into the existing lexicon is eased by offline periods of consolidation and by sleep in particular. Bakker et al. (2014) base their research on experiments that analyze the role of memory formation across written and oral modality learning and put forward the following findings. First, they note that although learners can recognize novel spoken words short after being learned, they still need a period of consolidation (of at least 24 hours for these authors) for these to become part of the mental lexicon. Second, in an experiment that compared concepts learned via oral versus written input, they found that new concepts first encountered in written form seem to require longer consolidation periods and more training than phonologically-met, new concepts, possibly indicating that printed input allows for weaker memory representations. Third, in an additional experiment they found, however, that 24 hours were enough for written-learned words to compete in response times in tasks that involved other words that were also learned through written input. Again, in this case, the process was still more efficient for phonologically learned elements when compared. This indicates, the authors state, that a contrast in the lexicalization mode (oral vs. written) appears to require longer periods of time for formation of cross-modal representations to take place if the word is encountered via written form, but not if it is encountered via spoken input. Bakker et al. (2014) mention that this may be taken to indicate that print input may not be optimal to allow strong formation of lexical representations.

In yet another study, Vitevitch and Goldstein (2014) attest to the existence of a set of keywords that occupy fundamental positions in the lexical network. The privileged positions of these key elements warrant them speedier access and greater accuracy during events that involve spoken recognition. These keywords, the authors say, hold the network together, avoiding breakage into smaller parts, and optimize the distance between nodes inside the network. This organization, which is based on phonological word forms, can in fact influence language processing, the authors point out. They argue that strengthening the representation of these keywords may derive from activation of neighboring elements, as well as the direct activation of the key elements themselves. Vitevitch and Goldstein (2014) add that findings on a network concept that is key in one domain can be used to explore dimensions of another domain that may appear distant at first. We can conclude that although these authors do not identify these so-called key concepts, the existence of such keywords support the need for stronger well-varied mental lexicon in second language learners. A much larger mental lexicon would inevitably result in a much greater chance of incorporating novel elements that add strength and value to the network.

Borovsky, Ellis, Evans and Elman (2016) present evidence regarding the contribution of the semantic networks surrounding a known word to language processing skills. They argue that the conceptual structures expressed through semantic networks develop as vocabulary develops and that these networks are organized in children very much the same way they are in adults. Borovsky et al. (2016) maintain that a key connection exists among vocabulary growth and the child's linguistic processing skills. They insist that this early established relationship between vocabulary and language processing proves to be crucial for language skills in the long run. These authors argue that this connection is evident in the fact that children who interpret speech faster are quicker at recognizing and acquiring new words. Their larger vocabulary enhances their ability to recognize lexical and sentential information faster, as well. This, Borovsky et al. (2016) insist, is a result of greater concurring activation of items with strong connections.

At the same time, the interference of items with weaker networks is limited in this type of network. They further maintain that linguistic processing skills vary not only in connection with the general knowledge that the child has, but also in regard to the knowledge and experience that the child might have with specific items in different domains. Borovsky et al. (2016) find that their study contributes evidence to the *preferential attachment* theory that indicates that children learn new vocabulary by adding elements to already existing words in a way of facilitating

the process by relying on previous knowledge, as mentioned above by Sailor (2013). They take this as an indicator that "organizing vocabulary training to items within semantic domains may yield benefits for processing skills and vocabulary growth" (p. 1906). Borovsky et al. (2016) conclude by pointing to the need to identify key vocabulary items that contribute to vocabulary growth, and to characterize the interaction among these elements for a better understanding of language acquisition.

Ellis, Burani, Izura, Bromiley and Venneri (2006) describe the importance of the role of the age of acquisition in connection with early- and late-acquired concepts. Ellis et al. (2006) argue that age of acquisition works as a predictor for early-acquired items. Aware of how controversial the age issue is in SLA, we will concentrate on the fact that early acquisition of elements is essential, without directing much attention to age itself. According to these authors, early-acquired concepts exhibit a more rapid recognition, greater likelihood of retention in the event of brain damage, much sounder connections to other items and a greater network configuration when compared with later learned concepts. If this information is extended to the role of concepts acquired early on in SLA, the function of the first 2000 words, as described by Nation (2013) and which are contained in the high frequency vocabulary band gains paramount importance in the service that they give to learners of second languages.

This brief reference points to the central role of semantic networks in the strengthening and growth of the mental lexicon. Key issues have been brought to our attention. The association of novel elements with already existing elements is deemed crucial to increase and strengthen the network, as pointed out by Sailor (2013) and Borovsky et al. (2016). Vitevitch and Goldstein (2014) argue for the existence of keywords by directing attention to the need to identify elements that may serve as a strong basis for a rich and robust network. Bakker et al. (2014) mention an issue that may directly affect methodology in second language classrooms, especially for those contexts where passive, reading practices overrule oral practices, as in the Costa Rican context. Phonological concepts have proven to be more effective to consolidate concepts in the mental lexicon. This, along with the idea of consolidation periods presented by Henderson et al. (2013) should be seriously considered in second language contexts, for new concepts to find a secure place in the learners' mental lexicons.

In sum, the support of semantic networks in language processing in general and second language processes in particular is undeniable. We should find ways for concepts to move between working memory and long-term memory in a way that

ensures a steady presence, with multiple connections in the mental lexicon. While the field of neurolinguistics continues its journey toward more definite results, SLA should follow those findings closely, to implement what theory seems to direct us toward. This would assure both neurolinguists and SLA practitioners a steady road to a better understanding of how our brain deals with learning a new vocabulary and eventually a language.

4.6. CONCLUDING REMARKS

The critical role of vocabulary knowledge in the second language acquisition process is undeniable. The influential role of L2 learners' vocabulary size is essential in the second language learning process. Given the key role of vocabulary in relation to the different language skills, vocabulary growth is often identified as one of the main goals of teaching and learning. Research suggests that the number of words a second language learner knows is a strong gauge to determine learners' proficiency (Nation, 1993; Hu and Nation, 2000; Laufer and Goldstein 2004; Stæhr, 2008). Despite the growth of vocabulary research, vocabulary-teaching practices seem to continue concentrating on the form-meaning link (Dóczi and Kormos, 2016). A combination of vocabulary techniques should be implemented to ensure appropriate vocabulary acquisition. Reading, writing, listening and speaking should all be used to foster vocabulary growth and all skills will benefit in return (Nation, 1996, 2007a; Nation and Chung, 2009). The importance assigned to vocabulary calls for additional direct and active ways of dealing with vocabulary in the classrooms as evidence shows that vocabulary learning should not be left to itself, but rather be given prominent attention in class through more explicit and implicit vocabulary learning practices. With the intention of ensuring presence of vocabulary in classroom activities, awareness must be raised in coursebook and material designers, teachers, and students toward the critical role that vocabulary plays in reading, writing, listening and speaking the second language. We must find a way to develop vocabulary knowledge that is reflected in students' fluency and ease in communicating. These are quintessential features in second language proficiency and language use.

CHAPTER 5

GENDER AND MOTIVATION IN SECOND LANGUAGE ACQUSITION

This chapter deals with two specific aspects of second language acquisition: gender and motivation. Although not the main focus of the present study, both aspects are often said to play a key role in any language learning process. Given the secondary roles assigned to these two topics, they will be explored here only to a limited extent and they will be analyzed as variables that are likely to intervene in the process of SLA and vocabulary development in particular. Gender is an area of second language acquisition that has received an increased level of attention in more recent decades. This area is currently undergoing steady development, as the focus of research in different institutions that are exploring different aspects of gender grows. In the case of SLA, gender studies have received considerable attention although results demonstrate that we are still in need of further investigation that may lead to more conclusive results.

On the other hand, motivation (along with other affective factors) is also considered a determinant factor in second and foreign language attainment. The present section should be taken as a concise introduction to the topic and should serve to provide a basis for possible explanations regarding the findings of the present study. Traditionally, motivation has credited the individual, social, and educational factors as key factors in second language acquisition. Particular attention will be given to Gardner's (1985, 2007) integrative motive and to other more recent developments in the area (Dörnyei, 2005, 2009a, 2017), with the intention of later identifying possible related characteristics of the population under analysis here.

5.1. GENDER AND SECOND LANGUAGE ACQUISITION

The issue of gender in connection with Second Language Acquisition has long been addressed as a determinant factor in second language proficiency and general second language attainment. In 2015, van der Slik, van Hout and Schepens published a detailed literature discussion regarding gender effects on language acquisition. They analyze factors ranging from memory access, nature/nurture issues, hormonal configuration, individual and societal characteristics to other features such as language skill performance or strategy use, all of which have been associated at some point with differences of gender achievements in the area of second language acquisition. Diverse studies mention results related to the variable *gender* as part of the account of their findings. It would appear that this practice is somewhat more evident now than it was decades ago. Sunderland (2000)

acknowledges the upswing in publications dealing with gender as a response to the growing interest in the topic and to an increase in the number of courses on the subject offered in different majors at the university level, as well as a marketing strategy on the part of publishing companies.

Nonetheless, Sunderland (2000, p. 203-204) notes "a lack of material from developing countries where (English) language teaching is widespread and where gender and educational opportunities is the subject of on-going debate." Years later, Sunderland (2010, p. 5) once again points to "a lack of research on gender in both foreign language classrooms and second language settings, […] especially in contexts in which English is not the target language." These quotes pinpoint the importance of research in the area of gender in connection with the English language, in the role of target and non-target language as well. Sunderland (2010) adds that the studies involving gender and SLA can be grouped into three main areas: 1) that dealing with the *subject matter* in the sense of use and usage along with the competence and performance of the language being learned; 2) the area dealing with the *processes* of learning and teaching of such languages (involving aspects such as motivation, attitude or achievement for each group); and 3) the area analyzing *classroom materials*, in their treatment of gender in published books, handouts or teacher guides.

The results of the present research, portrayed in the discussion section on each test, are framed with relation to the second group above because of its emphasis on the effects that the *process* of second language acquisition has on each gender group. The present research focuses on a more social (rather than biological) approximation of the concept of gender (Sunderland, 2010), with the intention of analyzing how classroom setting and methodology, as well as the social context surrounding the learners, may affect the language learning process of the different gender groups and the other way around.

While there may be a general belief directing people to the idea that female students tend to be more skillful at learning second languages, the results in this field are not yet conclusive. Authors such as Sunderland (2000) attribute this belief to possible aspects such as the type of tests used, an overwhelmingly high encouragement to girls in the field, or the fact that girls often are channeled away from other subjects and toward foreign language related careers and subjects. Others attribute this so-called female superiority to neuro-cognition factors of languages in connection to biological differences in male and female subject ability to perform certain language-related tasks, particularly a female advantage with declarative memory (Ullman et al., 2008). On the other hand, van der Slik, van

Hout and Schepens (2015) insist on a nature-based, gene-environmental correlational framework in which the interaction of genetic and environmental factors can mediate between the natural endowment of individuals and their ability to acquire a second language. Regardless of what causes gender differences in SLA, the fact remains that results have not been conclusive thus far.

With reference to foreign language vocabulary learning, Sunderland (2010) argues that many factors influence the mixed results yielded by the studies. Among these, she mentions the ranges of ages and levels that are often compared, as well as the different activities and social contexts, all of which are key in the process of teaching and learning foreign and second languages. Sunderland (2010) insists that while vocabulary research in the foreign language classroom may be able to identify both similarities and tendencies associated with gender groups, they ought to be seen simply as such. That is, for Sunderland (2010), the results of studies ought to be seen as *tendencies* exhibited by the groups under scrutiny, given that "there will always be 'gender overlap' and indeed 'intra-group differences' (among girls, among boys) as well as 'inter-group similarities' (between girls and boys)" (p. 12).

Gender as a variable is not always analyzed as such in SLA studies. The research addressed here to provide a background for the present analysis represents only a small sample of the research in the field and offers a glimpse of the findings that are often discussed in this area. For this analysis, these papers have been classified very broadly into three areas in which they share, to the best possible extent, similar features. The first group includes those that look at gender with relation to different aspects of SLA; they include issues such as motivation, strategy use, aspects related to language skills such as reading or writing, and general language proficiency issues. The second group describes research that deals with gender and different aspects of vocabulary use, either in connection to language skills or associated with the use of vocabulary for specific tasks. The third group includes studies that examine receptive and productive vocabulary size and that have analyzed the possible effects that the variable gender may have on them. Sunderland (2010, p. 5) summarizes this as follow: "In SLA […] sex (usually rendered 'politely' as gender) as an independent variable has been investigated less than other potentially relevant variables." The present section of this study attempts to contribute to this lack of research in the field, albeit only slightly.

5.1.1. SLA studies

Table 8 provides a summary of the first group of research which concentrates on different aspects of language proficiency. The table displays the scope of the study, the instrument that was implemented to collect the data and the findings with regard to gender. A preliminary look at the findings shows that different gender results are found depending on the activity or skills under scrutiny. Contrary to the common belief, males outperform females in many of the tasks in the present group of studies. There seems to be a connection between the type of activity and the outcomes of the different analyses on gender. For studies addressing results of language proficiency in a broader sense, some common ground is observed. For example, male participants seem to outperform female participants in vocabulary measures (amongst other things) in Lin and Wu (2003), Lynn, Ferguson and Horwood (2005) and Law, Rush, Parsons and Schoon (2013). The latter two show this tendency in children whereas the first study involved adult population. Male participants also outperform female participants in vocabulary knowledge in Ackerman and Ellingsen's (2014) work.

The research mentioned above shows female participants outperforming male counterparts in listening tasks (Lin and Wu, 2003), coding tasks (Lynn, Ferguson and Horwood, 2005), and non-verbal performance and reading (Law, Rush, Parsons and Schoon, 2013). Yet in other studies on general language proficiency, van der Slik, van Hout and Schepens (2015) found females outscoring males by far in speaking and writing skills. These authors found no gender-associated gap when reading and listening skills were evaluated.

Song, Cheng and Klinger's (2015) results favor female learners in overall language proficiency, particularly at the level of essay writing and cloze tasks performance. More significant is that some of the results discussed above often change when analyzed under different conditions. For instance, in the case of Law et al. (2013), the gender distinction disappears when the same population is assessed at adulthood. Furthermore, in studies such as Lin and Wu's (2003), no overall gender differences are observed when the tests results are analyzed as a whole rather than looking at specific item bundles. For Lynn, Ferguson and Horwood (2005), given that boys outscore girls at a non-significant level in the full IQ scale indicates that gender does not have an effect on general intelligence. This means that, overall, this research indicates differences in terms of participants' performance in specific tasks, but not at the overall level of language proficiency.

Table 8. SLA studies and gender findings

Author	Focus	Instrument	Findings
Baker and MacIntyre (2003)	Attitudes, orientations for learning and communication affect towards FSL	Survey and questionnaire (N = 71)	- While favoring spatial-related tasks, male non-immersion students presented the least positive attitudes towards learning French. - Female non-immersion students presented the highest endorsement of language learning orientations, especially verbal tasks. - In terms of 'travel' orientation, males show significantly lower means. - Gender differences are identified in attitudes towards and reasons for studying French, but the overall pattern is inconsistent.
Lin and Wu (2003)	General English language proficiency	English Proficiency Test in China (N = 4,459)	- Females outperformed males in the listening test. - Males outperformed females in the grammar, vocabulary and cloze sections. - With all bundles of the test combined, no overall gender differences appear.
Lynn, Ferguson and Horwood (2005)	General intelligence test with language proficiency assessment	WISC-R (Revised Wechsler Intelligence Scale for Children) (N = 897)	- Boys significantly outperformed girls in the areas of information, vocabulary, block design and object assembly. - Girls significantly outperformed boys on coding tasks. - Boys outscored girls (although not significantly) on the verbal, performance and full-scale IQs; gender does not have an effect on general intelligence.
Brantmeier (2010)	Reading literal, inferential and incorrect units recalled	Different reading passages (N = 78)	- Male outranked female participants in literal ideas recalled from familiar texts. - Inferential ideas were not always affected by gender and content. - Incorrect ideas recalled were not affected by gender or content.
Gorjian & Javadifar (2013)	Effects of gender and readings topic on reading comprehension test	Topic familiarity questionnaire, male and female oriented and neutral reading comprehension tests, vocabulary (N = 60)	- Male participants outscored female participants in their corresponding gender oriented test. - Male and female scores were associated with their prior knowledge and interests. - Neutral reading comprehension test neutralized the effects of reader's gender.
Law, Rush, Parsons and Schoon (2013)	Gender and receptive vocabulary, literacy and non-verbal proficiency	Measures were collected at age 5 and at age 34 from the British Cohort Study (N = 5,364)	- At age five, boys showed higher, statistically significant receptive vocabulary differences in comparison to girls while girls had statistically significant advantages in non-verbal performance and reading. - Gender is not associated with literacy factors in adulthood.

Ackerman & Ellingsen (2014)	Vocabulary over-claiming, self concept and verbal ability	Questionnaires for vocabulary knowledge, vocabulary assessment, verbal ability (N = 198)	- Male participants both claimed greater vocabulary knowledge and significantly outperformed females in average total vocabulary scores. - The study showed no statistically significant differences for over-claiming scores. In objective measures, gender differences were approximately equal.
Fernández Fontecha & Canga Alonso (2014)	Gender differences in motivation in CLIL and non-CLIL settings	An adaptation of Gardner's (1985) attitude/motivation test battery (N = 62)	- Boys and girls show similar means for intrinsic and extrinsic motivation. - Non-CLIL girls are slightly more motivated than non-CLIL boys. - CLIL boys are slightly more motivated than CLIL girls. - Non-CLIL boys are more motivated than CLIL boys. - Non-CLIL girls are more motivated than CLIL girls. - Males' intrinsic motivation in both groups is higher than females' although it is not statistically significant. - Non-CLIL girls and CLIL boys show non-significantly, higher extrinsic motivation.
Heras & Lasagabaster (2015)	Gender differences on affective factors (motivation and self-esteem) and CLIL	Background, motivation and self-esteem questionnaire and vocabulary test (N = 46)	- No gender-based differences in motivation were identified in CLIL group. - Non-statistically significant differences show females outperform males in instrumental and Ideal L2 self- motivation; CLIL- females show statistically significant differences in this factor. - Male participants outscored females in ought-to L2 self, self-esteem in the EFL setting and self-esteem in the PE-CLIL class, the first one being statistically significant. - Male non-CLIL students show superior, non-significant Ought-to L2 Self. - Non-CLIL females more strongly see themselves as future users of English. - Female students had higher scores, non-statistical differences in vocab.
van der Slik, van Hout & Schepens (2015)	Gender and Dutch as a second language general proficiency	State Examination of Dutch as a Second Language (STEX); data contained in a large-scale data base of L2 Dutch (N = 29,767)	- Female learners consistently and robustly outperformed male learners in speaking and writing proficiency, regardless of their country of origin and mother tongue. - No gender gap was identified in the skills of reading and listening. - Female learners profit more from higher education training than male learners, the higher the educational level the larger the differences.
Song, Cheng & Klinger (2015)	English general proficiency	Grad. School Entrance English Examination (N = 13,745)	- Overall, female participants outperformed male participants, in particular the tasks of essay writing and cloze tasks.
Lasagabaster (2016)	EMI, motivation, gender and L1.	Spanish university level students (N =189)	- Gender and L1 show no significant effect on participants' motivation towards L3 learning.

Studies evaluating reading also show male superiority on reading tasks involving recalling literal ideas from familiar topic texts (Brantmeier, 2010); however, he finds no gender differences when the tasks describe assessment of inferential ideas or incorrect ideas recalled. Gorjian and Javadifar (2013) found that prior knowledge and interests play a key role in the performance of gender groups in reading. They show that when the topic was gender oriented (based on each gender's selection of topics), male participants outperformed female counterparts. These results conflict with the findings that have identified female students' superiority in the reading skill.

To close the analysis of this first group of studies we can look at research related to attitudes and motivation. Baker and MacIntyre (2003), while acknowledging inconsistencies in their results, identified different tendencies, attitudes, and reasons that could be associated with gender in connection with studying French. They identified female students as portraying stronger language learning orientations and verbal-task abilities. Male participants, on the other hand, showed the least positive attitudes toward learning but performed better in spatial-related tasks. For Fernández Fontecha and Canga Alonso (2014), in general, both genders show similar means of intrinsic and extrinsic motivation. They do describe gender differences associated with the methodology followed in the school. For example, non-CLIL boys and girls are more motivated than CLIL boys and girls; and CLIL boys are more motivated than CLIL girls. They found that, in general, boys exhibit higher levels of intrinsic motivation, although at a non-statistically significant level. Heras and Lasagabaster (2015) results differ from the previous ones in the sense that they do not find gender-based differences in terms of motivation for CLIL students. If specific aspects are analyzed, CLIL female students outperform males in *instrumental* and *Ideal L2 Self* motivation, while CLIL male students outrank females in *Ought-to L2 Self* (statistically significant) and *self-esteem* (non-significantly). The only statistically significant difference identified in the non-CLIL group was the feature of *Ideal L2 Self* image portrayed by the female group. In a later study, Lasagabaster (2016) found no significant effect of gender and L1 in motivation toward learning a third language.

5.1.2. Studies on vocabulary use in specific tasks

The second group is comprised of research dealing with strategy use, written compositions and aspects such as lexical diversity and availability. As before, the

scope, the instrument, and the findings of each study are depicted in Table 9. The studies that analyze strategy use show that there are a lot of similarities in each gender group. Jiménez Catalán (2003) claims that male and female students share 8 out of the 10 most common strategies. Also, while females use a larger number of total strategies, these are more structured vocabulary strategies such as rule-based, rehearsal and planning strategies. Male students, on the other hand, prefer more *image* vocabulary learning strategies. Lee (2007) argues that no gender differences are associated with the pattern and the frequency of use of strategies. For Lee, both genders use the most common vocabulary strategies. Furthermore, Fernández Fontecha's (2010) study of motivation in connection with productive vocabulary determined that female students outscored male participants in vocabulary counts and motivation levels.

The group of studies on written compositions has similar results favoring female students. Ojeda Alba and Jiménez Catalán (2007) claim that females use more types and tokens of vocabulary and longer words showing better memory capacity. For Agustín Llach (2010b), males produce a larger number of word inventions when they write. Like Ojeda Alba and Jiménez Catalán (2007), Agustín Llach (2010b), as well as Jiménez Catalán (2010), assign statistically significant differences favoring female students in terms of the number of words they produce per composition. In yet another study, Díez Prado (2010) concludes that no gender differences are found for lexical density among the gender groups or language groups.

Agustín Llach (2009) (in Agustín Llach 2010a) and Fergadiotis, Wright, and Capilouto (2011) determined no statistically significant gender differences in lexical transfer in the former research, or in discourse type and lexical diversity, in the latter research. Finally, Agustín-Llach and Fernández Fontecha (2014) found statistically significant advantages for females in terms of words produced per word-cue.

This set of studies shows a general tendency toward the superiority of female students in production, particularly writing tasks. Studies such as those of by Ojeda Alba and Jiménez Catalán (2007), Agustín Llach (2010b), and Jiménez Catalán (2010) are in line with the findings discussed in the previous section where van der Slik, van Hout and Schepens (2015) argue that female students are better at writing. Given that writing is one of the foci of the present research, it is useful to compare its results to those discussed.

Table 9. Vocabulary studies and gender findings

Authors	Focus	Instrument	Findings
Jiménez Catalán (2003)	Vocabulary learning strategies	A questionnaire (N = 581)	- Both genders share 8 out of the 10 most common strategies used. - Female total strategy use percentage is higher than that of male students at a statistically significant level. - Female students favor the use of formal rule strategies, input elicitation strategies and rehearsal and planning as well as social elicitation strategies. - Male students use more image vocabulary learning strategies.
Lee (2007)	Vocabulary learning strategies in Korean students	Vocabulary Learning Strategy Survey (N = 466)	- No gender effect was noted for the pattern of use and frequency of strategy use. - Both male and female students used the same five most frequently used strategies while the least frequently used strategies were also very similar. - Only in the case of 'picture and grouping strategies' did male students outperformed females in their use.
Ojeda Alba & Jiménez Catalán (2007)	Children's use of words	Written composition (N = 271)	- Female students use more types and tokens of vocabulary than males. - Female use longer words and seem to exhibit a better memory capacity than males
Díez Prados (2010)	Lexical density and lexical variation in writing	Written compositions (N = 2,271)	- Non-native males outperformed female counterparts in lexical variation (no difference is found if errors are removed). - Germanic language, male participants outranked female participants in vocabulary range. - Romance language, male participants outperformed female participants in lexical variation. - No difference was found for lexical density among gender or language groups. - Female outperformed male population in grammatical metaphor use across all groups.
Agustín-Llach (2009, 2010a)	Lexical Transfer	--	- There are no gender differences in lexical transfer across different grades.
Agustín-Llach (2010b)	Lexical creations in writing	Written compositions (N = 298)	- Male students show slightly more lexical inventions and maximum values across compositions although neither of these is statistically significant. - Female students produce a statistically significant higher number of words per composition.

Fernández Fontecha (2010)	Effect of Motivation in vocabulary production	Lexical availability test (N = 250)	- Female students outscore male participants both in vocabulary counts and motivation measures.
Jiménez Catalán (2010)	Vocabulary measures variation across different tests	1,000 Word Test (Nation, 1983), 2,000 VLT (Schmitt et al., 2001), a written composition and a cue test (N = 210)	- No gender differences are found in the 1,000 and 2,000 vocabulary tests. - Female students outperform males with statistically significant differences in the composition and in the cue word test.
Moreno Espinoza (2010)	Productive lexical profiles through word association	Lex30 (Meara and Fitzpatrick, 2000) (N = 225)	- Girls provide higher means in numbers of types and tokens although these are not statistically significant. - Both genders present similar productive vocabulary. - In terms of word associations, the results show not only similarities but also differences in the tendencies exhibited by boys and girls.
Fergadiotis, Wright, Capilouto (2011)	Discourse type on lexical diversity	Discourse tasks and language samples (N = 86)	- No significant differences are found based on the variable gender.
Agustín-Llach and Fernández Fontecha (2014)	Lexical availability	Production task where students generated as many items as possible in response to a cue (N = 190)	- Females write significantly more words per word-cue than their male counterparts. - Both gender groups increment the number of words they retrieve in the second data collection after three additional years of instruction.

5.1.3. Receptive and productive vocabulary studies

This much smaller final group of analyses (summarized in Table 10) concentrates on the receptive and productive vocabulary size of learners, which have considered the variable *gender* in their analyses. Almost all used the VLT and the PVLT as the instruments to collect the data. This last group may be the least consistent in terms of tendencies favoring one particular gender group. This lack of consistency may indicate that gender is really not the determinant factor in the vocabulary size of learners. Contrariwise, other factors (methodology, topic, motivation) could have a stronger effect on word-size estimates.

Six out of the eight studies claim no gender-related differences in connection with vocabulary size: Jiménez Catalán and Terrazas Gallego (2005-2008), Jiménez Catalán (2010), Moreno Espinoza (2010b), Agustín Llach and Terrazas Gallegos (2012), Canga Alonso and Arribas García (2014), and Özönder (2016). In contrast, Canga Alonso (2013a) is the only study describing statistically significant differences favoring male students. On the other hand, the one paper that shows a female advantage in total word counts (Canga Alonso, 2013b) also determines that this preliminary numerical advantage is not statistically significant. Both studies measuring productive vocabulary claim that female students produce higher word counts although this advantage is not statistically significant in either case. This tendency is in line with findings linking better productive skills with the feminine gender (Ojeda Alba and Jiménez Catalán, 2007; Agustín Llach, 2010b, van der Skil et al., 2015; amongst others).

Considering these studies, we are led to believe that the findings are not yet conclusive. They do show a tendency toward gender not being a determinant factor in connection with the word size of learners, but more research is still necessary. Despite a noticeable increase in research that uses the VLT and the PVLT in the last 15 years, not all offer results regarding the role of the variable *gender*. More work is required in this field and the present study aims to contribute in this area. Prior results show a tendency toward gender not being a controlling factor in word sizes; we will see below whether that tendency is maintained in the present research.

Table 10. Receptive and productive vocabulary and gender findings

Study	Focus	Instrument	Findings
Jiménez Catalán and Terrazas Gallegos (2005-2008)	Receptive vocabulary levels of primary students	VLT (Version 2, Schmitt et al., 2001) (N = 240)	- No statistically significant differences were found in relation to gender despite the numerically higher word counts of female students.
Jiménez Catalán (2010)	Productive vocabulary and word association	VLT (Version 2, Schmitt et al., 2001) (N = 210)	- No gender differences are found in the 1,000 and 2,000 vocabulary tests.
Moreno Espinoza (2010)	Productive lexical profiles	Lex 30 (Meara and Fitzpatrick, 2000) (N = 210)	- No statistically significant gender differences are found, although female students produce a larger number of types and tokens.
Agustín-Llach and Terrazas Gallegos (2012)	Receptive vocabulary size of Spanish EFL students	VLT (Version 2, Schmitt et al., 2001) (N = 176)	- No significant difference between genders regarding receptive vocabulary was identified.
Canga Alonso (2013a)	Vocabulary size of Spanish, secondary students	VLT (Version 2, Schmitt et al., 2001) (N = 192)	- Male learners outperformed female participants with a means that is statistically significant.
Canga Alonso (2013b)	Vocabulary size of Spanish, primary students	VLT (Version 2, Schmitt et al., 2001) (N = 79)	- Female students outscored male counterparts in terms of higher scores and word estimate values, although these differences are not statistically significant.
Canga Alonso and Arribas García (2014)	Productive vocabulary size young learners	PVLT (Laufer and Nation, 1995, 1999) (N = 38)	- No gender differences are found in the productive vocabulary size of the participants, despite higher word estimate exhibited by female students.
Özönder (2016)	Vocabulary size for teaching students	VLT (Version 2, Schmitt et al., 2001) (N = 104)	- No significant differences are seen between males and females regarding receptive vocabulary size.

5.1.4. Concluding remarks

As Fernández Fontecha (2010, p. 105) puts it: "[w]hether supported by evidence or merely driven by uninformed stereotypical opinion, languages have been traditionally labeled as a female subject." Although some of the above data do support this idea, others clearly show opposing results. Sunderland (2010) credits the frequently mentioned mixed results to aspects such as age, language level, classroom-setting activities and individual contexts. She finds it "unsurprising" that all those features render these mixed results. All in all, the scarcity of studies describing receptive vocabulary size and gender is evident. They are even scarcer for productive vocabulary size and gender. The present study aims to contribute information to academic research on both fields.

Gender differences can be identified in many abilities; they may vary, and their roles may be reversed in different studies. That is, whereas gender does not seem to be the driving factor that influences the capacity to learn a second language, there are gender tendencies in certain abilities that may lead subjects to perform better on certain tasks, given the conditions under which tests are provided. This should also serve as an indication that the type of task and the social features of the context surrounding the research may influence the role that gender plays in the diverse learning groups. Gender appears to be only one of the many variables affecting the myriad of aspects involved in the process of second language acquisition. As researchers, we must make certain that the specific aspects of the measures used in our studies do not accentuate the differences in connection with gender groups. This practice, along with additional research, may help to settle the controversy surrounding gender and second language acquisition.

5.2. MOTIVATION IN SECOND LANGUAGE ACQUISITION

Dörnyei (1994) and Dörnyei and Chan (2013) describe the evolution of the study of motivation in SLA as a process that started with Gardner and Lambert (1959, 1972) establishing motivation as a socio-psychological phenomenon. At that time, Gardner and Lambert introduced crucial concepts such as the distinction between integrative and instrumental motivation. Later, according to Dörnyei (1994) and Dörnyei and Chan (2013), the next step in the evolution of research was taken by Clément (1980) who brought forward the idea of self-confidence, and by

Gardner (1985) and Gardner and MacIntyre's (1992, 1993) socio-educational model, the basis for the socio-educational model of motivation in SLA which has led research in the area for decades.

According to Dörnyei (1994) and Dörnyei and Chan (2013), the authors mentioned above have paved the way for a stronger, better-defined, educational analysis in motivation research. Dörnyei (2009b) argues that the concept of integrative motivation put forward by Gardner and Lambert (1959) has played a central role in L2 motivation research. The *L2 Motivational Self System* proposed by Dörnyei (2005) is capturing most of the attention in the current research arena, and it contains features inherited from the integrative motive theory. The L2 Motivational Self System concentrates on the self-perception of L2 learners and their aspirations for future self-states.

Based on the above, simplified description of the major changes in the development of studies on motivation, it could be concluded that although there is evidence of certain changes in the study of motivation, it is still possible to identify a permanent existence of a strong socio-educational orientation of that research through the years. To analyze the most common current trends of motivation research, we will look at two important contributors. The work of Gardner (1985, 2007), and to a larger extent, that of Dörnyei (2005, 2009a, 2014, 2017) will be the focus of the present section.

5.2.1. Gardner's integrativeness

Gardner (1985), in his book *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*, set the basis for most of the research in motivation in the past three or four decades. He put forward key concepts that even today serve as the groundwork for more contemporary theories. Gardner is credited for his contribution in defining distinctions such as the dichotomies of integrative/instrumental or intrinsic/extrinsic motivation, which have been part of research on motivation in SLA for decades.

Gardner (1985) has explored several key aspects in depth. He studies individual differences such as aptitude and the corresponding individual abilities. Gardner analyses the effect of personality, as it is reflected in multivariate studies. He also considers attitude, in general and specific terms, and its effect in connection with the target language, the L2 community, and the language classroom. Gardner

(1985) posits that second language achievement is highly influenced by attitudinal and motivational factors, leading to an integrative motive in L2 learning. Particularly important is his consideration of the socio-educational model which represents a dynamic interplay that describes the interactions of individual differences as they relate to the environment and the acquisition context in which the individual is immersed. This set of interactions results in either linguistic or non-linguistic outcomes. This model expresses the concept of *integrativeness* that remains in Gardner's later work even now.

According to Gardner (1985), having as a basis the cultural belief system of the individual, the integrative motive part of the theory serves as a space where features such as an individual's integrativeness and the attitude toward learning the language have an effect on motivation while interacting with it. These features are, at the same time, influenced by that individual's aptitude and the context (formal or informal) where language learning is taking place. As a result of the interaction of these factors, we obtain either a linguistic or a non-linguistic outcome. In the context of SLA, this would result in the individual attaining a given level of second language expertise or in failing to do so. The focus of integrativeness is on individuals and their interest in learning the language to interact with members of the target community and/or to learn more about this community.

As is the case of other concepts in the area of second language learning, defining motivation is complex. Gardner (1985) recognizes that difficulty and proposes the following definition:

> Motivation in the present context [SLA] refers to the combination of effort plus desire to achieve the goal of learning the language plus the favorable attitudes toward learning the language. That is, motivation to learn a second language is seen as referring to the extent to which the individual works or strives to learn the language because of a desire to do so and the satisfaction experienced in this activity. Effort alone does not signify motivation. (p. 10)

The definition points to a clearly delineated combination of elements (i.e. effort and desire to achieve) that give way to motivation, as the moving source of power to reach a goal. Gardner (2007) explains that it is easier to indicate characteristics of a motivated individual than it is to define the concept with greater precision; and describes a motivated individual as a goal-directed, persistent and attentive person who has desires and expectations, shows effort, positive affect and

self-confidence, and has reasons. We can recognize that all these attributes are easily identifiable in individuals that exert effort and are striving to achieve the skills of a second language.

In addition, Gardner acknowledges that multiple factors are involved in the concept of motivation. Gardner (2007) identifies two fundamental contexts that function as roots of motivation to learn a second language in the school setting. The first one is the *educational context*, defined as the type of educational system in which the student is enrolled, particularly that conformed by the language classroom. The educational context includes, but is not limited to, the set of expectations that are created in connection to the school setting, the properties of the program students are part of, the interests that are pursued, the instructor's enthusiasm and skills, and the appropriateness of the materials. The educational context includes the idea of *attitudes toward the learning situation*; that is, if we assume that the latter are, in turn, derived from the former. The second context mentioned by Gardner (2007) is the *cultural context*. This context, on the other hand, covers socially relevant variables, known as *integrativeness*, and what Gardner (2007) later calls *Openness* or *Openness to Cultural Identification*. For Gardner (2007), integrativeness serves as a reflection of an individual's attributes that can be traced back to the cultural context of the language.

Moreover, according to Gardner (2007), integrativeness entails three basic aspects. First, it means that the individual has the motivation to learn the other language. Second, the learner has a genuine interest in interacting with members of the community who speak the language. Third, the individual demonstrates a positive attitude toward the language-learning situation. He maintains that the integrative motive plays a fundamental role in terms of possible levels of language attainment. Gardner (2007) argues that, beyond the dichotomies of integrative/ instrumental or intrinsic/extrinsic motivation,

> it is the intensity of the motivation in its broadest sense, incorporating the behavioral, cognitive, and affective components, that is important. Because the Integrative Motive includes the Openness to Cultural Identification (i.e. Integrativeness) as a component it seems likely that it may be more associated with attaining the ultimate level of achievement (i.e. Thought and Automaticity) than other motivational types that lack this component. Classroom learning motivation may promote the acquisition of individual elements of the language, but more is needed to achieve a true mastery of the language. From my perspective the integrative motive serves this need. (p. 19)

From the description above, we gather that there are a number of aspects intervening in the process of second language acquisition when it comes to motivation. For Gardner, having an integrative motive, however, seems to serve as the engine that can take learners to higher levels of second language attainment. It follows that for a well-rounded, integrative motive, students must rely on educational and cultural contexts that support the effort and promote the desire of second language attainment. Schools, curricula planning experts, and instructors should ensure that the educational context and the program provide the students with the best conditions for language development. The cultural context, which does not play a fundamental role in most other school subjects, becomes of paramount importance in SLA, where crucial aspects of motivation find the support needed for students to strive for language development. Gardner's integrativeness theory was also used as part of the basis for the conceptualization of Dörnyei's (2005) L2 Motivational Self System. This well-known theory will be discussed in the following section.

5.2.2. Dörnyei's L2 Motivational Self System

Dörnyei (1998) agrees that motivation is a critical element in second language achievement. He recognizes the complexity of motivation as a determinant element in human behavior, one that shapes and directs human action. For Dörnyei (2014), motivation represents the reflection of the interaction between cognitive and emotional issues in the form of an individual's thoughts, beliefs, and emotions that are visible through that person's actions. For Dörnyei (1998):

> Motivation has been widely accepted by both teachers and researchers as one of the key factors that influence the rate and success of second/foreign language (L2) learning. Motivation provides the primary impetus to initiate learning the L2 and later the driving force to sustain the long and often tedious learning process, indeed, all the other factors involved in L2 acquisition presuppose motivation to some extent. Without sufficient motivation, even individuals with the most remarkable abilities cannot accomplish long-term goals, and neither are appropriate curricula and good teaching enough on their own to ensure student achievement. On the other hand, high motivation can make up for considerable deficiencies both in one's language aptitude and learning conditions. (p. 117)

One of the most recent trends in second language motivation research is guided primarily by the theory of L2 Motivational Self System put forward by Dörnyei (2005). Dörnyei (2009a) maintains that this new approach brings together core, influential L2 theories (i.e. Gardner's integrative motive), with findings that have derived from the area of psychology. The latter has to do with interest in the active, more dynamic characteristics of the self-system. For Dörnyei (2009a), the selves have to do with the particular images that individuals may have about their future conditions. These future selves are usually represented through vivid images and senses that need to be associated with realistic situations. Dörnyei (2009a) and Dörnyei and Chan (2013) argue that there are three basic sources of motivation to learn a second language: first, learners' vision of themselves as successful L2 speakers as expressed through their inner desire to become successful L2 users; second, the social pressures surrounding learners and pushing them to master the L2, which may be related to the family or the society; and third, the role of the learning experience that learners enjoy; that is, the actual experience of learning the language. Dörnyei (2009b) maintains that, traditionally, L2 motivational research has recognized the fact that L2 learning is strongly associated with the learners' personal "core." He maintains that the L2 Motivational Self System continues this tradition as the selves represent the learners' inner characteristics.

Dörnyei (2010), in describing the importance of motivation, maintains: "[…] learners with sufficient motivation *can* achieve a working knowledge of an L2, *regardless of* their language aptitude, whereas without sufficient motivation even the brightest learners are unlikely to persist long enough to attain any really useful language […]" (p. 74, emphasis in original). That is, motivation can change the outcome of a learning process. Aside from the individual endowment at one's disposal to face the language challenge, motivation has a powerful effect on directing learners toward their goal. Additionally, Dörnyei (2014) argues that, by definition, motivation has to do with "why people behave as they do […] Motivation is responsible for *why* people decide to do something, *how long* they are willing to sustain the activity, and *how hard* they are going to pursue it" (p. 519, emphasis in original). That being the case, it is worth considering the possible role played by motivation while conducting research in the second language classroom.

Furthermore, Dörnyei (2009a) describes two concepts that play a substantial role in the L2 Motivational Self System theory: the *ideal self* and the *ought-to self*. The ideal self refers to the representation of traits that one desires to have in the future, ones' personal hopes and dreams. The *ought-to self*, on the other hand, describes the characteristics that one feels is more obligated to have and which are

oriented toward personal or social duties or responsibilities. With that basis, Dörnyei (2009a) describes the three basic tenets of the L2 Motivational Self System theory, all involving the selves mentioned above. For Dörnyei (2009a), the three constructs at the basis of the theory are the following:

> * Ideal L2 Self which concerns the L2-specific facet of one's ideal *self*, if the person we would like to become speaks an L2, the ideal L2 self is a powerful motivator to learn the L2 because we would like to reduce the discrepancy between our actual and ideal selves.
>
> * *Ought-to L2 Self* which concerns the attributes that one believes one *ought-to* possess to *avoid* possible negative outcomes, and which therefore may bear little resemblance to the person's own desires or wishes.
>
> * L2 Learning Experience, which concerns situation-specific motives related to the immediate learning environment and experience (e.g. the positive impact of success or the enjoyable quality of a language course). (p. 18)

Additionally, Dörnyei (2009b) argues that the first component contains the integrative and instrumental motives (Gardner 1985, 2007) that have traditionally been associated with motivation. Dörnyei (2010, 2014) clarifies that this is the *L2 specific facet* of the theory. That can also be associated with the traditional concept of intrinsic motivation, the motivation coming from within the learner. As for the second component, Dörnyei (2009b) mentions that it covers aspects that are traditionally associated with the extrinsic, more instrumental properties of L2 learning. He adds that the last component is more "executive" in the sense that it concerns the learning environment and is thus conceptualized at a different level than the other two components are. Regarding the applicability of the L2 Motivational Self System theory, Dörnyei (2009b, 2010) offers information about a number of studies that have found validity in the theory through empirical studies. Dörnyei (2010) discusses results of other studies that show a highly statistical correlation between aspects of Gardner's integrativeness theory and aspects of the Ideal L2 Self where both concepts tap into the same kind of constructs. This serves to show evidence for the presence of traditional theories through this newer one.

Dörnyei (2009b) argues that there are a number of pre-requisites to create and maintain a strong Ideal L2 self: 1) First, the learner has to construct the Ideal L2 self through the creation of an image whose existence is fundamental for the concept to work. If learners do not possess a clear image of the future self, the

required L2 motivation may not develop. 2) The learner needs to have a continuous capacity to imagine the future ideal L2 self. This capacity has to be constantly boosted for it to become sufficiently effective. 3) Learners need to create an ideal L2 self that fits their possibilities in terms of being realistically plausible; if the self-image is unrealistic, learners will not pursue it. 4) The self-image has to be nurtured for it to stay alive; teachers and their activities in the language classroom can achieve this. 5) The ideal L2 self vision should be operationalized through a plan of action that ensures the vision will become true in the future. 6) Learners should also consider a plan of action that shows them what would happen if they do not succeed. This would warrant greater effort in trying to reach the goal. Dörnyei (2014) further analyzes the topic and offers a series of strategies that would help learners and teachers to construct and maintain the motivational power of an ideal L2 self, future vision. Multiple strategies can be applied by students, at the individual and group level, by teachers and by the context itself (school or institution) to enhance the learners' ability to create and maintain a vivid image of their ideal L2 self vision.

More recently, Dörnyei (2017) has extended the limits of L2 Motivational Self System and applies this theory to the idea of *narrative identity*, where it refers to "the ways in which people organize and understand their experiences and memories in the form of various narratives, such as stories, excuses, myths or explanations, thereby making autobiographical stories the foundations of their self-concept" (p. 89). From this definition, it follows that narrative identity represents the internalized version of any individual's life story that includes factual events from the past and desires for the future. In terms of the applicability of this concept to SLA, Dörnyei (2017) explains that it is the unique feature of the learner's interior ongoing narrative that creates a link between learning and using a second or foreign language. This narrative is, for Dörnyei, in charge of processing past L2-related experiences and constructing future goals associated with L2 learning. Dörnyei's (2017) explanation of the L2 narrative identity is portrayed through Figure 5.

Dörnyei (2017) argues that the language learner's narrative is the essential managerial mechanism for learners' characteristics. Because it drives and regulates all changes in the individual, it serves L2 learners as the element that influences and drives L2 learning. As learners understand the changes that derive from the characteristic adaptations and the learning environment, the language learning narrative will be strengthened, and new adaptations will, in turn, take place.

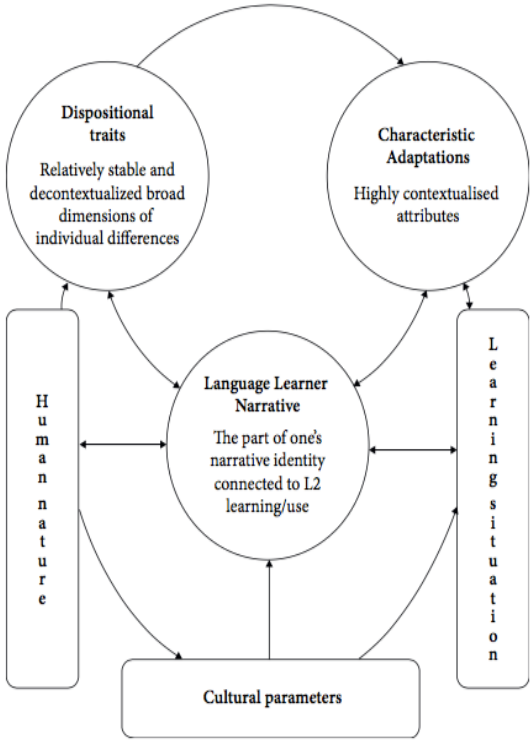


Figure 5. The psychology of the language learner represented in the narrative identity. (Dörnyei 2017, p. 88)

As illustrated in the figure, multiple variables intervene in the process, from the cultural traits to the individual differences and the natural endowment of the individual. While all of these aspects interact, the narrative identity of the individual is developed. If a person's L2 experience and desire for a strong development of L2 skills in the future are part of this development, second language attainment stands a fairly good chance of becoming a reality.

The representation of the learner's narrative identity maintains a close association with the idea of L2 Motivational Self System because they both share several elements and reinforce the second language acquisition process. Both theories take into consideration the individual characteristics of the learner (the ideal L2 self and the human nature), both consider external influences (ought-to self and characteristic adaptations and cultural parameters), and both acknowledge the role of the instructional setting (L2 learning experience and learning situation). Once more, our attention is drawn toward the numerous factors intervening in the second language process and the need for a well-balanced distribution of the roles

and effects on each one of them. Of particular interest in the present study, the role of the L2 Learning Experience/Learning situation will be analyzed with the intention of determining possible effects on the second language learning outcomes.

This possible effect of the educational context and motivation has been analyzed in several studies involving the comparison between CLIL and non-CLIL settings. This comparison could allegedly extend to CBT (presumably a form of CLIL) and FLT (as the non-CLIL form). Navarro Pablo and García Jiménez (2018) conclude that CLIL learners show higher levels of motivation which interact with language achievement levels. These authors claim that results are more consistent in primary than in secondary levels, suggesting progressive reduction of motivation as time progresses. They further maintain that different affective factors seem to have varying effects on academic performance, greater than motivation alone. For them, CLIL versus non-CLIL methodology plays a more significant role in the outcomes of the study, with differences favoring CLIL methodology (see Pfenniger, 2016).

Contrariwise, Fernández Fontecha and Canga Alonso (2014) found that students from non-CLIL settings showed greater motivation than those from CLIL settings. Overall, they claim that the teaching methodology exerts different effects on diverse types of motivation. Lasagabaster (2011) assigns higher motivational and language achievement levels to CLIL students; Lasagabaster (2016) claims that the students' ideal L2, their attitude toward learning English as a medium of instruction, and the family play a significant role in the effort they make to learn an L3. In an older study involving immersion students, Baker and MacIntyre (2003) found that non-immersion students showed lower levels of motivation toward learning French. All in all, the debate about the effects of methodology on students' motivation is still ongoing. This debate cannot be separated from the effect of motivation on their academic achievement; hence, its crucial importance. This is partly the intention of the present study.

5.2.3. Concluding remarks

Albeit brief, this section provides pertinent insights on traditional and current trends in the study of motivation in SLA. The concept of integrative motivation or integrativeness described by Gardner (1985, 2007) still remains

applicable today. Dörnyei's (2005, 2009a, 2009b, 2014) proposal has been credited with validity and applicability in several studies (Dörnyei, 2009b, 2010). Both theories discussed here offer a tripartite constitution that shares several similarities and that focuses on individual aspects, the surrounding or target context and the learning context. These constructs assign a good portion of the success or failure in SLA attainment to the role played by the educational context. The number of elements conforming the different theories evidences the multiplicity of factors involved in the variable motivation. These aspects are important in any second language acquisition setting. This theoretical review adds one more element to the present analysis. Given that in the sample of this study there is variation in terms of the learning context, it is worthwhile to consider whether that factor has an effect on the results of this research.

PART II: THIS STUDY

CHAPTER 6

VOCABULARY LEVELS TEST THEORY AND ANALYSIS

6.1. RECEPTIVE VOCABULARY MEASURES: VOCABULARY TESTS

Vocabulary tests have been found to be highly predictive of students' performance in different language skills (Webb and Nation, 2017). Given their availability and practicality, it could be expected that SLA instructors would take advantage of the benefits that may be derived from the application of this type of test. The predictive factor of vocabulary would be useful in this context to provide students with the tools they need to enhance their knowledge for the functional use of the language, either through listening, speaking, reading, or writing. This section offers an overview of some analyses of vocabulary profiles of students at their receptive level and it provides the results of the present study.

6.1.1. Breadth of vocabulary tests

Several tools are currently available to profile speakers' receptive vocabulary, such as the English Vocabulary Test (Meara, 2010), and its refined version known as Eurocentres Vocabulary Size Test (Meara and Buxton, 1987; Meara and Jones, 1990), the Vocabulary Levels Test (Nation, 1983; Schmitt et al., 2001), and the Vocabulary Size Test (Nation and Beglar, 2007). All of these tests measure written receptive vocabulary and may serve as an indication of students' performance in reading where vocabulary is a crucial factor in the development of reading skills.

The Eurocentres Vocabulary Size Test (Meara and Buxton, 1987; Meara and Jones, 1990) is a *Yes/No* test that serves as a placement instrument for learners' receptive vocabulary knowledge. Its tests assess Level 1 and Level 2 for the first 1000 and 2000 core words of English. Students must score 75% correct answers or higher to be qualified as mastering each level. They are given a list of 60 single words, and have to provide a *Yes* (Y) or *No* (N) answer next to each one of the words depending on whether they know the meaning of the word (*Yes*) or not (*No*). For Meara (2009), this makes the test very easy to apply as students can take a test in less than six minutes because it requires only a couple of seconds for them to decide whether they know a word or not. To control for students claiming to know a word when they actually do not, these tests include a number of non-existing words that help determine the number of words a testee is guessing. Through the

application of a formula, Meara (1992) claims that an adequate estimation of learners' real vocabulary knowledge can be achieved.

In addition, the Eurocentres webpage offers a wide range of tools and electronic language tests that can guide researchers, teachers, and learners in determining the learners' knowledge of vocabulary in multiple languages. In his webpage, http://www.lognostics.co.uk, Meara also offers a wide range of tools related to vocabulary acquisition research. According to Nation (2013), the Eurocentres Vocabulary Size test (Meara and Buxton, 1987; Meara and Jones, 1990), was originally designed for placement purposes. It is a good option as a vocabulary profiler test because it is "quick to administer, is easily computerized for speedy marking, can contain many items thus increasing reliability, and is easy to make" (p. 518). While this test appears as a viable option as a vocabulary diagnostic instrument, the Vocabulary Levels Test (VLT) seems to be more widely used in the area of vocabulary research. We turn now to the VLT.

The VLT (Nation, 1983; Schmitt et al., 2001) has been used in multiple studies to profile students' vocabulary knowledge. Although it was originally created as a diagnostic test (Nation, 1983), it has been widely used as a tool to establish students' vocabulary size of students in many languages. This test measures high frequency and academic vocabulary at different vocabulary levels (Webb and Nation, 2017) while being the best-known and most widely used tool (Read, 2000). According to Schmitt et al. (2001, p. 72), this test is "quick and easy to score, can be readily photocopied, needs no special equipment and gives a more complete picture of a learner's vocabulary than most other tests." These features, along with the fact that it has been proven to be a valid and reliable measure, make this test a very attractive instrument to use in research that involves vocabulary size measures. With this instrument, therefore, we have attempted to determine students' vocabulary size at the receptive level. This test provides the information required to profile them in terms of their estimated receptive vocabulary size. It is our intention to determine the number of words that students can understand without the help of contextual cues. The test, and its original version, has been used for over 30 years in vocabulary research and it has now been translated into multiple languages.

For the VLT, Schmitt, et al. (2001) established a criterion of mastery for each level in terms of a score of 26 or higher out of 30. For these authors, this would indicate that students know the level relatively well. It has been claimed that vocabulary knowledge develops incrementally and that students acquire vocabulary progressively, starting by basic vocabulary words and moving toward more

advanced forms of vocabulary (Schmitt and Meara, 1997; Schmitt, 2007, 2010a; Laufer and Goldstein, 2004; Doczi and Kormos, 2016). Thus, establishing the vocabulary profile for the students at the basic level can help us create a better image of the language development of the students in the sample. This information may contribute to establishing a connection between basic vocabulary knowledge and its contribution to the development of other language skills such as reading and writing.

Finally, the Vocabulary Size Test (Nation and Beglar, 2007) is similar to the VLT in that it is based on sample words taken from word lists based on the British National Corpus and the Corpus of Contemporary American English. This test offers several versions and assesses speakers' vocabulary up to the 14,000 and 20,000 word levels. According to Nation (2013, p. 524), this test requires test takers to have a "moderately developed idea of the word. This is likely to make it a slightly more difficult test than the Vocabulary Levels Test (Schmitt et al., 2001) because the correct answer and the distractors usually share elements of meaning." Given the characteristics of this test, it is clear that it represents a bigger challenge for test takers in comparison to the VLT. Aware of the fact that participants in this study had a limited command of the language, we considered that the VLT (Schmitt et al., 2001) was more appropriate for this population.

6.2. PREVIOUS STUDIES

Table 11 provides a review of studies that explored the development of receptive vocabulary in students. This summary is relevant here because the participants in these studies share a number of similarities with those of the present study. First, in all cases, the same instrument (VLT) was used to measure the learners' vocabulary size. Second, several of the tests describe the vocabulary profile of learners at the secondary level of EFL education. Third, most of the studies include participants whose first language is Spanish. The upper portion of the table presents studies where secondary or upper-level students are involved. In the lower portion of the table, the studies describe the vocabulary profiles of students in elementary levels of EFL.

Table 11. Studies on the development of receptive vocabulary

Study on vocabulary	Size (VLT) in words	Hours of instruction	L1	Learning context EFL/ESL, ed. level
Laufer* (1998) (N = 48)	1,471 1,855	1,080 (10th grade) 1,260 (11th)	Hebrew	Secondary EFL
Qian (2002)** (N = 217)	7,224 6,663	-- --	Korean Chinese	Secondary and upper level EFL
López Mezquita (2005)**	941	1,049	Spanish	Secondary EFL
Olmos (2009)* (N = 49)	1,019	--	Spanish	Secondary EFL
Canga Alonso (2013a) (N = 92)	935	1,049	Spanish	Secondary EFL
Fernández Fontecha (2014) (N = 183)	985	839	Spanish	Secondary EFL
Agustín-Llach & Terrazas Gallegos (2012) (N = 176)	1,206	944	Spanish	Primary and secondary EFL
Fernández Fontecha & Canga Alonso (2014) (N = 30)	1,558 1,658	350 (1st year) 400 (2nd year)	Spanish	Adolescent and adult EFL, Official School of Language
Özönder (2016)* (N = 104)	1,848	--	Turkish	University-level EFL
Jiménez Catalán, Ruiz de Zarobe & Cenoz Iragui (2006) (N = 130)	802 601	-- (Content-based, CB) -- (Non-CB)	Spanish	Primary EFL
Jiménez Catalán & Terrazas Gallegos (2005-2008) (N = 240)	737	419	Spanish	Primary EFL
Agustín-Llach & Terrazas Gallegos (2009) (N = 274)	1,106	629	Spanish	Primary EFL
Terrazas Gallegos & Agustín- Llach (2009) (N = 224)	817	734	Spanish	Primary EFL
Canga Alonso (2013b) (N = 79)	903	944	Spanish	Primary EFL
Fernández Fontecha (2014) (N = 55)	705	839	Spanish	Primary EFL
Agustín-Llach (2015) (N = 68)	592	419	Spanish	Primary EFL (non-CLIL)
Agustín-Llach (2015) (N = 72)	479	700	Spanish	Primary EFL (CLIL)
Agustín-Llach & Canga Alonso (2016) (N = 49)	421 622 827	419 (4th grade) 524 (5th grade) 629 (6th grade)	Spanish	Primary EFL (non-CLIL)
Agustín-Llach & Canga Alonso (2016) (N = 58)	430 649 914	714 (4th grade) 839 (5th grade) 944 (6th grade)	Spanish	Primary EFL students (CLIL)

* Although other levels of the VLT were reported in these studies, the numbers reported here correspond to the 2000 word level; ** As reported in Canga Alonso (2013).

The word counts for students in Laufer (1998) and Qian (2002) stand out as far superior to the results in the other studies, regardless of the students' age. In both cases, the researchers report overall word counts in their studies rather than presenting word counts for each VLT band. The results reported in Table 11 for the 2000 word band in Laufer (1998) are calculated based on the mean scores described by Laufer (1998, p. 262). Laufer (1998) also uses a specific formula to calculate overall word counts in her study. For the results presented in Table 11, however, we used the formula proposed by Nation (1990, p. 78): "Vocabulary size = N correct answers multiplied by total N words in dictionary (the relevant word list) divided by N items in test." In the case of Qian (2002), calculating a separate score for the 2000 word band alone is not possible as he does not refer to the item mean score. In studies such as those, when test takers are assessed across the different level tests (2,000, 3,000, AWL, 5,000, 10,000), comparisons would be more accurate if the results of each word bands were analyzed independently. Webb and Nation (2017) discuss the importance of providing a score for each level of the test, rather than assigning an overall score, which could give the impression that all word bands are known by the learners when this might not be necessarily the case. Although these two studies often appear in research on vocabulary profiles using the VLT, it is not feasible to establish comparable parameters with the present one.

The second group of analyses in Table 11 is composed of studies carried out with secondary and upper-level students mainly in Spain. In the cases of López Mezquita (2005), Canga Alonso (2013a), and Fernández Fontecha (2014), word counts show that students are well below the level of mastery for this vocabulary band. Although in the first two the number of hours is higher than in other studies in Table 11, the word counts are limited. The results in Olmos (2009), Fernández Fontecha and Canga Alonso (2014), Agustín Llach and Terrazas Gallegos (2012), and Özönder (2016) are the ones where students have the largest number of words. In the last two, word counts are above 1,500 words, indicating that students are near mastery levels of this word band. Using Nation's formula in Özönder (2016), students could be said to show mastery of this word band. However, participants in Özönder's (2016) study are already university students. We do not know their exact number of hours of instruction, but we can assume that they are higher than the others in Table 11.

The last set of studies is made up of those whose population came from elementary school settings. There is a wide range of total word counts and total hours of instruction in this group. The highest number of words is 1,106 (Agustín Llach and Terrazas Gallegos, 2009) and the lowest is 421 (Agustín Llach and Canga

Alonso, 2016). At least upon superficial inspection, there does not always seem to be a correlation between the number of hours of instruction and the total word counts. For example, Canga Alonso (2013b) reports the highest number of hours (944), but not the highest word count (903 words). Agustín Llach and Terrazas Gallegos (2009), on the other hand, report lower total hours of instruction (629 hours) and higher total word counts (1,106 words). In another study, Agustín Llach (2015) reports 700 hours, and only 479 words.

Contrary to Agustín Llach and Terrazas Gallegos's (2009) claim, this seems to indicate that there is no direct relation between the number of hours and larger word counts. It would appear that the higher the number of total hours of instruction, the greater the influence of other factors in the total word counts in the profile of EFL learners. At least for some of the cases depicted in Table 11, the type of methodology is not necessarily responsible for the total word counts either. For example, in Agustín Llach (2015), CLIL students show lower total word counts than their non-CLIL counterparts, whereas in Agustín Llach and Canga Alonso (2016), the CLIL students have higher total word counts than do the non-CLIL students. This advantage is, however, reduced if one looks at the much higher number of hours of instruction in the former group.

To conclude this brief review, we must acknowledge that there is great variety in the results obtained by similar groups of students when the VLT has been used to measure these word counts. Interestingly, the Spanish-speaking population displays rather low word counts when compared with students from different L1 backgrounds. All of these studies include Spanish speakers from Spain. The present analysis will contribute to this discussion in providing results from Spanish L1 students from a different context and region.

6.3. THE PRESENT STUDY[5]

6.3.1. Participants

A total of 185 students (55 from the CBS, and 130 from the FLS) completed this receptive vocabulary test. In the CBS group, 21 participants were male and 34 were female. In the FLS group, there were 64 male and 66 female students.

[5] Part of this analysis was discussed in D. Castro-García (2017), Receptive vocabulary measures for EFL Costa Rican high school students. *IJES,* 17(2), pp. 81-99.

Regarding the total number of hours of instruction at high school level, the CB school students have received approximately 1,368 hours, and the mainstream EFL school students have received approximately 1,140 hours of instruction at this level.

6.3.2. Instrument

To measure the students' receptive vocabulary, the Vocabulary Levels Test (VLT), 2000 band, Version 2 (Schmitt et al., 2001) was used. In this test, the participants must complete a total of 30 items. For each one, they match a word to its corresponding definition (with 30 extra words for which no definition is given). According to its authors (Schmitt et al., 2001), in this version of the test, the 2000 level contains words from the first 1000 words and from the second 1000 words in a ratio close to 1:1 (28 from the first 1000 frequency level and 32 from the second 1000 level). Research has shown the VLT to be valid (see Beglar & Hunt, 1999; Cameron, 2002; Laufer, 1998; Schmitt et al., 2001); reliable (Read, 1988; Schmitt et al., 2001); and practical (Schmitt et al., 2001).

6.3.3. Procedure

The data to measure receptive vocabulary were collected in one session in each school, during the English class, in all six sub-groups: two groups in the CBS and four groups in the FLS. Students were allowed 15 minutes to complete the VLT, although many of them finished sooner.

At the beginning of the VLT, both Spanish oral and written instructions were given to explain what the students would be asked to do. A sample item set was provided with the corresponding answers for students to have a clear understanding of the format and mechanics of the test (as in Figure 6). Once the tests were collected and checked, the results were analyzed using SPSS 20 (IBM Corp. 2011) to calculate descriptive values and to measure differences among the different groups.

Vea el siguiente ejemplo:

business		
clock	_____	part of the house
horse	_____	animal with four legs
pencil	_____	something used for writing
shoe		
wall		

Usted debe responder de la siguiente forma:

business		
clock	6	part of the house
horse	3	animal with four legs
pencil	4	something used for writing
shoe		
wall		

Figure 6. Sample item for VLT set.

6.3.4. Results

In the following section, we describe the results for each of the schools in this study. First, we look at the general results for each item in the test. Then, a description based on the statistical analysis of the results follows. Finally, the distribution of words according to gender is discussed for each institution. Appendix 1 offers access to the VLT version used for this task.

6.3.4.1. Item knowledge

Table 12 displays the results relating to correct, incorrect and missing percentages for each of the items in the test. First, we will begin by analyzing the results about the CBS. Four items stand out as posing greater difficulty for students in this group. Those are items 1, 10, 11 and 29. For item 1 the given definition "end or highest point" corresponded to the word "tip." Only 23.6% provided the correct answer for this association. Item 10 presented the word "grow," which was supposed to be associated with the concept "develop." The correct percentage, in this case, reached 60.0%. Item 11 presented the definition "put in order" for the concept "arrange." This item reached a percentage of correct answers of 54.5%. Finally, item 29 provided the word "small." which students had to associate with

the word "slight." In this case, only 49.1% of the answers were correct. There were three more items that presented moderate difficulty for students, "debt" ("something you must pay"); "wage" ("money for work") and "manufacture" ("make"), whose degree of correct answers reached 83.6%, 85.5%, and 83.6% respectively. For the rest of the items in the test, the CBS population reached over 90% of correct answers, with many of them standing well above 95%.

Table 12. Correct, incorrect, and missing percentages per item in the VLT

	Content Based Teaching			**Foreign Language Teaching**		
	Correct	**Incorrect**	**Missing**	**Correct**	**Incorrect**	**Missing**
Item 1	23.6%	70.9%	5.5%	13.7%	64.9%	21.4%
Item 2	98.2%	--	1.8%	100.0%	--	--
Item 3	96.4%	1.8%	1.8%	75.6%	16.8%	7.6%
Item 4	96.4%	--	3.6%	69.5%	16.0%	14.5%
Item 5	83.6%	14.5%	1.8%	67.9%	26.7%	5.3%
Item 6	90.9%	5.5%	3.6%	62.6%	22.1%	15.3%
Item 7	85.5%	12.7%	1.8%	61.1%	19.8%	19.1%
Item 8	96.4%	1.8%	1.8%	94.7%	3.8%	1.5%
Item 9	96.4%	1.8%	1.8%	96.2%	1.5%	2.3%
Item 10	60.0%	36.4%	3.6%	32.8%	45.8%	21.4%
Item 11	54.5%	43.6%	1.8%	21.4%	51.9%	26.7%
Item 12	90.9%	5.5%	3.6%	77.1%	16.0%	6.9%
Item 13	96.4%	1.8%	1.8%	77.1%	12.2%	10.7%
Item 14	92.7%	3.6%	3.6%	59.5%	22.9%	17.6%
Item 15	96.4%	1.8%	1.8%	80.9%	12.2%	6.9%
Item 16	83.6%	10.9%	5.5%	65.6%	19.8%	14.5%
Item 17	98.2%	--	1.8%	98.5%	--	1.5%
Item 18	94.5%	1.8%	3.6%	70.2%	19.8%	9.9%
Item 19	96.4%	1.8%	1.8%	93.9%	4.6%	1.5%
Item 20	96.4%	1.8%	1.8%	75.6%	8.4%	16.0%
Item 21	98.2%	--	1.8%	86.3%	9.9%	3.8%
Item 22	98.2%	--	1.8%	98.5%	1.5%	--
Item 23	98.2%	--	1.8%	89.3%	6.9%	3.8%
Item 24	98.2%	--	1.8%	90.8%	5.3%	3.8%
Item 25	94.5%	1.8%	3.6%	67.2%	23.7%	9.2%
Item 26	92.7%	5.5%	1.8%	76.3%	13.7%	9.9%
Item 27	96.4%	1.8%	1.8%	92.4%	6.9%	0.8%
Item 28	90.9%	7.3%	1.8%	80.2%	15.3%	4.6%
Item 29	49.1%	49.1%	1.8%	43.5%	48.9%	7.6%
Item 30	96.4%	1.8%	1.8%	88.5%	9.9%	1.5%

The situation in the FLS is very different from the one observed in the CBS with many more items presenting greater challenges for this group of students. Interestingly, the same four items that were more challenging for the CBS students

represented the lower levels of correct responses in the FLS. The difference lies, however, in that the percentages of correct answers are much lower for these subjects and in that many other items represented greater difficulty for this group, and thus joined the list of challenging items. In the FLS, the percentages of correct answers for these four items are the following: 13.7% for item 1; 32.8% for item 10; 21.4% for item 11; and 43.5% for item 29. Item 14, "clerk" ("office worker"), reaching only 59.5% of correct answers, is added to the list of difficult items in this school.

While the previous items represented the greatest challenges, from the items on the list multiple other items posed relatively high difficulty. In this list, the meanings of "debt" and "wage," were also difficult for the CBS students. The following are the cases of very difficult items in the FLS: item 3, "copy" ("things made to be like another"), 75.6% correct answers; item 4, "roar" ("loud, deep sound"), 69.5% correct answers; item 5, "debt" ("something you must pay"), 67.9% correct answers; item 6, "pride" ("having a high opinion of yourself"), 62.6% correct answers; item 7, "wage" ("money for work"), 61.1% correct answers; item 12, "prefer" ("like more than something else"), 77.1% correct answers; item 13, "wine" ("a drink"), 77.1% correct answers; item 18, "melt" ("become like water"), 70.2% correct answers; item 20, "dozen" ("twelve"), 75.6% correct answers; item 25, "stretch" ("make wider or longer"), 67.2% correct answers and item 26, "introduce" ("bring in for the first time"), 76.3% correct answers.

On the other hand, the words whose definitions were best identified by the students in this group were: "motor" ("this moves a car"); "skirt" ("a piece of clothing"); "justice" (using the law in the right way"); "elect" ("choose by voting"); "opportunity" ("chance"); "difficult" ("not easy"); "holy" ("related to God"); and "admire" ("have a high opinion of someone"). We cannot help but notice that several of the words that obtained the highest correct percentages keep a resemblance to Spanish terms, although that is not necessarily the case for the words in their definition. It is worth noting, however, that there are items in this same list of concepts that were very challenging for this group. They are transparent, and we would have expected them to be identified by students whose first language is Spanish, such as "copy," "prefer," and "dozen," but this did not occur here. Many students, despite the resemblance of these words to Spanish words, still failed to identify the definition of these words. Schmitt et al. (2001) claim that a Romance L1 provides a distinct advantage particularly for the Academic Level of the vocabulary levels test (not necessarily at other word levels). The results of this study seem to add to the evidence provided by Schmitt, et al.

(2001) in regard to the fact that this test does not necessarily favor speakers of Romance languages at the 2000 word level. As illustrated by the words mentioned above, other factors, aside from similarities between L1 and L2 words, mediate when it comes to students' choosing the definition of a concept. Regardless of the similarities between the cognates and the Spanish words, students did not always identify the correct definition of these concepts.

In general, the preliminary results show that students from the CBS obtained much better outcomes in the VLT. Although the CBT students easily recognized most of the concepts in the test, a few items are difficult for a good portion of this population sample. In the case of the FLT school, the 2000 word band test is still highly challenging, with just a few items being fairly easy for this population group. In the next section, we will explore the results in more depth.

6.3.4.2. Results in relation to word counts

Table 13 presents the first results of mean and standard deviation. When we analyze the mean from the perspective of the 30 items that conform the test we see that it is higher for the CB school; it is placed at 26.91 against 22.15 for the mainstream FL school. The standard deviation numbers show that there is a major concentration of the data for the CB school, while the range of data for the FL schools shows greater dispersion.

These results may be analyzed in greater depth in relation to the 2000 words that are measured in the VLT 2000. Following Nation's (1990, p. 78) formula "Vocabulary size = N correct answers multiplied by total N words in dictionary (the relevant word list) divided by N items in test," we convert the total of correct items into total word counts for each student in each school.

Table 13 also shows that the CB school students achieve a mean of 1,793.94 words out of the 2000 possible words and that the FL school students reach a mean of 1,475.85 words. Although there are students in both schools that reach the maximum 2000 words, the minimum of words is much lower (600) for students in the FL school. Hence, the higher spread of data expressed by the standard deviation observed for this school (349.7) when compared to that of the CB school (152.9) where the minimum number of words was 1,400.

Table 13. Mean, Standard Deviations and word estimates for schools

	VLT 2000 (n = 185)			
	Item count information		Total word count information	
	CB School	FL School	CB school	FL school
Number of items	30	30	2,000	2,000
Mean	26.91	22.15	1,793.94	1,475.85
Median	27	23.50	1800	1,566.60
Max items/words	30	30	2,000	2,000
Min items/words	21	10	1,400	600
SD	2.29	5.24	152.9	349.7

The range of distribution of the data for both schools is more clearly illustrated in Figure 7. As mentioned above, while we can see that for the CB school the lowest score falls around 1,400 words, we can also observe that most of the distribution is clustered above 1,600 words and even more tightly clustered between 1,800 and 1,950 words. On the other hand, this figure shows that the range distribution for the FL school is much more spread along the continuum of the vocabulary band. As mentioned above, it starts well below the 750-word range and unevenly moves up to 2,000 words. It must be noted though that the figure shows a greater clustering of data starting around 1,250 and continues up to 2,000 words, the highest point in this portion of the data seems to cluster around 1,700 words.

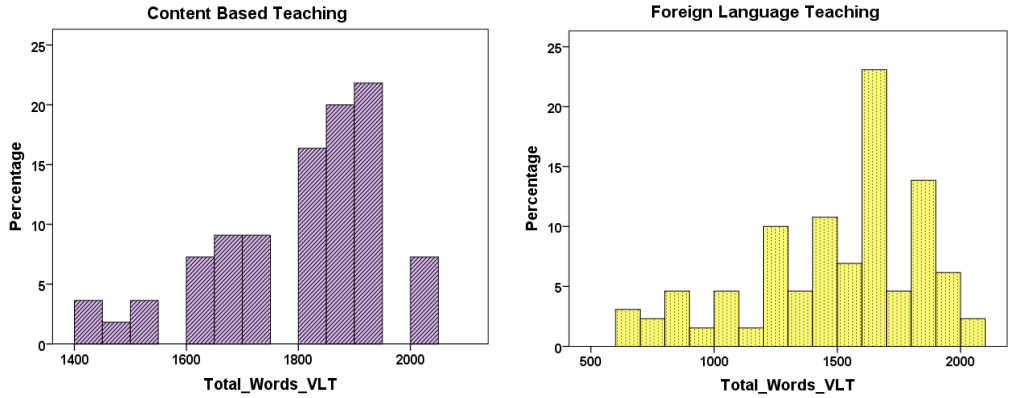


Figure 7. Frequency of distribution of words for CB School and FL school.

Another way of looking at this data, now closely considering the median for each school is through the use of a box plot in Figure 8. The maximum numbers of

words reached as well as the distribution of the data within the 2000 word band discussed above are illustrated in this box plot. We can see that the median shows that a distribution for the CB school slightly leans toward the lower quartiles, whereas for the FL school it shows a more abnormal distribution, where the lower quartiles receive a greater percentage of the data.

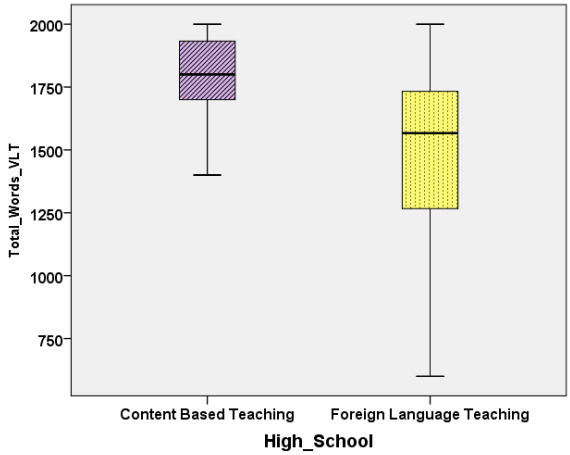


Figure 8. Box diagram of 2K levels in both schools.

Undoubtedly, the spread covered by the number of words for the FL is much wider than the one for the CB school. The spread of distribution is abnormal for the FL school, much more so than for the CB school. The asymmetry of the upper and lower extremes points to an evident higher performance by the CB school participants. Although in both schools the lower extremes indicate that the results are negatively skewed, with the majority of scores concentrated toward lower quartiles, the FL school clearly points to the lowest total word counts. To determine normality assumptions in our data, with the purpose of defining the nature of the differences between the results of both schools in the VLT, a Kolmogorov-Smirnov test was used. The results of this test are presented in Table 14.

Table 14. Parametric test for school for VLT

	High School		**Kolmogorov-Smirnov**	
			gl	**Sig**
Rec_Num_Wrds	CB School	.174	55	.000
	FL School	.139	130	.000

According to this test, the data follow an abnormal distribution, which calls for non-parametric tests to compare medians for two independent variables. The Mann-Whitney U test is thus conducted to test inferential statistical differences among the groups. The results of this test are presented in Table 15. This test reveals statistically significant differences in favor of the CB school, which shows a median rank of 131.70 over 76.63 for the FL school.

Table 15. Results of inferential statistics for the 2K VLT

N=185	**2000 VLT**
Mann-Whitney U	1,446.500
Wilcoxon W	9,961.500
Error	331.762
Z	- 6.416
P (two-tailed)	.000

6.3.4.3. Results in relation to gender

In the following section, the results are analyzed from a gender perspective. The general results are presented in Table 16. The data in this sample provide evidence that the overall receptive vocabulary for boys is slightly higher than that of girls in both schools. Although the average word-count difference is low (a difference of under 100 words between boys and girls in each school), the data still evinces an advantage for boys in terms of total word counts. The standard deviation information is also relevant since we can observe a slightly broader diffusion in the data of word counts in the girls' group for the CB school and a larger dispersion in the results for boys in the FL school.

Table 16. Mean and Standard Deviation based on gender in both schools.

	CB school		**FL school**	
	Boys	**Girls**	**Boys**	**Girls**
Number of items	2,000	2,000	2,000	2,000
Mean	1,828.57	1,772.53	1,512.45	1,440.35
Median	1,867	1,800	1,600	1,533.20
SD	142.35	157.43	361.44	336.88

A box diagram considering the median word counts for boys and girls in each school is presented in Figure 9. In this diagram, we can appreciate the spread of total word counts along the 2000 word band continuum for boys and girls in each school.

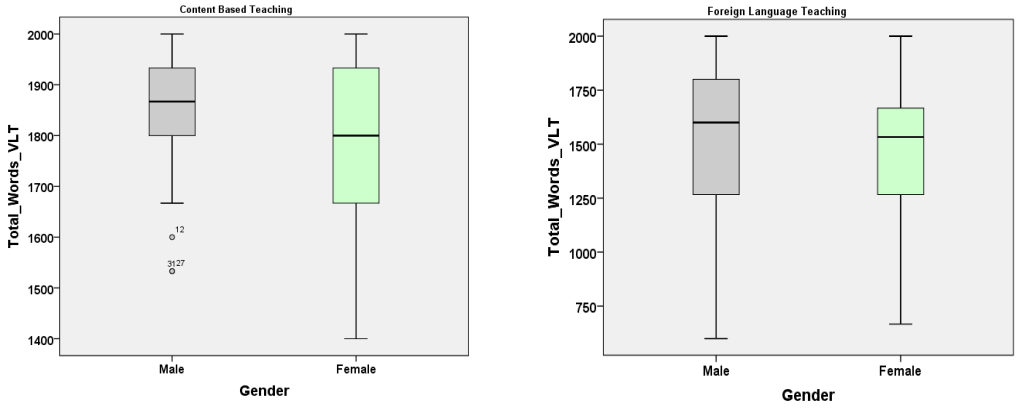


Figure 9. Box diagram for total word counts in the VLT for girls and boys in each school.

Comparing the word counts for boys and girls in the CB school we find that the median shows a slightly homogeneous distribution. The spread of distribution, however, is wider in the lower quartiles than in the upper quartiles, and even more so in the case of the girls. In the case of the boys, the plot provides evidence of the presence of three outliers in the data. The situation is different for the FL school where the median leans toward the upper quartiles leaving a major concentration of data in the upper quartiles. The spread is different for this school, since the word counts are much more widely spread toward the lower quartiles, even more so for the boys. In summary, participants' total word counts show clear differences between the two groups and the two genders. While boys at the CB school show a more concentrated spread in distribution, they exhibit a wider spread of distribution in the FL school. The FL school girls, on the other hand, present slightly more concentration in the spread of distribution in relation to boys than do the girls from the CB school in relation to boys in that school. This figure confirms once more the advantage of the CB school participants in relation to the FL school, this time from a gender perspective.

Once more, normality assumptions are checked for each gender group in each school to establish the characteristics of the differences found. The results are depicted in Table 17. Again, the Kolmogorov-Smirnov was used.

Table 17. Normality distribution test for gender word counts in each school.

		Kolmogorov-Smirnov		
School	**Gender**		**gl**	**Sig.**
Content Based	Masculine	.225	21	.007
	Feminine	.1157	24	.032
Foreign Language	Masculine	.158	64	.000
	Feminine	.139	66	.003

Given the not-normal distribution of the data according to gender, inferential statistics are calculated with the Mann-Whitney U test. The results of the median rank in this test are the following: female 25.76, male 31.62 in the CBS and female 60.51, male 70.65 in the FLS. Based on the Mann-Whitney U test, no statistically significant differences are found in either school or gender group. This means that the apparent difference initially found in word counts favoring male students is not statistically significant in the end. The results of the Mann-Whitney U test are displayed in Table 18.

Table 18. Results of inferential statistics for the 2K VLT in gender in both schools.

	CB School	**FL School**
Mann-Whitney U	281.000	1,782.500
Wilcoxon W	876.000	3,993.500
Error	281.000	1,782.500
Z	-1.333	-1.538
P (two-tailed)	.182	.124

6.3.5. Discussion

The results show that while the CBS obtained an item mean of 26.91, the FLS students obtained a mean of 22.15. Following Schmitt et al. (2001), for whom a criterion of mastery of 26 items or higher is needed, we now determine that the CBT school meets this criterion, albeit barely crossing the line. The FLS, on the other hand, falls short and cannot be said to have mastered the 2000 word band of

the VLT. The information provided by the SD measures also leaves us with worth noting information as it shows that the students in the CB school appear to have better cohesion in terms of vocabulary knowledge. The fact that their results are more concentrated can be taken as an indication that these students share a similar knowledge of this word band. On the other hand, a greater dispersion is observed in the FLS, indicating that students do, in fact, possess different degrees of knowledge of this vocabulary. Overall, the results are evidence that the preliminary numerical differences in the word count of the participants, CBS (1,793.94) and FLS (1,475.85) are, indeed, statistically significant. Thus, the CBS shows an advantage for receptive vocabulary knowledge when compared with the FLS counterpart. The comparison of the present study against results in others seems to indicate that students in the present study may be progressing adequately.

When comparing these results with those presented in Table 11 above, we find that they are close to those stated in Laufer (1998). The participants in the present study, however, have received more hours of instruction (FLT, 1,140 hours; CBT, 1,368 hours) than those in Laufer's (1998) study (10^{th} grade, 1,080; 11^{th} grade, 1,260 hours approximately). In terms of total word counts, Laufer reports 1,471 and 1,855 words for students in 10^{th} and 11^{th} grade respectively. Those numbers are fairly close to the 1,475.85 (FLS) and 1,793.94 (CBS) obtained in the present study. Also, the results of the CBS in the present study are close to those reported by Özönder (2016) for Turkish students, although they are still not as high as those. We do not know the number of hours of instruction for participants in Özönder's study. However, she states that her participants are university students; that could result in greater number of hours of instruction for those participants.

The results of the FLS are close to those of Fernández Fontecha and Canga Alonso (2014), 1,558 and 1,658 for 1^{st} and 2^{nd} year respectively. They are also similar in the sense that none of the participants in either study reached a mastery level. Fernández Fontecha and Canga Alonso (2014) report 350-400 hours of instruction for their study. Given that the adolescent and adult participants in that study come from the Official School of Language context, it is not clear whether that number of hours is the absolute total number of hours that students have been exposed to in English or if students have been further exposed to English through their secondary schooling. As for the rest of studies mentioned in Table 11, in which secondary students participated, the results of the present study are much higher. Agustín Llach and Terrazas Gallegos (2012) have the highest word count, with 1,206 words, followed by Olmos (2009) with 1,019 words. In the rest of the cases, the numbers fall below the 1000 word range.

As expected, the difference in word counts is even more marked when compared with studies involving primary students. In those cases, Agustín Llach and Terrazas Gallegos (2009) report the highest word count of 1,106 words for a group of primary EFL education with 629 hours of instruction. In the rest of the cases, the total word counts are distributed in the scale and go down to around 400 words. This is to be expected as students in the present study are already receiving secondary education.

Regarding the findings on the effect of gender, overall results of mean word counts seem to indicate an advantage for male students, albeit small, in both the CBT and FLT schools. When statistical calculations are applied, however, we find that the apparent numerical advantage presented in total word counts is not significant at the statistical level for either one of the schools. That is, the gender variable is not shown to influence students' performance on the VLT for this sample population. These results are in line with findings of similar studies reported in Jiménez Catalán and Terrazas Gallegos (2005), Jiménez Catalán (2010), Agustín Llach and Terrazas Gallegos (2012), and Özönder (2016). All of them used the VLT, and after analyzing gender effects found no significant difference between male and female participants in receptive vocabulary measures. On the other hand, Canga Alonso (2013a) and Canga Alonso (2013b) did find differences favoring male students in the first case and female students in the second case. All in all, the empirical evidence available through these studies seems to indicate that the gender variable exerts no effect on learners' capacity in connection with receptive vocabulary counts.

We have to keep in mind that the VLT serves to provide a profile of learners' vocabulary. As such we must consider the possibility that the students overall profile for vocabulary may be higher. Especially if we consider that we did not measure vocabulary knowledge at other word bands and that it is very likely that students know (at least some of the words) that belong to other word levels. As the most basic word level, however, we have to conclude that while students attending the CBT setting master this level, students in the FLT setting do not. It has been established that there is need for frequent and repeated exposures to vocabulary for it to be learned (Nation, 1993; Doczi and Kormos, 2016). Based on the present results, it seems that the FLT school (and a percentage of the CBT students) still need more and more effective exposure to vocabulary for them to be able to master this word band.

According to Dóczi and Kormos (2016), the breadth (i.e. size as we have called it) of vocabulary knowledge usually develops very quickly in initial levels

and then it maintains congruence with general language development. Given this important initial development, learning high frequency before low frequency vocabulary is only logical because the former represent the highest percentage of words in common use. The 2000 word band represents between 80 and 90% of the vocabulary that appears frequently in common texts, conversations and novels and it has the powerful effect of either promoting or hindering the development of essential language skills (Nation, 2013). From this, it follows that students would benefit greatly from diligent learning of this word band; the benefits of knowing this word band would be enjoyed immediately. Nation (2013), maintains that knowing this word band warranties access to a great share of written and spoken information; it allows greater comprehension of texts; and it allows production of written and spoken language.

Considering the previous information and the results of the present study we have to point to the dire need of vocabulary learning at this word level in beginning stages of ELF learning. While the students from the CBT context seem to be on the right track to achieve this goal, students from the FLT school need to work much harder to obtain better results. Given the binding benefits of vocabulary in connection with the development of major language skills such as reading and writing, it is only rational that resources and effort are directed toward this goal.

This first part of the analysis has begun to create an image of the vocabulary knowledge of the population under study. In terms of the receptive vocabulary of the sample, we have determined that the students in Content Based Instruction appear to show advantages in terms of passive vocabulary knowledge when compared with students in Foreign Language Instruction. The numerical differences obtained in the VLT are statistically significant in favor of the CBS. The results obtained in the Costa Rican context seem to be higher when compared with the results of studies carried out in Spain with Spanish L1 students. That is not the case when compared with students whose L1 is Turkish and Hebrew. Concerning gender differences, the present portion of the study concludes that there are no statistically significant differences amongst the female and male populations in either one of the school settings. Overall, the apparent advantage of the CBS over the FLS could be associated with the teaching methodology in general. Students in the CB school appear more coherent as a group when it comes to the knowledge of passive vocabulary assessed through the VLT.

CHAPTER 7

READING COMPREHENSION THEORY AND ANALYSIS

7.1. READING COMPREHENSION

Authors such as Hu and Nation (2000), Laufer (2001), Nation (2006a), and Schmitt (2008) insist that vocabulary is essential for the appropriate development of reading skills. In the area of vocabulary and reading, research has been divided into different foci, for example:

- depth of vocabulary knowledge in reading as in Schmitt (2008, 2010a);
- comparison of the results and identification of effectiveness of explicit versus incidental vocabulary learning through reading as in Pellicer-Sánchez (2016), Schmitt (2010b), and Elgort and Warren (2014);
- determining the vocabulary threshold for adequate reading comprehension, as in Nation (2006c), Laufer (1992, 1996), and Laufer and Ravenhorst-Kalovski (2010).

The first two foci were discussed in Chapter 4 where vocabulary theory was discussed; the latter is briefly referred to below. To a more limited extent, research has also looked into the relation between vocabulary size and reading comprehension. This is one of the main interests of this section of the present study.

While it is true that students sometimes need only partial knowledge of the words they are reading, they usually need a wide vocabulary and more in-depth knowledge of these words to comprehend and process information in a text. As Hunt and Beglar (2005, p. 24) contend, "at the heart of language comprehension and use is lexicon." Vocabulary; however, as these authors insist, is often undervalued by teachers and course designers. The purpose of analyses such as this one is to identify the weaknesses that the participant population may have in processing information from a reading to later find ways of providing the tools required to reduce difficulties. Having limitations to understand the content of reading material would demotivate students who could then be discouraged from reading in general. Knowing that reading plays a key role in education, we aim at finding ways to obtain better results from the process so that students read as much as possible to advance in their educational endeavors as well as to entertain themselves through reading as a leisure activity. Because vocabulary carries a heavy load in the reading process, it is important to examine reading results in this study, and subsequently establish how they are related to receptive vocabulary measures.

In this section, we will first review key aspects leading to successful reading comprehension. Then, different studies defining an appropriate vocabulary threshold are covered, and an analysis is provided of the results of the reading task for the participants in this study. Finally, research on the relation between vocabulary levels and reading comprehension is discussed before going on to examine the relation between receptive vocabulary size and reading in the present study.

7.1.1. Reading skills development

Specialists in the field have analyzed key aspects associated with reading skills development and reading comprehension. Some of the findings are discussed in the following paragraphs with the purpose of recognizing the role that vocabulary plays across the different aspects required for adequate reading comprehension. Schmitt (2008) points to the importance of automatic lexical recognition that ensures an appropriate rate of decoding which, in turn, warrants an appropriate flow of information. That means that if our speed of recognition is fast enough, then, reading will progress without problems and will represent a pleasurable and rewarding activity. To be able to achieve rapid reading skills, a large receptive vocabulary size is required. Because we want our students to enjoy the reading process, we ought to provide them with the tools they need to learn the vocabulary that will lead them to do so.

Just as with other language skills, mastery of vocabulary plays a key role in enhancing reading ability. Along this same vein, reading plays a key role in successful academic attainment in school. Bourgoin (2014) notes that some literacy notions transfer across languages and further argues that phonological processing skills and decoding abilities, both of which are fundamental in the development of reading skills, are correlated in L1 and L2 (see Cummin's Interdependence Hypothesis, section 1.1.1.2). In addition, Elgort and Warren (2014) point out that L2 vocabulary learning through reading is affected by the same factors that affect SLA in general. Age, initial L2 lexical proficiency, L1, learning strategies and even motivation become active factors in the process of vocabulary learning. They further argue that strategies such as dividing the readings up into smaller chunks as well as practicing thematic reading can aid the quality of L2 vocabulary development. In sum, multiple aspects intervene in the development of fruitful

reading skills. The following studies discuss several practices that can enhance reading abilities, and they show how vocabulary remains a key factor in the process.

Bingham Wesche and Paribakht (2000) carried out an investigation using think-aloud protocols to determine how different types of text-based vocabulary exercises promoted lexical processing and learning. Comparing the results of students who concentrated on thematic reading for comprehension (*reading only*) with those of students who completed a number of vocabulary exercises in addition to reading (*reading plus*), the authors highlight the importance of vocabulary tasks in fostering elaboration with this vocabulary as well as strengthening different types of knowledge of these words. For Bingham Wesche and Paribakht (2000), this is facilitated by the fact that students encounter words several times and in multiple functions; thus, enrichment through the elaboration of the knowledge of the concept they are learning takes place and students learn the meaning and the form of the words. Also, these authors argue that if students do not get the meaning of a word in one exercise they have the chance to get it in another exercise. They insist that the repeated and varied mental processing that takes place when students perform different tasks and see the word in different environments allows for automatization and association of a given concept with different contexts, and that this does not necessarily happen if students only read a text.

Bingham Wesche and Paribakht (2000), by comparing the reading only with the reading plus groups, found that the difference in the outcome of learned words in the reading plus treatment exhibited a number of advantages. First, the reading plus tasks augmented the saliency of words. Second, learners were further encouraged to explore word uses on their own. Third, students were motivated to work more with the words and this higher mental involvement was conducive to a higher chance of learning the words. Fourth, completing exercises often requires students to re-read the text many times, and may require more in-depth sentence processing that leads to discourse processing. This additional activity resulted in better comprehension of the reading and in other benefits from the tasks. Fifth, students profited from the additional practice that involved dealing with the words in different contexts. Finally, learners had an awareness of the type of knowledge they had developed about the concepts they were using, albeit precise or imprecise knowledge, and they had the possibility of adjusting this knowledge.

In general, Bingham Wesche and Paribakht (2000) state that their study serves as further evidence for the idea that multiple encounters with vocabulary are necessary to learn new concepts. They insist that the reading plus condition goes beyond creating a form-meaning connection and develops several aspects of word

knowledge and use. For them, students are aware of the type of knowledge they have gained and seem to be ready to use this knowledge in production. The authors conclude that the summative and diverse nature of exercises results in retention that is foreseeable and more effective. They argue that these guided tasks help students concentrate their attention on specific concepts, comprehend general and specific information, and recognize different word forms, functions and meanings as well as the connections that are established among them. In sum, we believe that this study serves as further evidence that students need additional guidance to acquire and use new concepts effectively. Alone, reading is not as beneficial as reading supplemented by exercises designed to promote students' awareness and deep processing of concepts that will eventually lead to a productive use of the language.

Along these lines, Laufer (2001) explores how much reading comprehension can be enriched if this task is complemented with additional tasks. Her study challenges the idea of words being acquired through exposure to (reading) input alone. She insists that if only small amounts of reading take place, then word-focused activities are required to ensure a growth in vocabulary. In this analysis, Laufer reports on a number of studies that compared reading with isolated activities such as sentence writing, fill-in tasks, composition writing, and in some cases reading combined with these activities as well. In all cases Laufer has found that sole reading results were surpassed by results that involved some additional kind of word-focused activity. She concludes that the most beneficial activities to support vocabulary learning in combination with reading include dictionary use, fill-in tasks, and use of the words in sentence and composition production. Finally, she states that, if we expect our learners to acquire a word without trying to memorize it, then, word-focused activities are a viable option that warrants vocabulary learning. She insists that reading alone does not represent the most effective source of vocabulary acquisition. Again, we find that studies such as this one should serve as a strong indication for adequate practices in SLA classrooms where reading is seen as an autonomous activity. SL practitioners should make sure that they provide students with enough additional activities so that learners can reap the highest rewards out of reading practices in terms of vocabulary achievement.

In a different vein, Hunt and Beglar (2005) contribute information through the description of a reading framework. They put forward a framework for FL instructors and administrators with the intention of guiding learners to a faster lexical growth. The key for the successful application of this framework, the authors insist, is to establish a "pedagogically sound balance between explicit and implicit activities for L2 learners at all levels of their development" (p. 23). Hunt

and Beglar (2005) argue that lexical objectives should always be pursued, along with an increase in vocabulary breadth, a more consolidated, elaborated vocabulary knowledge and greater fluency with familiar vocabulary. For them, the previous objectives and the appropriate assessment techniques would result in a successful vocabulary and foreign language program. While they insist that extensive reading is a key to developing high levels of reading vocabulary, they acknowledge that the features described in the framework would also aid learners to achieve high vocabulary proficiency at a much faster rate.

Their framework consists of two important pillars. On the one hand, it promotes explicit lexical instruction and learning strategies that include learning of vocabulary out of context (which is subsequently contextualized in meaning-based exercises and readings) and using dictionaries to learn new concepts and inferring meaning from context. The authors argue that these elements combined with the right amount of reading result in great vocabulary development for learners. The second pillar is related to the encouragement of implicit lexical instruction and learning strategies, among which they mention task-sets linked to readings and narrow as well as extensive reading. Reading should be present in large quantities for it to be effective. Only repeated lexical encounters would ensure actual learning of vocabulary; thus, students need to read both frequently and extensively. Hunt and Beglar (2005) add that the application of metacognitive and cognitive strategies, such as monitoring and self-evaluation, helps learners plan how they approach classroom tasks and it helps them monitor their understanding and use of language and vocabulary in particular. The use of strategies gives learners more independence and responsibility in their learning process. According to Hunt and Beglar (2005), a detailed balanced combination of explicit and implicit vocabulary instruction that implements the correct strategies can guarantee fast lexical development. Their work provides additional information on adequate techniques that can be implemented in SLA classrooms.

In another study, Pigada and Schmitt (2006) looked at different aspects of vocabulary enhancement that can be gained through reading. They addressed the effect of incidental learning through extensive reading on depth of vocabulary knowledge and found different evidence. To determine the number of encounters necessary for these words or for their information to be retained, they investigated three aspects: spelling, meaning and grammatical knowledge. The treatment consisted of the participant (an adult, single individual) selecting and taking one month to read four short books that represented around 30,000 words (228 pages). The target words were encountered at different rates from 1 to 19 times across the

four readers. These authors found that knowledge of 65% of the words in the study improved to some extent during the treatment process. As for the aspects of depth of knowledge, spelling turned out to be strongly enhanced for all nouns in the study and almost for all verbs. Improvement in spelling was observed for some of the words even after a single exposure.

Meaning was also enhanced but not to the same extent as spelling knowledge was. There is no clear number of exposures to identify a steady improvement in levels of meaning. Even after 20+ exposures the participant was unable to recall the meaning of some of the target words. Regarding grammatical knowledge, all frequency groups showed improvement. In particular, the knowledge of articles that should accompany nouns was strongly enhanced, whereas the knowledge about prepositions that should accompany verbs was severely limited. The authors insist that two thirds of the words in the study showed improvement in at least one of the word knowledge aspects. They emphasize that the effectiveness of this knowledge is greater if one considers that the participants were evaluated through productive tasks, although he had only receptive input. They conclude that over 20 encounters are required to ensure a good chance for all three types of word knowledge to be enhanced through extensive reading. Pigada and Schmitt argue that extensive reading that is part of an organized reading program can lead to sustained vocabulary acquisition. For them, spelling is the aspect that benefits the most from this type of input, while the other aspects of depth would require further specific attention. This study serves as an indication of the need for multiple encounters for the students to actually profit from reading in terms of vocabulary learning.

7.1.2. Vocabulary threshold for reading

Laufer (1992) establishes the minimal vocabulary threshold for reading comprehension at 3000 word families. For Laufer (1996), the vocabulary threshold is identified as the minimal number of words that readers must be able to recognize automatically, regardless of the context, when they are facing a text. For Laufer (1992), once L2 learners have achieved this 3000-word threshold, they are able to make use of the reading strategies that they use in their L1. If the 3000-word threshold is not reached, however, L1 strategies for reading comprehension would not be effective due to lack of vocabulary, contends Laufer. Laufer (1996) notes that this 3000 word family threshold contains approximately 4,800 lexical items.

She further insists that 3000 word families account for a coverage of 90-95% of any text, which she maintains is sufficient for *reasonable* text comprehension. Laufer (1996) concludes that if instructors and administrators want to improve students' reading proficiency levels, they must work toward providing students with the tools to enhance vocabulary breadth; reading comprehension will follow.

In a subsequent consideration, Laufer and Ravenhorst-Kalovski (2010) reanalyze the reading threshold and discuss the effect that the researcher's interpretation has on what is *adequate* reading comprehension. This time, the authors propose two different thresholds that describe different scenarios. An "optimal threshold" (of between 6000 and 8000 word families) with a 98% text-coverage, represents adequate, independent reading where no additional help is provided to the reader. The other threshold is a "minimal threshold" (of between 4000 and 5000 word families) which offers between 93.4% to 95.5% text coverage, describes adequate reading skills in which some external guidance and aid could be given to the reader. The authors mention yet another threshold of around 3000 words which allows a text-coverage of between 90.56% and 92.66% for first year university students who will be later trained to improve these levels. This case, they note, does not represent an ideal scenario. Although they put forward different threshold possibilities, the minimum (not the best) option stands at 3000 word families. As Nation (1983, 2001, 2013) has pointed out, to reach the 3000 word level students must know the high-frequency 2000 word level band. This continues to be the foundation on which the rest of vocabulary would develop. While students can apply compensating strategies to make up for lack of vocabulary, these are not always effective, and vocabulary ends up being the determinant factor for reading comprehension.

Along these lines, Nation (2006c) offers an extensive analysis of the number of words necessary for adequate reading and listening comprehension. In agreement with Laufer and Ravenhorst-Kalovski (2010), Nation argues that we need to know at least 98% of the words in a text for unaided reading comprehension. A reader would require a word-family vocabulary of between 8000 and 9000 to comprehend a written text without the help of a dictionary, a teacher or any other source of external help. Nation calculated these figures based on the vocabulary needed to read unsimplified novels whose audience is usually teenage, native speakers of English. When this is applied to a novel, Nation argues that knowing 9000 words plus the proper nouns appearing in the novel, readers would be able to identify 98.22% of the tokens in the text and only one unknown word in every fifty running words would elude the reader. In the same study, Nation

analyzes how many words would be needed for newspaper articles and concludes that the same numbers apply. Nation (2006c) insists that 98% is taken as the required vocabulary knowledge assuming that no external help is sought or received during the reading process. If external help were available, the requirements for readers would probably be a little lower. He warns, though, that even for a 95% text coverage, readers would encounter one unknown word in every two lines; that is, one unknown word in every twenty words.

In earlier studies, Hirsh and Nation (1992) had suggested a vocabulary of at least 5000 word families to read short novels. They showed that 2000 words were not enough to provide vocabulary coverage for pleasurable reading, claiming that 5000 words would suffice for the readers to enjoy the novels without struggling to understand the content. Hu and Nation (2000) estimated a 98% vocabulary coverage as the required measure for comprehension without additional aid in texts such as fiction texts (which they do not consider highly demanding). They found that an 80% text coverage did not result in any gain in reading comprehension in their study and that a 90-95% vocabulary text coverage results in comprehension for only some of the participants in their study. It is clear that this threshold is supported by new evidence brought forward in Nation (2006c).

Analysis of this information leads us to the realization of how significant this lexical threshold is and how much a lack of vocabulary may hinder successful reading comprehension. Vocabulary knowledge represents a heavy load in reading comprehension and in assessment through reading. In general, it seems that for learners to enjoy the flow of a text without hindrance resulting from lack of vocabulary they need to have reached a threshold of 98% of vocabulary coverage in the text. The following section will provide an analysis of the reading comprehension of students as measured by a reading test, where we can examine how well they use word knowledge mostly at the 2000 word level when these words appear in the context of short texts.

7.2. THIS STUDY: READING TEST RESULTS

7.2.1. Participants

A total of 185 participants performed the reading task. From the total number of students that participated in the complete study (186), in this particular

task there were 131 students from the FLS and 54 students from the CBT school. One student from the CB school did not attend class the day that the reading test was administered. In terms of gender groups, there are 21 male students and 33 female students in the CBS, and 64 male and 67 female participants in the FLS.

7.2.2. Instrument

With this reading test, we attempt to measure the receptive vocabulary knowledge of students when vocabulary that belongs mainly to the 1000 and 2000 word bands is immersed in the context of short reading passages. For this purpose, the paper and pencil Cambridge ESOL Entry Level Certificate in ESOL Skills for Life (Entry 3)/Reading (Corporate Author Cambridge ESOL, UCLES 2011) was used. This is a subset of the standardized language level tests used worldwide in studies determining language level. The reading test consists of 5 texts and 32 questions. There are 22 multiple-choice questions where students must check (√) one answer, 9 questions where students have to write ONE-WORD answers, and 1 question where students write a ONE-SENTENCE answer. Out of these 22 multiple-choice questions, 3 questions contain sub-questions. That is, from the multiple-choice questions, questions number 3 and 16 contained 4 items each, and question 28 contained three items. That is, in total there were 30 multiple-choice items. From the short-answer questions, question 21 contained three items. In total, short-answer questions add up to 9 questions and 11 items. Finally, one question (question 29) requires a ONE-SENTENCE answer and contains 3 items. In total there are 32 questions and 44 items.

Each part of the test also indicates the approximate time that students should spend on that part. Part 1 suggests 20 minutes and includes the first two texts (A and B). Part 2 is allotted 25 minutes and includes Texts C and D. Part 3 is allocated 30 minutes and includes Text E. All of the items are equally weighted and are worth 1 point each, except for the items in question 29, in which case each item is worth 2 points. That is, the test is worth 47 points in total. The students' overall scores for the test are then calculated on a scale from 0 to 100. The total number of points that the student obtained is added up and the result is multiplied by 100 (maximum score) and divided by 47 (maximum points). The students' overall score is the resulting number of this operation rounded up to the nearest whole number. A sample of the reading test can be found in Appendix 2.

Before the test was administered, all the texts in the test were checked to determine the word level band of the words used in them, to ensure that the content of the texts was within the reach of students' knowledge. Table 19 depicts the word type distribution for the five different texts in the reading test as well as the total number of words that appeared in each text. Table 19 shows that between 80 and 89 per cent of the words in the text fits within the 2000 word level band. Between 2 and 6 per cent of the words belong to the AWL and the remaining percentage belongs to words that are not part of these lists. We must keep in mind that words labeled *Not-on-the-list* include proper names (that is mainly the case in these texts) and words that are part of word lists above the 2000 word level.

Table 19. Percentage of word types and total words in each text in the Reading Test

Texts	Type-1000	Type-2000	AWL	Not on the list	Total words
Text 1	78.70%	10.19%	3.70%	7.41%	176
Text 2	78.70%	9.26%	5.56%	6.48%	181
Text 3	72.44%	9.45%	2.36%	15.75%	192
Text 4	73.64%	6.98%	6.20%	13.18%	187
Text 5	79.50%	8.50%	4%	8%	440

To illustrate the elements that are outside the 2000 word level we proceed to analyze the case of Text 3. In this case, the words labeled *Not-on-the-list* include proper names such as *Dave*, *Jasna and Khan, Ozgur* or others such as *DVDs*, *CDs*, *café*, *superstore*, *headache* or *magazine*; all of these are highly likely understood by students, as they are either proper names, transparent or highly frequent words in everyday use. As for the AWL, the three words that are not in the 2000 word band are *assist*, *contrast* and *relax*. Once again, there is a high probability that students are familiar with these elements. The texts are, thus, deemed appropriate for this population. Table 19. shows how many total words appeared in each text. Four of the texts were fairly short, under 200 words, whereas the last text was a little longer, with a total word count of 440.

Based on this analysis we consider that the content of the texts in the test seems to be within the grasp of the students' reading ability. We must recall that reading is the skill that receives more attention in the Costa Rican curriculum and that students start developing reading skills since the beginning stages of their high school English instruction. With these premises in mind, we proceed to the phase of the study that examines students' reading comprehension skills.

7.2.3. Procedure

For this task, the students had 1 hour and 15 minutes to complete the test. Most of the students in both schools finished in less than an hour. At the beginning of the session, instructions were given in Spanish to ensure all students knew exactly what was expected of them. Written instructions were provided in English, for each of the different sections of the test.

The heading of the test was slightly adapted to a format that would result more familiar to students in the Costa Rican setting. Key information (i.e., name of the test, author of the test, student information, general guidelines, and time allotted) are all given in the heading. Throughout the test, spelling and a few words were changed to comply with the American English system and its vocabulary, which is the variety of English that Costa Rican students are learning and are more familiar with. These changes are summarized as follows. In Text A, we labeled each of the paragraphs to guide students because question 5 asks them to go to paragraph 1. Word spelling for "organised" was changed to "organized." Also, throughout the text, the word "tick" was changed to "check"; and the symbol (✓) was always included. In Text B, we changed the acronym "ESOL" to "English learners," because students may not be familiar with the technical concept indicated by ESOL. In Text C, *UK* was changed to *US*, and "loathe" was changed to "hate," again, to provide students with more familiar vocabulary. In Text D, the following words were changed: "crèches" to "child care centers," "colour" to "color," and "crèche" to "child care." In Text E, "mad" was changed to "crazy." UK was not changed in this text because the context mentions other countries, so it was easier for students to create a parallel between UK and other countries.

7.2.4. Results

7.2.4.1. Item analysis

In general terms, the behavior of the results, as displayed in Tables 20 and 21 for each one of the items in the test, seems to be in accordance with the overall means obtained in this same test by each school. These results will be discussed in detail below.

Table 20. Percentages per response for multiple-choice items in the Reading Test

	Content Based Teaching N = 54			**Foreign Language Teaching N = 131**		
	Correct	**Incorrect**	**Missing**	**Correct**	**Incorrect**	**Missing**
Item 1	85.5%	12.7%	1.8%	68.7%	31.3%	--
Item 2	76.4%	21.8%	1.8%	64.1%	35.9%	--
Item 3.1	96.4%	1.8%	1.8%	84.0%	16.0%	--
Item 3.2	89.1%	9.1%	1.8%	88.5%	11.5%	--
Item 3.3	83.6%	14.5%	1.8%	63.4%	36.6%	--
Item 3.4	81,8%	14.5%	3.6%	68.7%	31.3%	--
Item 6	98.2%	0%	1.8%	95.4%	4.6%	--
Item 7	67.3%	30.9%	1.8%	66.4%	33.6%	--
Item 8	14.5%	83.6%	1.8%	19.1%	80.9%	--
Item 12	83.6%	14.5%	1.8%	79.4%	20.6%	--
Item 13	94.5%	3.6%	1.8%	88.5%	11.5%	--
Item 14	96.4%	1.8%	1.8%	91.6%	8.4%	--
Item 15	83.6%	14.5%	1.8%	65.6%	34.4%	--
Item 16.1	96.4%	1.8%	1.8%	82.4%	17.6%	--
Item 16.2	61.8%	36.4%	1.8%	52.7%	47.3%	--
Item 16.3	94.5%	3.6%	1.8%	90.1%	9.9%	--
Item 16.4	78.2%	20.0%	1.8%	83.2%	16.8%	--
Item 17	96.4%	1.8%	1.8%	91.6%	8.4%	--
Item 18	94.5%	3.6%	1.8%	90.8%	9.2%	--
Item 19	89.1%	9.1%	1.8%	67.9%	32.1%	--
Item 20	72.7%	25.5%	1.8%	63.4%	36.6%	--
Item 22	90.1%	7.3%	1.8%	87.0%	13.0%	--
Item 23	81.8%	16.4%	1.8%	62.6%	37.4%	--
Item 24	92.7%	5.5%	1.8%	79.4%	20.6%	--
Item 25	87.3%	10.9%	1.8%	65.6%	34.4%	--
Item 26	96.4%	1.8%	1.8%	92.4%	7.6%	--
Item 27	94.5%	3.6%	1.8%	74.8%	25.2%	--
Item 28.1	89.1%	9.1%	1.8%	78.6%	21.4%	--
Item 28.2	85.5%	12.7%	1.8%	84.0%	15.2%	.8%
Item 28.3	56.4%	41.8%	1.8%	53.4%	45.8%	.8%

To illustrate the behavior of the results, let's consider that the mean score for the CB school was 86.13 and for the FL school it was 74.18. We see that this same difference prevails across many of the items. That is, in many of the cases, the CB school maintains an advantage that is around +/– 12 points over the FL school. Throughout the test, we note that the items that were easier or more difficult for students in the CB school were also easier or more difficult for students in the FL school. In items such as 6, 14, 16.c, 17, 18, 26, 29.b and 29.c, students from both schools obtained higher than average results; over 90–95% correct responses. In items such as item 2, 8, 9, 11, 20, 28.c, both schools scored percentages 10-20% or more below their means. Yet in the items 3.b, 5, and 7, the correct percentage

for both schools is almost the same. There are only 5 items (8, 16.d, 21.a, 21.b, 21.c) for which the FL school achieved results that are higher, though only slightly, than the ones attained by the CB school. The last three items required short-sentence answers.

Item 8 is a special case to discuss here in more depth; 83.6% of the CBS students and 80.9% of the FLS students chose the wrong answer. This question does indeed yield lower scores. During the analysis of the results, a problem was identified in the transcription of this item. This explains the drop in correct answers in this item in comparison to the others. To determine whether this problem affected the overall consistency of the test, a reliability test was conducted. An item analysis revealed that this item does not affect the overall reliability and internal consistency of the test, as this is not improved considerably if the item is deleted. The Cronbach's alpha reliability changes from .716 to .720 in the CB school and from .887 to .891 for the FL school; the latter figures do not include item 8. If the alphas were rounded to the nearest whole number, the figures would not change; that is .72 for the CBS and .89 for the FLS. Item 8 does not affect the overall reliability of the test substantially and it was, thus, kept throughout the analysis.

Table 21. Percentages per response for production items in the Reading Test

	Content Based Teaching N = 54			**Foreign Language Teaching N = 131**		
	Correct	**Incorrect**	**Missing**	**Correct**	**Incorrect**	**Missing**
Item 4	60%	38.2%	1.8%	43.5%	56.5%	--
Item 5	80.0%	18.2%	1.8%	80.9%	19.1%	--
Item 9	72.7%	25.5%	1.8%	54.2%	45.8%	--
Item 10	83.6%	14.6%	1.8%	63.4%	36.6%	--
Item 11	56.4%	41.8%	1.8%	40.5%	59.5%	--
Item 21.1	85.5%	12.7%	1.8%	86.3%	13.6%	--
Item 21.2	74.5%	23.7%	1.8%	77.1%	22.9%	--
Item 21.3	78.2%	20.0%	1.8%	80.2%	19.8%	--
Item 29.1	96.4%	1.8%	1.8%	80.9%	19.1%	--
Item 29.2	98.2%	0%	1.8%	91.6%	.4%	--
Item 29.3	98.2%	0%	1.8%	90.1%	9.9%	--
Item 30	80.0%	18.2%	1.8%	69.5%	30.5%	--
Item 31	87.3%	10.9%	1.8%	80.2%	19.8%	--
Item 32	90.9%	7.3%	1.8%	54.2%	45.8%	--

As mentioned above, in general, the results per item of the test present a behavior that is in consonance with the overall results in this and the other tests that students took along this study. Table 20 contains the information in terms of correct, incorrect and missing percentages for each one of the multiple-choice items. Table 21 contains the percentage of correct, incorrect and missing data for each one of the items that demanded students to write ONE-WORD or ONE-SENTENCE answers.

One student from the CB was absent from class on the day students were given this test; that is reflected in the portion for the missing data for that school. In the case of the missing data for the FL school, that reflects the cases for which one student did not answer two of the questions. As mentioned above, the mean scores are 86.13 and 74.13 for the CB and FL school respectively.

7.2.4.2. Test score analysis

Table 22 contains the descriptive statistics for this test. The mean for the CB school is higher (86.13) than the mean for the FL school (74.18).

Table 22. Scores for the Reading Test in each school

Reading scores (N= 185)		
	CB School	**FL School**
Score	100	100
Mean	86.13	74.18
Median	87	79
Max scores	98	98
Min scores	57	23
SD	8.564	17.327

While students in both schools reached a maximum grade of 98 out of 100, the minimum grades behave differently. These, however, are still higher for the CB school (57) when compared with the minimum scores of the FLS (23). The spread of distribution for the FL school is twice as wide (SD 17.327) than the one for the CB school (SD 8.564) where grades are clearly more concentrated. This preliminary information about results gives advantage to the CBS over the FLS. This could be expected given the results analyzed in the previous section for the VLT.

Figure 10 presents a visual representation of the distribution of the results for the reading test. We can observe that the CBS scores are clustered mainly over the 80-point mark. Over 20 percent of school's scores are grouped close to the 90-point mark and another large group is well over that mark. In the case of the FLS, the distribution of scores is spread to a considerable extent below the 80-point mark, as low as the 20-point mark. This spread of course has an effect on the mean for this school. The majority of the distribution, however, appears concentrated between the 60-point mark and the 100-mark with the higher concentration in this portion of the scores between 80 and 90, and then it slightly drops again. A quick glance at the distribution of these figures already anticipates that the data does not follow a normal distribution in either school.

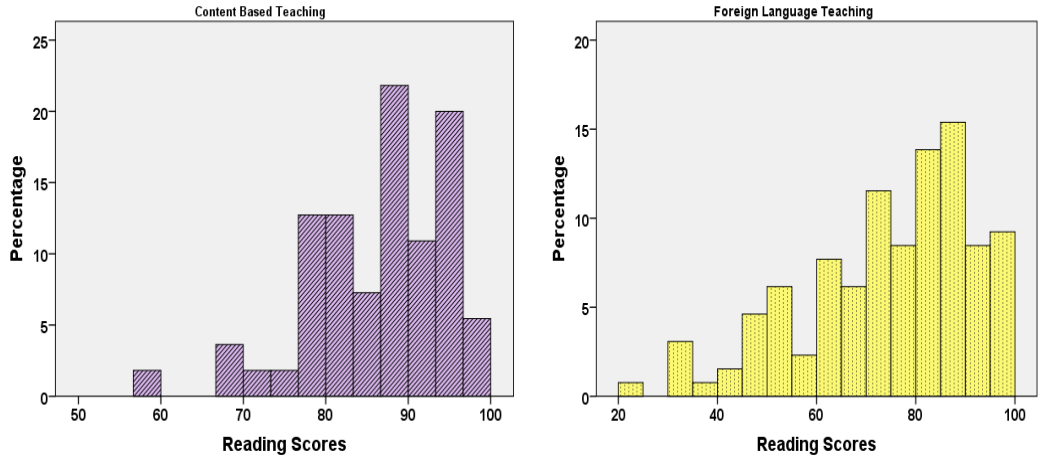


Figure 10. Score distribution in the Reading Test per school.

Additional information is provided by Figure 11, which illustrates the score distribution for each school using a box plot. To begin, in the CBS the distribution seems to be negatively skewed. While the distribution is slightly asymmetrical inside the box, the whiskers show a clear asymmetrical distribution with a great dispersion toward the lower part of the scale. As mentioned above, the CBS grades are concentrated around 60 and close to 100.

One outlier, though, falls outside this spread range. For the FL school, the distribution is clearly asymmetrical both inside the box and for the whiskers. The data is negatively skewed as well, also occupying a much wider spread of

distribution in direction of the lower values of the scale. An outlier is also spotted well below 40, close to 20 on the scale.

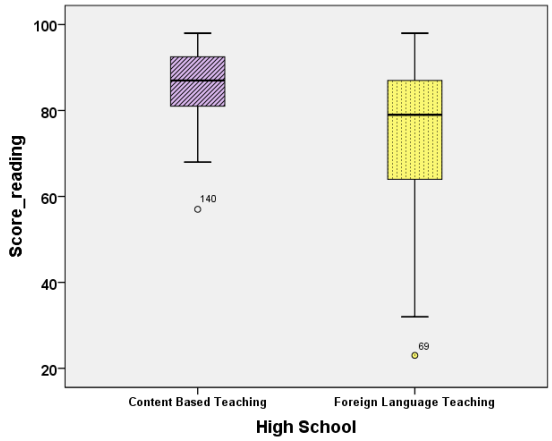


Figure 11. Box plot of score distribution per school.

The Kolmogorov-Smirnov test was used to determine normality assumptions in our data, with the purpose of defining the nature of the distribution between the results of both schools in the Reading Test.

Table 23. Parametric test for Reading Test

	High School	**Kolmogorov-Smirnov**		
			gl	Sig.
Reading Test	CB School	.122	55	.039
	FL School	.122	130	.000

The results from the parametric test are presented in Table 23. According to this test, as we anticipated from Figure 10, the data follow a not-normal distribution which calls for non-parametric tests of means comparison for two independent variables. The Mann-Whitney U test was, therefore, conducted to measure inferential statistical differences between the groups. As seen in Table 24, this test reveals statistically significant differences favoring the CB school, which shows a median rank of 121.19 over 81.38 for the FL school.

Table 24. Mann-Whitney U test results for the Reading Test

N = 185	**Reading**
Mann-Whitney U	2,014.500
Wilcoxon W	10,660.500
Error	330.626
Z	- 4.605
P (two-tailed)	.000

As briefly mentioned above, to estimate the internal consistency associated with the scores in the test, a Cronbach's alpha test was calculated for each school. For both schools, Cronbach's alpha shows that the test is internally consistent. Cronbach's alpha assigns the CB school an alpha of .72 and for the FL school the test determines a variance of $\alpha = .89$. In both schools the variance can be considered sufficiently reliable.

7.2.4.3. Gender results analysis

The test scores were also compared using the variable gender, first at the intra-school level and after that at the inter-school level. Table 25 presents the results for gender scores in each institution. At the CBS, boys exhibit a higher mean at 88.81, and girls at 84.58. In comparison, the girls' scores are slightly higher (74.66) than the boys' scores (73.78) at the FL school. Both gender groups have similar maximum scores while the FLS shows clearly lower minimum scores. The spread of distribution is also evidently larger for the latter school, with a spread that is more than double the one in the CB school for boys and almost twice as large in the case of girls. Figure 12. presents the distribution of scores based on gender.

Table 25. Descriptive statistics based on gender in both schools for the Reading Test

	CB school		**FL school**	
	Boys	**Girls**	**Boys**	**Girls**
N	21	33	64	67
Score	100	100	100	100
Mean	88.81	84.58	73.78	74.66
Median	89	87	78	79
Max	98	98	98	96
Min	68	57	23	32
SD	7.181	9.138	18.267	16.441

As depicted in the numerical description in Table 25 and in Figure 12, in the CBS the boys' scores are clustered in the higher levels of the scale. One outlier' score falls short below the 70-mark on the scale. The data for boys seems to be more dispersed toward the lower levels while showing a positive tendency. The whiskers also evidence the asymmetrical distribution of the data with one of the whiskers clearly more disperse toward the lower values of the scale, thus allowing greater concentration in the upper levels. For the female group, the scores are more widely distributed although most of them concentrate above the 70-mark. This distribution is asymmetrical both inside the box and to a lower degree in the whiskers of that box plot. The data in the box shows a more negatively skewed tendency. One outlier is also present in this data at a point that falls below the 60-mark on the scale.

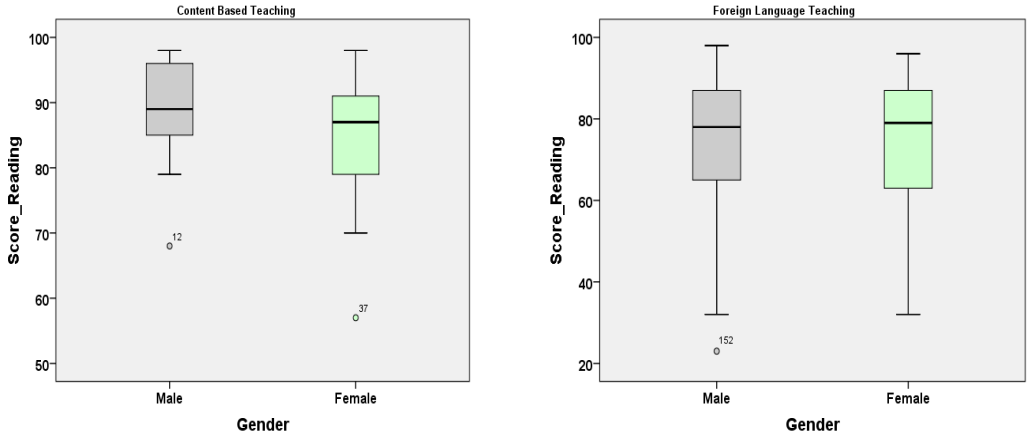


Figure 12. Score distribution based on the gender variable.

The distribution for the FL school is clearly different from that of the CB school, but it has more similarities between the gender groups for this population. Both ends of the whiskers in this representation are placed at very similar end points for the genders, indicating some similarity in terms of the general spread of distribution of the data. For each gender group the distribution seems to be negatively skewed and is clearly asymmetrical. One outlier is present in the male group, a little above the 20-mark on the scale.

Normality assumptions for distribution were checked for both gender groups in both schools. For this purpose, Table 26 shows the results of a Kolmogorov-Smirnov test. According to this test, while gender groups in the CBS

follow a normal distribution, the FLS data for gender groups follow a not-normal distribution, as predicted by the box distribution in Figure 12.

Table 26. Normality distribution test for gender and Reading Test scores in each school

School	Gender	Kolmogorov-Smirnov	gl	Sig.
Content Based	Masculine	.155	21	.200
	Feminine	.120	33	.200
Foreign Language	Masculine	.115	64	.035
	Feminine	.143	67	.002

Further tests were performed to determine the statistical significance of the differences found so far. A *t*-test was applied to the CB data to determine the nature of the differences between gender groups in this school. Since the CBS data show that genders groups follow a normal distribution, the Independent sample *t*-test is used to determine the nature of those differences. Assuming equal variances, the *t*-test shows non-significant statistical differences between gender groups. The analysis follows.

The male group ($N = 21$) was described as having a superior numerical mean of 88.81 (*SD* 7.181). By comparison the female group ($N = 33$) showed a lower numerical test score with a mean of 84.58 (*SD* 9.138). As mentioned above, to test the hypothesis that the male and female students actually showed statistically significant different test scores in Reading, a *t*-test was performed. The *t*-test results are presented in Table 27.

Table 27. Use of a *t*-test to determine gender score differences for the Reading Test in CBS (N = 54)

		Levene's test		*t*-test for equality of means		
		F	Sig.	*t*	df	Sig. (2-tailed)
Score on Reading Test	Equal variances assumed	.1777	.188	1.797	52	.078

The Kolmogorov-Smirnov test already indicated that the data follow a normal distribution. Also, the male and female distributions were sufficiently normal for a *t*-test to be performed; statistical descriptors indicate that their skew is

less than 2.0 and their kurtosis is less than 9.0. In addition, Table 27 shows how homogeneity assumptions were tested and satisfied via Levene's *F* test. Equal variances were assumed in this test $F(52) = .177$, $p = .188$; where *p* is greater than .05. The independent *t*-test is thus associated with a statistically non-significant effect, $t(52) = 1.797$ $p = .078$. This means that the male and female test scores for reading are thus an indication of a non-significant statistical result for gender groups in the CB school setting.

To determine gender differences at the FLS and given the not-normal results provided by the Kolmogorov-Smirnov test, a Mann-Whitney U test was performed. According to this test, the gender variable shows non-significant differences for the gender variable in the FL school. See Table 28, for statistical results.

Table 28. Mann-Whitney U parametric test for gender differences in FLS in Reading Test.

N = 131	**Reading**
Mann-Whitney U	2 153. 500
Wilcoxon W	4 431.500
Error	216.936
Z	.044
P (two-tailed)	.965

As seen in Table 28, this test reveals that the slight numerical superiority of female students with a mean of 74.66 (median rank of 66.14) when compared to the boys' mean of 73.78 (median rank of 65.85) is actually not statistically significant for the gender variable in this school and the distribution for the variable score is assumed to be the same regardless of gender differences.

Further tests in connection to gender were implemented at the inter-school level. Homogeneity assumptions were not satisfied to perform an ANOVA, nor were normality assumptions met. A non-parametric Kruskal-Wallis test was then implemented as a viable choice. This test, however, allows for the analysis of two factors; given that there were two high schools and two genders in each high school, the creation of a new variable containing both features was necessary to perform the test. The original variables were combined to create four new variables, comprising gender and school, as shown in Table 29.

Table 29. Combination of variables to perform Kruskal-Wallis.

	Original Variables	New Variable
Gender	1 Masculine	1. Masculine-CB
	2 Feminine	2. Feminine-CB
High School	1 Content Based (CB)	3. Masculine-FL
	2 Foreign Language (FL)	4. Feminine-FL

In this case, we have one independent variable which is categorical and combines the variable of gender and school, and one dependent variable that is ordinal and represents the scores in the reading task. Considering that the shapes of the distributions are not equal, we test the hypothesis that there are no differences across the median ranks of schools in connection to the scores obtained in the Reading Test. As mentioned above, to test that hypothesis, a Kruskal-Wallis test was implemented. The comparison of gender and school results shows that there are highly statistically significant differences (considering $\alpha = 0.05$) between gender at interschool level when students from different genders are compared but not across schools when students of the same gender are compared for the scores in the Reading Test (α greater than 0.05).

The null hypothesis of equal median rank distribution in Reading Test scores for gender and schools is thus rejected on the basis of cross-gender comparisons. If we compare girls with boys' results, the results favor masculine (median rank 134.24) and feminine students (median rank 81.22) in the CBS in comparison to feminine students (median rank 81.22) and masculine (median rank 114.62) in the FLS. Same gender comparisons, however, yield no statistically significant results. Table 30 shows the descriptive statistics of the Kruskal-Wallis test and Figure 13. depicts the distribution of mean ranks for gender and school.

Table 30. Krustal-Wallis test to analyze Reading Test and gender between samples (α= 0.05)

Sample1-Sample2	Stat Test	Typical Error	Statistical Dev.	Sig.	Sig. (2-tail)
4.Feminine-FL/ 1.Masculine-CB	54.071	13.396	4.036	.000	.000
2.Feminine-CB/ 1.Masculine-CB	53.019	14.446	3.943	.000	.000
4.Feminine-FL/ 3.Masculine-FL	34.451	11.287	3.052	.002	.014
2.Feminine-CB/ 3.Masculine-FL	-33.399	11.347	- 2.943	.003	.019
3.Masculine-FL/ 1.Masculine-CB	19.620	14.840	1.322	.186	1.000
4.Feminine-FL/ 2.Feminine-CB	1.052	9.380	.112	.911	1.000

In Figure 13, we can easily identify the comparisons for which statistically significant differences are identified and which are connected by yellow lines in the graph. These comparisons represent variables involving students of different gender groups (contrasts Feminine-FL/Masculine-CB, Feminine-CB/Masculine-CB, Feminine-FL/Masculine-FL, Feminine-CB/Masculine-FL are all checked for statistically significant differences).

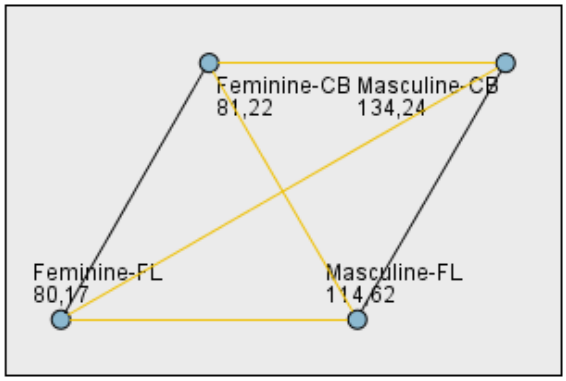


Figure 13. Kruskal-Wallis test for Reading scores for gender and school.

As mentioned above, same gender comparisons (connected by black lines) show no statistically significant differences. That is, when male students in the FLS are compared to male students in the CBS (3.00-1.00) and female students in FLS are compared to female students in the CBS (4.00-2.00), the statistically significant difference disappears. In these cases, it seems that the school, rather than the gender variable, appears to have had an effect on the students' scores. Gender alone has not influenced the scores obtained on the Reading Test.

7.2.5. Discussion

In interpreting the results different considerations should be made. To begin, strictly adhering ourselves to the objectives of the curriculum we could argue that these are partly met. If we assume that one of the tacit objectives of the curriculum is to prepare students to pass the English reading test that they have to take at the end of high school, we could conclude that most students do so. The passing grade

for the English test administered by the Costa Rican Board of Education is 70. Both schools seem to achieve that goal with means above that requirement. The mean at the CB school is 86.13 and that of the FLT school is 74.18. While students in the CB school seem to obtain the necessary score with more ease, FL school students also do so, though to a more limited extent. We should take this information, however, as an indication that a fair portion of the students in the FL (and a smaller portion in the CBS) school may risk not reaching that goal.

On the other hand, there is one more aspect to consider. The final overall passing grade for high school in the Costa Rican system is made up of two parts. There is a 60% that depends on this exam only, and a 40% that represents a cumulative percentage derived from the students' performance in that subject in their last two years of high school. So, even if students obtained a score that is a little below the mean in the present study in that school, they still have a fairly good chance that the overall grade would be enough to pass their final school year and then move on to college. Taken together, this information seems to indicate that; in general terms, the population represented in this study has a very good chance of passing the compulsory exam, an implicit objective of the curriculum according to our interpretation.

Here we could look back at the discussion that Laufer (1992, 1996) and Laufer and Ravenhorst-Kalovski (2010) maintain regarding what an "adequate" level of reading comprehension is and how it changes depending on the researchers and the context. On the one hand, considering the compulsory test at the end of high school as a reference and based on the previous results, from the perspective of teachers, parents, students and maybe even the Board of Education, the reading comprehension of these students is probably deemed *adequate*. On the other hand, thinking about the notions of *independent* on *unassisted* reading comprehension, then we must face the fact that these students are not anywhere near these possibilities; their reading comprehension is, thus, still *inadequate*. Assuming the scores obtained in the test as a sign of overall reading comprehension, a high percentage of the content of the reading still eludes the students' understanding. In the CB school, almost 14% of the information was missed while in the FL school that percentage reaches over 25%. Both figures represent a large proportion of content that was not handled by the participants in this study.

As discussed in section 6.3.4, the vocabulary size of the participants in this study is far below the minimal vocabulary threshold established for reading comprehension. The mean for the CBS stands at 1,793.94 words, while for the FLS it was established at 1,475.85. Laufer (1992, 1996) established 3000 words; Laufer

and Ravenhorst-Kalovski (2010) suggest an optimal threshold of 6000 to 8000 words families and a minimal threshold of 4000 to 5000 thousand; Nation (2006c) speaks of 8000 to 9000 for unassisted reading comprehension and Hirsh and Nation (1992) mentioned 5000 words to read short novels. Regardless of the threshold and the genre in these previous studies, it is evident that the students in the present study have a more limited command of the vocabulary required to excel at reading authentic texts, and they only somewhat succeed when these texts have been partially modified to aid for comprehension.

Laufer (1996) warns us that if the 3000-word threshold were not met, lack of vocabulary would hinder the effectiveness of application of L1 reading strategies. Considering the results in this test, we could hypothesize that effective strategies must have been implemented by a large number of students in the study to understand most of the text. These however, are not enough to deal with the full load of information presented in these readings. Students have worked on developing reading strategies in the L2, however, neither these, nor the ones they use in L1 reading comprehension can aid them to fully deal with texts in the test, as is aptly argued by Laufer (1996). Furthermore, when considering that students work on strengthening reading skills (and thus its strategies) and aware that the content of the texts is well within the level that students should have (between 80 and 89% of words are in and below the 2000 word band and the rest is represented by highly familiar words outside this range as well as proper nouns), then, vocabulary is one of the factors that most likely aids or hinders comprehension for these subjects. Further analysis will be conducted in the next section to confirm or reject this possibility.

Another consideration that should be made has to do with the kind of practice that should be taking place in SL classrooms to ensure vocabulary learning through reading. Authors such as Bingham Wesche and Paribakht (2000), Hunt and Beglar (2005) and Pigada and Schmitt (2006) insist on the importance that repeated lexical encounters have to guarantee vocabulary learning through reading. Because the Costa Rican context places so much emphasis on the reading skills of students throughout their secondary education, there is a dire need that this practice yields as many positive results as possible. Vocabulary learning should be one of these results; however, the current findings do not seem to support the idea of this really happening. Pigada and Schmitt (2006) suggest that 20+ encounters are necessary to have a good chance of learning vocabulary items. If the right planning and course material selection take place, a large percentage of the first 2000 and 3000 word families could easily be acquired during 5 years of secondary education. The results

of the present study, however, do not back up the idea of students learning these word bands.

These authors also point to the type of tasks that facilitate vocabulary learning through reading. Bingham Wesche and Paribakht (2000) and Laufer (2001) describe the importance of additional exercises that accompany reading and their role for reading comprehension and vocabulary growth. Most of the tasks that Costa Rican students do when they analyze readings require students to complete fill-in tasks, write one-word or short sentence answers and do multiple-choice tasks. Laufer (2001) describes fill-in tasks and use of words to produce sentences as highly rewarding tasks. If the participants in the study are already used to some of these practices, one of the elements that seems to be missing is a more coordinated planning that offers a structure that allows multiple encounters with vocabulary items across different readings. Dividing the 3000-word band across five years and offering consistent exercises on key topics that promote learning of specific vocabulary families may result in large percentages of vocabulary gain per month and year.

Another aspect that has been mentioned before has to do with explicit and implicit vocabulary teaching as well as development of learning strategies that lead to vocabulary growth (Schmitt, 2008; Doczi and Kormos, 2016). The Costa Rican context does not escape this scenario. Explicit vocabulary teaching is merely absent from the classrooms and if it exists it is limited to pointing at the vocabulary lists that precede the lessons in a book. Authors such as Doczi and Kormos (2016) describe the need for a more prevalent role of vocabulary in course materials. This need is extensive to teachers and course designers as well. The Costa Rican context represents yet another example where this need is made evident. Instructors and students need to become aware of the importance that vocabulary plays in attaining high levels of reading comprehension. For this to be possible, explicit and implicit vocabulary learning is required. We believe that for students to be aware of the importance of implicit vocabulary learning, explicit vocabulary learning should be assumed by instructors and should guide students in the best practices to attain vocabulary growth.

The contribution of reading in regard to the enhancement of spelling, meaning, and grammatical word knowledge has been demonstrated (Schmitt and Pigada, 2006). While these aspects of expertise develop at different paces for different individuals, the benefits for students will always be there. Instructors should make sure that students have multiple exposures (over 20 according to Schmitt and Pigada, 2006) for vocabulary items to be maintained in memory. The

incidental learning of vocabulary that can take place through reading, however, needs to be supplemented through explicit vocabulary learning as both of these contribute to strengthening the general knowledge of words (Schmitt, 2008).

Also, the present study contributes to the ongoing debate about gender and second language learning performance. The different tests have shown that there are no significant differences among gender groups when it comes to reading ability for the subjects in the present study. The slight numerical advantages that are suggested by the mean scores in the reading test disappear when these differences are analyzed in closer inspection. In sum, no statistically significant advantages can be attributed to gender differences for the reading comprehension skill.

Finally, we must point out that type of methodology also has an effect on the reading capacity that students have achieved. As demonstrated in the exploration of the gender and school variables, the analysis showed that the school, and not the gender, seems to play a role in the ability that students demonstrated in reading comprehension. This leads us to conclude that learning language through content serves a better purpose when it comes to the application of this knowledge to the reading skill. While students in the FL school achieve some level of reading comprehension, when compared with counterparts in the CB school, the latter seem to maintain an advantage in this skill.

A second aim in evaluating the students' reading proficiency is to establish a possible relation between the vocabulary level of students (based on the results obtained on the VLT) and their reading comprehension ability (based on the scores obtained on this reading test). The possible connection between VLT and reading comprehension is explored in the following section.

The fact that Costa Rican students receive a high degree of reading practice where they have to complete reading comprehension exercises seems to work in their favor when it comes to the development of the reading skill up to a certain extent. As claimed by Bingham Wesche and Paribakht (2000), when students work on activities that require higher concentration on the information contained in the reading passage, they have higher chances of learning words and they also train in more in-depth sentence processing.

Laufer (2001) also points at the importance of word-focused activities, especially if students only have access to small amounts of reading, as is the case in the Costa Rican EFL classrooms. She lists dictionary use, fill-in tasks and sentence or paragraph production as the best ways to support vocabulary. As it happens the first two, and to a more limited extent the third one, are also very

common in the Costa Rican classrooms. However, a more explicit approach to vocabulary teaching and learning is still needed in this context.

The balance between explicit and implicit instruction that Hunt and Beglar (2005) call for is not truly part of this country's teaching methodology, and the learning of vocabulary is usually left up to the individual interest of students. This should change. If we are not ensuring that students receive the necessary number of encounters (at least 20 encounters for Pigada and Schimitt, 2006) they need to learn the different vocabulary words in the 2000 word band, then explicit teaching of vocabulary should take place (along with the implicit exposure that may already be taking place).

While the level of reading comprehension may be enough to reach certain objectives, more adequate capacities in the development of this skill are still necessary. There are still a percentage of students, especially in the FL school who do not reach adequate levels of reading comprehension. A strong vocabulary baseline may be the answer for these deficiencies. Further studies involving larger population samples and longitudinal studies are necessary to provide more answers in this area.

7.3. VOCABULARY AND READING

Several studies have looked at the connection existing between vocabulary and reading comprehension. Qian (2002) investigated the relation between certain vocabulary aspects and reading comprehension. In a study involving 217 students from 19 different L1s who took part in an intensive ESL course at the University of Toronto, Qian used four different instruments to measure diverse aspects of vocabulary and reading comprehension. He used the Test of English as a Foreign Language, Reading for Basic Comprehension (TOEFL-RBC) to examine students' performance in basic reading comprehension. Qian also used the Depth of Vocabulary Knowledge (DVK) Measure to determine students' depth of word knowledge regarding polysemy, synonymy, and collocation aspects. In addition, he used the Vocabulary Size (VS) test (Nation, 1983) to determine vocabulary size; and finally he used the TOEFL-Vocabulary Item Measure (VIM) to check for synonymy knowledge in isolated sentences, a type of modified vocabulary depth measure.

While Qian (2002) found that all the different types of vocabulary measures mentioned above are directly associated with reading comprehension to similar extents, he also notes that any two variables combined would convey much higher results. He further concludes that DVK and VS are strongly associated with one another and with reading comprehension. Of particular interest are the results of the VST and its effect on reading comprehension. Qian (2002, p. 528) determined that according to R^2 statistical results, the "VS alone explains about 54% of the variance in TOEFL-RBC." That means that the VS test serves as a strong predictor for a good portion of the variance of the reading scores. Qian (2002) also insists that well-designed measures of breadth and depth of vocabulary represent valid predictors for reading comprehension. This study serves as evidence of the value that tests such as the VS test, now called VLT, have to create proper projections of the results students may obtain in reading comprehension tests.

In another study dealing with vocabulary breadth and reading comprehension Jiménez Catalán and Terrazas Gallego (2005) estimated the vocabulary size and its relation to reading in Spanish, primary school, EFL children. These children age 10.3 received English for approximately 419 hours. Using the Vocabulary Size Test the authors determined that the vocabulary size for these learners is set at 737 words, most of which belong to the first 1000 word band. Jiménez Catalán and Terrazas Gallego report 559 words for the first 1000 level, and 178 words for the 2000 level band (p. 184). After examining the relation between the vocabulary size of the students and the scores in the reading test, the authors report a positive significant correlation between the reading test and each one of the tests that measures vocabulary size (1000 level, $r = 0.282$, $p = <.001$; 2000 level, $r = 0.208$, $p = <.001$). They have noted that the higher the scores were for the reading test, the higher these were also for the VLT. Regarding gender, Jiménez Catalán and Terrazas Gallego (2005) only discuss gender differences in regard to the VLT but not in connection with the reading test. In sum, the results presented by Jiménez Catalán and Terrazas Gallego (2005) are in line with the idea that vocabulary knowledge can be associated with SL proficiency and with reading in specific, which is the particular interest of their study.

Further evidence is contributed by Stæhr (2008) in a study involving the VLT and reading comprehension, where students were evaluated in four levels: 2000, 3000, 5000 and 10,000 of the VLT (Schmitt et al., 2001). No total vocabulary counts are presented in his study, but the author notes that only 20 out of the 88 participants in the study mastered the 2000 VLT. Twelve of these mastered only the 2000 level; three participants mastered the 2000 and 3000 word levels; and five

participants mastered the 2000, 3000, and 5000 word levels; none of them mastered the 10,000 word level. Participants also took a reading comprehension test containing multiple-choice and multiple-matching items that evaluated reading comprehension of several text types, a very similar format to the one used in the present study.

Stæhr found a positive correlation of 0.83 between vocabulary size in the VLT and reading comprehension. Further regression analysis established that as much as 72% of the results of the reading test could be predicted by the VLT (R^2 = 0.722). Stæhr (2008) concludes that the 2000 word level "represents some sort of probabilistic vocabulary threshold level above which learners will perform adequately in the reading test" (p. 147). Although he notes that a fair number of students (38%) did not master the 2000 VLT and still obtained above average scores in reading, he insists that knowing the 2000 word level leads to significantly better results on the reading test.

Lervåg and Aukrust (2010) present additional information on the subject. They discussed the critical role that vocabulary has in reading comprehension both for L1 and L2 languages. The study involved 198 native speaking Norwegian children and 90 Urdu children who were L2 Norwegian learners. The study used the Norwegian version of the Woodcock Reading Mastery Test-R, Passage Comprehension (WRMT-PC) and the Neale Analysis of Reading Ability II tests to assess reading comprehension and a Norwegian adaptation of the Peabody Picture Vocabulary Test III was used to calculate vocabulary breadth. Lervåg and Aukrust (2010) insist that while decoding skills and vocabulary can predict reading comprehension skills in early stages of reading skills development, both for L1 and L2 students, vocabulary is essential to predict the progress of the reading skill to a greater extent in L2 learners than in L1 learners.

These authors identify lack of vocabulary as the factor responsible for L2 students lagging behind L1 learners in the development of adequate skills for reading comprehension. In addition, they claim that growth in reading comprehension ability "could be fully accounted for by initial differences in vocabulary," and demonstrates the heavy load that vocabulary represents in connection to reading skills. As they point out, having a well-established vocabulary permits students to acquire more words through reading, and contributes to reading comprehension. What these results indicate is that L2 vocabulary instruction as well as its development and growth ought to receive special attention in SLA classrooms from the beginning stages in the language learning process given the critical role that vocabulary plays.

As can be established based on the aforementioned evidence, previous studies point to the critical connection between vocabulary knowledge and reading comprehension. With this information in mind, we turn now to the analysis of the connection between the results of the VLT and the reading comprehension test in the present study.

7.4. THIS STUDY: VOCABULARY LEVELS AND READING

To determine the relation between the effects that the vocabulary level of the students exerts on their reading ability we implemented a regression analysis. We will analyze the results for the case of the CBT first. The assumption of residuals normally distributed around the mean is confirmed through a histogram in the output of the test. According to the results, the two variables are related with a moderately high Pearson correlation coefficient of .730, a positive correlation with significance .000. A significant regression equation was found (F (1, 53) = 60.474, p <.000) with an R^2 of 0.533 (N = 55). According to this R^2, 53% of the variance in Reading scores can be predicted by the results on the VLT. Based on the coefficient information, the participants' predicted score in reading is equal to 12.924 +.819 (from the score in the VLT) for the reading score when the reading ability is measured according to the score that students obtained on the VLT. The score for the reading test increased .819 for each word that students recognized in the VLT.

As for the FLT school, a residual normal distribution is also confirmed in the data. In this school, the correlation is also significant, (F (1, 128) = 14.806, p = <.000), with an R^2 of .104. The Pearson correlation shows a much lower relation of .322 for this school, although still positively significant p = .000. The students' reading score is equal to 50.642 + .317 (VLT) for the reading score if this is based on the students' scores for the VLT. For the FLT students, the reading test increased .317 for each word students were able to recognize on the VLT. Based on the R^2 for this school, 10.4% of the variance in the reading score could be determined through the VLT results. The distribution is illustrated in Figure 14.

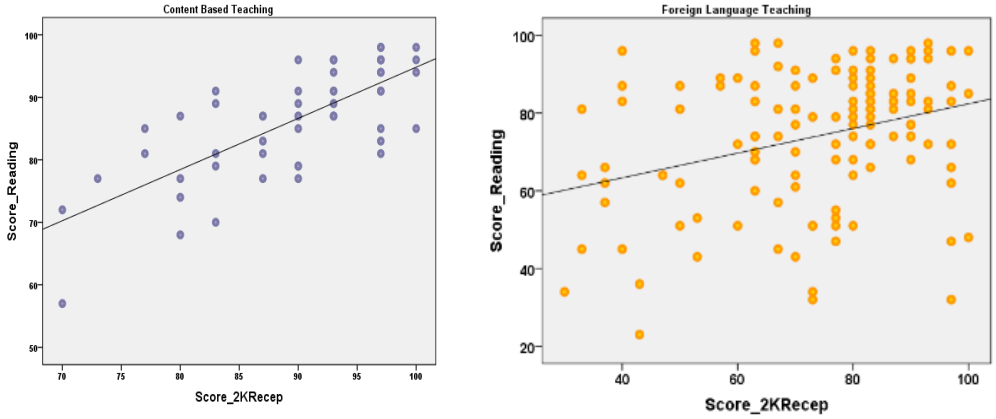


Figure 14. Regression analysis for the VLT and Reading Test.

Figure 14 illustrates how the variance in scores is better accounted for by the VLT in the CBS than in the FLS. For the CB school, we see the distribution of the points scattered much closer to the regression line than they are for the FL school. The farther the points are scattered from the regression line, the lower the variance is accounted for, in our case, by the VLT. To summarize, the reading ability of the students, measured through the scores students obtained on the Reading Test, can be predicted through the VLT, albeit to different degrees. For the CBS, 53% of the students' score variation could be predicted based on their results on the VLT. In the FLT School, 10.4% of the variance in the scores in the reading test could be foretold based on the results of the VLT.

7.4.1. Discussion

The results of the present study are in line with others that show that the size of the vocabulary of the students is key in the capacity for reading comprehension (Qian, 2002; Stæhr, 2008). Based on the present results, we could argue that the greater the vocabulary size of the students, the better the VLT can be used to explain the capacity that students have in terms of reading comprehension. It appears that when the students' vocabulary size is very limited, the VLT does not help us explain the students' performance in a reading test to a very large extent. The fact that students in the FL school do not have full knowledge of the 2000 word level band

limits the capacity of the test to predict how well students could perform on a reading test. This is supported by the fact that the ability of students in the CB school, who (barely) mastered the 2000 word band, is better explained through the VLT. The results of the Pearson correlation coefficient also seem to indicate that something else (besides vocabulary knowledge) is associated with the FL students' performance in the reading test. It could be argued that many of these participants seem to have developed enough reading comprehension strategies to make up for lack of vocabulary and that these strategies may be serving their purpose for many of the students.

The results for the CB school indicate that 53% of the variance can be explained by the results of the VLT. These results are very much in line with those obtained by Qian (2002), who found 54% of the variance in his reading test could be predicted by the VS test, a former version of the VLT. They are much lower than the 72% mentioned by Stæhr (2008), who also used the VLT. In the case of the FLS, however, we see a similar pattern as that described by Stæhr for some of his participants. Stæhr (2008) describes a 38% of students who do not handle the 2000 word band and yet obtained adequate reading comprehension. That seems to also be the case for many students in the FL school who have a mean word count of 1475.85 in the VLT and obtain a mean score in reading of 74.18. As discussed above, this score would represent a passing grade in the Costa Rican system and, to that extent, would be arguably representing an "acceptable" level of reading comprehension in that context. As Laufer and Ravenhorst-Kalovski (2010) argue, levels of acceptability for reading comprehension vary across different contexts, and in this case, although it does not seem ideal, it seems to be serving the purpose of passing reading tests for many students in this scenario.

Knowledge of the 2000 high frequency words as determined by the VLT seem to serve a predictive effect in reading comprehension. Lack of knowledge of this vocabulary band hinders the establishment of stronger, more direct correlations between vocabulary and reading. While students could make up for some of this lack of vocabulary through the use of their own strategies, knowledge of this high frequency vocabulary seems to be key for overall reading comprehension. This analysis has shown that better mastery of the 2000 word band results in higher scores in the reading comprehension test. Evidence shows that vocabulary knowledge is the most critical aspect for reading comprehension (Beglar and Hunt, 2005). The present study substantiates this claim.

CHAPTER 8

PRODUCTIVE VOCABULARY THEORY AND ANALYSIS

8.1. PRODUCTIVE VOCABULARY

The ultimate intention of learning a language is to use that language for communication. To achieve that in an L2, learners require enough vocabulary to convey the meaning they wish to express. The productive vocabulary also represents a major keystone in second language development. The number of words in our vocabulary, learned during the SLA process, would (ideally) increase gradually. However, Laufer (1998) comments that despite evidence of vocabulary growth in L2 learners, the totals are radically lower for second language speakers when compared to native speakers' word counts. Moreover, Laufer (1998) insists that word knowledge advances from shallow to deep as students' learning progresses. That means that a strong receptive vocabulary basis leads to the solid development of active vocabulary. This process takes time and effort and develops differently for different individuals under different conditions. Productive vocabulary has also proven to be essential for young children.

There is evidence that demonstrates that productive vocabulary is as important in our mother tongue as it is in our second language development. Vagh et al. (2009), for example, provide evidence for faster growing, higher-quantity, productive vocabulary in monolingual rather than bilingual children up to 36 months of age. This shows how specific conditions may affect productive vocabulary development and calls for consideration of the consequences of these differences. Likewise, Bleses, Makransky, Dale, Højen and Aktürk Ari, (2016) provide further rationale on how children's early productive vocabulary measures can predict future educational outcomes in decoding and reading comprehension abilities. They establish a link between early vocabulary measures and language development, and then make a meaningful association between language and literacy (also present, to a lesser extent, for mathematical skills). Bleses et al. (2016) propose using oral skills, for which productive vocabulary plays a key role, as a strong basis for reading. These studies signal the long-term significance and impact of vocabulary on the educational future of monolingual and bilingual students. These are additional reasons to justify the need for further productive vocabulary research. These findings from younger learners may be extended to learners of other age groups as well.

8.1.1. Productive vocabulary measures and their tests

The number of studies that describe productive vocabulary measures is meager. Authors agree that measuring productive vocabulary knowledge is indeed much more difficult than measuring receptive vocabulary knowledge (Laufer and Nation 1995, 1999; Meara, 2009; Webb and Nation, 2017). Meara (2009) assigns this difficulty to the fact that the vocabulary produced by a learner is subjected to the specific context of the test; as such, it does not offer a full picture of the complete productive knowledge that the speaker possesses. In SLA studies, Laufer (1998) acknowledges that the process of vocabulary learning moves from receptive to productive expertise and that receptive vocabulary is usually more extensive than productive vocabulary.

Any language learner would agree that both types of vocabulary are, indeed, essential during the learning process. By establishing the size of vocabulary (receptive and productive), teachers and program designers can trace students' progress in any given language program. Laufer and Nation (1999) insist on the importance of determining learners' vocabulary knowledge to design the vocabulary component of teaching programs. While studies on receptive vocabulary have become more frequent in the last 10 years or so, there is an evident necessity for more studies dealing with productive vocabulary measures. The latter are not found in academic research as frequently as the former.

Laufer and Nation (1999, p. 36) insist that the ability to produce vocabulary "implies degrees of knowledge" and claim that students can use vocabulary forms in some educational situations (i.e., when required by the teacher) and not in others (i.e., free writing). Laufer and Nation (1999, p. 37) establish a distinction between "the ability to use a word at one's free will as 'free productive ability' [and] 'controlled productive ability' for the ability to use a word when compelled to do so by a teacher or researcher." The PVLT measures the latter.

Controlled productive ability is of particular importance because we can predict (given the idea of progressive degrees of knowledge) that students should be able to use vocabulary in these controlled settings before they can actually use it in activities that provide a freer context requiring deeper vocabulary knowledge such as writing. Thus, this section of the present study deals with *controlled productive ability*. The intention is to determine the students' ability in this controlled context as preliminary information to support the analysis of students' production in free writing.

Nation (2013) notes that productive knowledge of vocabulary, the knowledge required to speak and write, implies form recognition and form recall. Through the knowledge of these aspects learners reflect their ability to actively use vocabulary knowledge in language production. According to Nation (2013, p. 561), procedural knowledge of vocabulary refers to the "learners' ability to use words receptively and productively when their focus is on the message that they are receiving and conveying." This represents the main goal of learners when they are learning a second language. To be able to understand and respond to any given message in the target language would mean that speakers are communicating through that language. Webb and Nation (2017) further argue that for learners to develop productive ability would demand students to previously possess the corresponding receptive knowledge; the opposite does not necessarily occur. They maintain:

> Productive learning is more difficult than receptive learning, mainly because it requires more precise knowledge of aspects of the form, meaning, and use of words than receptive learning. It also requires the learners to give attention to aspects of vocabulary knowledge that are not so critical for receptive use. (p. 180)

With this information in mind, along with that analyzed in the previous section about the receptive vocabulary knowledge of the students in the present study, we can address the state of productive vocabulary knowledge in this population.

8.1.1.1. The Productive Vocabulary Levels Test (PVLT)

The PVLT (Laufer and Nation, 1999) was originally created as a diagnostic test and it has subsequently been used to help profile students' ability to use productive vocabulary in context. The Productive Vocabulary Test is a discrete-point vocabulary test that assesses productive vocabulary knowledge at different levels of frequency: 2000, 3000, UWL, 5000, and 10,000 word levels. Each level test contains 18 items. In these tests, students are given an incomplete word immersed in the context of a sentence and they must provide the completion of the word that better fits the context. The number of letters in the cue is just enough to disambiguate the concept and to prevent students from providing words from other frequency levels. The test has been validated with students from different language

backgrounds and it has been proven to offer reliable results and to be practical as well (Laufer and Nation, 1999).

Meara (2009) ascribes certain shortcomings to the PVLT; namely, that it is more effective to measure vocabulary in low-level language students. Second, at high frequency bands (i.e. 10,000 word band) it is hard for the PVLT to extrapolate the vocabulary measured through 18 items when assessing bands that contain thousands of words. Third, he further argues that the PVLT, due to the clues that it provides, may limit the type of response that students may give. Despite these alleged limitations, the PVLT has been validated in multiple studies. Given that the subjects of the present study are low-level second language speakers, the PVLT still seems to be the best available option as a tool to profile the vocabulary of these students. Since we will be assessing students in the lower level frequency bands, Meara's alleged limitations do not appear to posit a problem for the present study.

8.1.1.2. Lex30

Lex30 (Meara, 2009) represents another available instrument to measure students' productive vocabulary; however, it has not been as commonly used as the PVLT. Meara (2009, p. 33) claims that "there are no well-established and easy-to-use tests of productive lexical skills" and offers Lex30 as an option that "generates rich vocabulary output from testees, […] is easily administered, and can be scored automatically using a computer program" (p. 36). Lex30 serves as a word association task where students are given a set of 30 stimulus words and then learners respond to these stimuli producing words that they associate to the triggering concept. There is no set number of response words to be produced. Also, while the stimulus words may be said to exert some kind of constraint on the responses given by learners (due to the nature of the topics the words may be associated with), the test still offers some freedom for students to produce the words they know. The stimulus words are high frequency words and they elicit multiple response words, which are often part of upper level vocabulary bands. The elicited words are, then, processed in a program similar to Nation's *VocabProfile*, an earlier version of *AntWordProfiler* used in the present study. In this program, the words are divided according to the word level they belong to. Meara (2009) concludes that the test offers the opportunity to trigger a wide variety of vocabulary through simple stimuli while it also results in a large number of words produced by the learners. We have opted here for the PVLT as it seems to be more suitable for the population

of this study. It offers the practicality necessary to deal with a moderately large number of participants and a format that is familiar for the present population.

8.2. PREVIOUS STUDIES

Table 31 summarizes the only studies the author of the present analysis has located on productive vocabulary measures where the PVLT has been used. Lex30 was also applied in one study by Moreno Espinoza (2010a); however, no total word counts were provided.

Table 31. Previous Studies using the PVLT

Study	Participants	Hours of Instruction	Vocabulary Size
Laufer (1998)	Israeli, high school, 16 and 17 years old	1,080 1,260	1,314 words 1,667 words
Laufer and Nation (1999)	Israeli, high school, 16 to 19 years old	1,080 (10th grade) 1,260 (11th grade) 1,365 (12th grade) 1,560 (university)	1,311 words 1,667 words 1,800 words 1,889 words
Moreno Espinoza (2010b) in Canga Alonso & Arribas García (2014)	Spanish, high school (age not reported)	Not reported	645 words
Canga Alonso & Arribas García (2014)	Spanish, high school, 15-16 years old	1,049	644 words

In Table 31, four different reference groups are shown; they are presented here along with the test performance, hours of instruction and age. The first groups come from Laufer (1998), and Laufer and Nation (1999); for these first groups, Hebrew is the L1. For the other two studies, Castilian Spanish is the L1. All groups are studying English as a Foreign Language. Two important distinctions can be drawn from these data. First, the number of words for the Israeli students is over double that of the Spanish students. Second, the number of hours of instruction seems to have a different effect on the results of word counts for each language background group.

In reference to gender, studies discussing productive vocabulary measures and the gender variable are even scarcer. Of those presented in Table 31, only Canga Alonso and Arribas García (2014) address this issue, reporting no gender-

based differences. Despite initial numerical differences favoring female students, Canga Alonso and Arribas García (2014) conclude that no statistically significant differences prevail. In another study, Moreno Espinoza (2010a), using Lex30, also found no differences in terms of gender effects. In that case, female students also produced higher means in numbers of types and tokens, although these do not represent statistically significant differences.

8.3. THE PRESENT STUDY[6]

8.3.1. Participants

A total of 182 students sat this test: 54 students from the CB school (21 male and 33 female); and 128 from the FL school (64 male and 64 female students).

8.3.2. Instrument

The Productive Vocabulary Levels Test (PVLT) (Laufer and Nation 1999) is used to determine the "controlled productive vocabulary" of the participants; particularly, the parallel version 1 (Version C) of the PVLT-2000 band. The test consists of 18 items where students have to complete a word that fits the context of a given sentence. A number of letters are provided in each case with the idea of triggering the target word. As mentioned above, for Laufer and Nation (1999), this provision of letters serves a two-fold purpose. First, it provides as many letters as are required to disambiguate the cue, thus eliminating other possible alternatives; second, it has the intention of preventing learners from providing words from a different frequency band. This test has been found to be reliable, valid and very practical to implement (Laufer and Nation, 1999, p. 44). The following is an example of the format of the items in this test:

Sample: 1. I'm glad we had this opp_____ to talk.

[6] Part of this analysis was discussed in D. Castro García (2017), Are we preparing secondary students for a productive use of vocabulary in English as their second language? *Porta Linguarum*, 28, 141-155.

8.3.3. Procedure

Data were collected in one session per group in each school, during the English class in all six sub-groups (two groups from the CB school and four groups in FL school). At the beginning of the PVLT, both Spanish oral and written instructions were given to explain what the students were being asked to do. A sample sentence was also provided. Students were allowed 15 minutes to complete the PVLT. Regardless of the institutional setting, some students completed the test in less than 15 minutes while others needed the full 15 min.

Following Laufer and Nation (1999), the answers were checked based on semantic correctness. This means that they were graded as correct or incorrect, and that minor spelling mistakes (i.e., "lovley" instead of "lovely") and grammatical mistakes (i.e., "introduce" instead of "introduced") were ignored.

Descriptive statistics were calculated and differences between schools and gender groups were compared by using SPSS Version 20 (IBM Corp, 2011). During the analysis of the data, a Cronbach's alpha reliability test was run, and it reported internal consistency of .75 for the CB school and .85 for the FL school.

8.3.4. Results

This section offers the analysis of the results for the PVLT. First, an overview of the general results of the test is presented with the analysis of response per item. Then, a statistical analysis of the outcomes in the test for each school is provided, and the results of the consideration of the variable gender are analyzed. Appendix 3 offers access to a sample of the PVLT used in this study.

8.3.4.1. Item analysis

Table 32 presents the results for correct, incorrect and missing data for each item in each one of the schools.

Table 32. Correct, incorrect, and missing percentages per item in the PVLT

	Content Based Teaching			Foreign Language Teaching		
	Correct	**Incorrect**	**Missing**	**Correct**	**Incorrect**	**Missing**
Item 1	94.5%	3.6%	1.8%	74.8%	22.9%	2.3%
Item 2	98.2%	--	1.8%	71.0%	26.7%	2.3%
Item 3	83.6%	14.5%	1.8%	48.9%	48.9%	2.3%
Item 4	76.4%	21.8%	1.8%	40.5%	57.3%	2.3%
Item 5	63.6%	34.5%	1.8%	27.5%	70.2%	2.3%
Item 6	50.9%	47.3%	1.8%	37.4%	60.3%	2.3%
Item 7	70.9%	27.3%	1.8%	75.6%	21.4%	3.1%
Item 8	40.0%	58.2%	1.8%	23.7%	74.0%	2.3%
Item 9	25.5%	72.7%	1.8%	7.6%	90.1%	2.3%
Item 10	89.1%	9.1%	1.8%	61.1%	36.6%	2.3%
Item 11	89.1%	9.1%	1.8%	77.9%	19.8%	2.3%
Item 12	90.9%	7.3%	1.8%	74.0%	23.7%	2.3%
Item 13	29.1%	69.1%	1.8%	7.6%	90.1%	2.3%
Item 14	89.1%	9.1%	1.8%	72.5%	25.2%	2.3%
Item 15	78.2%	20.0%	1.8%	44.3%	53.4%	2.3%
Item 16	38.2%	60.0%	1.8%	20.6%	77.1%	2.3%
Item 17	98.2%	--	1.8%	68.7%	29.0%	2.3%
Item 18	98.2%	--	1.8%	92.4%	5.3%	2.3%

For the CBS, the table shows that there are four items that proved to be the most difficult for students in this school. The item with the lowest correct score was item 9, for which students had to provide the word “pupils”; this one reached only 25.5% of correct responses. For item13, “burn,” the correct percentage came to 29.1%. In the case of item 16, “slim,” reached 38.2% of correct responses. Finally, item 8, “wealth,” obtained 40% of correct responses. From these four items, we find that the word “pupils” may not be as frequently used in American English as it is used in British English. Given that students are more familiar with American English, students might not be as familiar with this word and this might be a possible explanation for the low results obtained in this item. A second group of words represents moderate difficulty for students from this school. The words “lack,” “charm,” “treasure” and “delivered” obtained correct responses of 50.9%; 63.6%; 70.9% and 78.2% respectively. The rest of the words obtained between 83.6 and 98.2% of correct answers: “opportunity,” “dozen,” “tax,” “stretched,” “introduced,” “admire,” “improve,” “lovely,” and “popular.” Although the results suggest an adequate degree of knowledge for some of the words in the test, the percentages are fairly low for over half of the items in the test.

In the FLS the results are even more notorious. The percentages of correct responses are markedly inferior. The word “popular” was the only one showing

results above 90% of correct responses, at 92.4%. In the range of 70% we find: "introduced" (77.9%), "cream" (75.6%), "opportunity" (74.8%), "admire" (74.0%), "improve" (72.5%), and "dozen" (71.0%). Between 70 and 40%, we have the items "lovely" (68.7%), "stretched" (61.1%), "tax" (48.9%), "treasure" (40.5%), and "delivered" (44.3%). The most difficult items for students in this school are: "lack" (37.4%), "charm" (27.5%), "wealth" (23.7%), "slim" (20.6%), and the words "pupils" and "burn" with 7.6% respectively. Needless to say, these percentages are extremely low.

These results show that the FLS students have a limited knowledge of the 2000 word band. The four concepts that proved to be the most difficult for the CBS are also the ones that obtained the lowest correct percentages for the FLS: "pupils," "burn," "slim" and "wealth." Further analysis will be discussed below to determine more specific findings in regard to the results in this test.

8.3.4.2. Results in relation to word counts

Table 33 presents the results for correct items and total word counts in this test. Following Nation's (1990, p. 78) formula, "Vocabulary size = N correct answers multiplied by the N words in dictionary (the relevant word list) divided by N items in the test," a relation was established between the number of items (maximum 18) and the total number of words that these represent in the 2000 word band. For the total of 18 items, the CB school reaches a mean of 13.20 while the FL school's mean is 9.41. Although in both schools there are students who reach the maximum possible (18); the minimum for the CB school is 8 whereas for the FL school it is zero. This difference is reflected in a much wider spread expressed in the SD of 4.284 for the FL school as opposed to 2.836 of the CB school.

Table 33. Item and total word counts in the PVLT

PVLT 2000 (n = 182 (54 CBT, 128 FLT)				
	Item count information		**Total word count information**	
	CB School	**FL School**	**CB School**	**FL School**
Number of items	18	18	2,000	2,000
Mean	13.20	9.41	1,467.33	1,046.80
Median	13	10	1,444	1,111
Max words	18	18	2,000	2,000
Min words	8	0	889	0
SD	2.836	4.284	312.075	477.644

Regarding total word counts, the mean is 1,467.33 words for the CB and 1,046.80 for the FL school. We also see that students in both schools reach a maximum of 2000 words. The minimum, however, is clearly different: it stands at 889 words for the CB school and zero for the FL school (one subject did not know any of the words in the test). Figure 15 illustrates the distribution of these results.

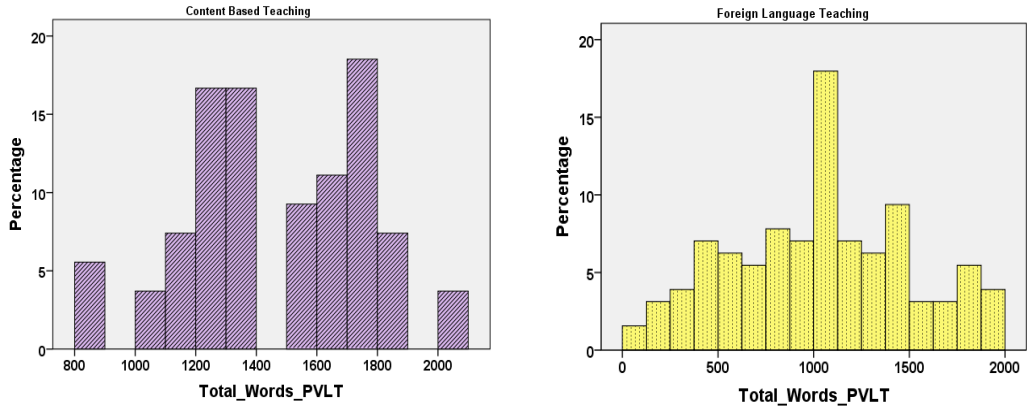


Figure 15. Word distribution per school in the PVLT.

In an analysis of distribution, the frequency of distribution for the CB school presents two clearly separate word-level size groups. The first group of learners, in the lower levels, ranges from 800 to 1400 words with a peak between 1200 and 1400 words; then the data show a marked halt. The second cluster of learners' vocabulary size grows from the 1600 word level with the highest peak at the 1700 to 1800 word range. It then slopes down to the 2000 word range. The distribution for the FL school appears more evenly dispersed. The word sizes steadily grow from zero to the highest peak of 1000 words; they then descend to 2000 words.

Figure 16 offers an additional perspective on the distribution of word counts for these two schools. The spread of the distribution as well as the difference in the median are more easily identified in this figure. Additional possibilities of interpretation exist, such as eliminating the data from the individual that obtained a zero score, considering this as an outlier. At this point, all results are included to offer a more complete picture of the findings. The figure shows that the data for the FL school is distributed across the 2000 words in the band, thus evidencing the spread of knowledge along this word band in individuals from this school. The fact

that the spread of distribution is more concentrated in the CBS appears to indicate a more cohesive degree of knowledge of words in this band for the CBS.

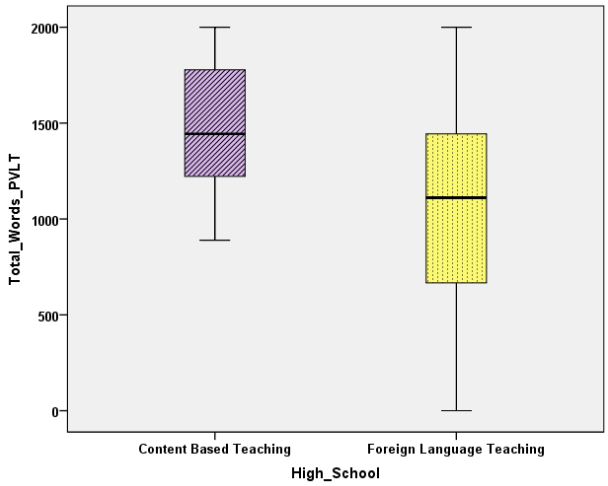


Figure 16. Box plot word distribution per school in the PVLT.

To determine normality assumptions in our data, with the purpose of defining the nature of the differences between the results of both schools in the PVLT, the Kolmogorov-Smirnov test was used. It confirms that the data follow a non-normal distribution for the CB school (p = .001), while the FL school data follow a normal distribution (p = .200). Table 34 shows the results of a Mann-Whitney U test that was then implemented to determine differences between the two school groups and the p value (.000) shows that there are significant differences favoring the CB school. The test shows a median rank of 124.38 for the CB school and 77.63 for the FL school.

Table 34. Results of Mann-Whitney U tests for inferential statistics in the 2K PVLT

N = 185	**2000 PVLT**
Mann-Whitney U	1,680.500
Wilcoxon W	9,936.500
Error	323.900
Z	- 5.482
P (two-tailed)	.000

8.3.4.3. Results in relation to gender

According to Table 35, the data in this sample population provide evidence that the overall productive vocabulary of male students is higher than that of girls in both schools, even if only slightly higher in the case of the FL school. Both male and female students achieve the maximum word scores in the CB school, whereas only boys reach that number in the FL school.

Table 35. Descriptive statistics for gender groups in the PVLT

	CB school		**FL school**	
	Male (N = 20)	**Female (N = 33)**	**Male (N = 64)**	**Female (N = 63)**
Total words	2,000	2,000	2,000	2,000
Mean	1,592.52	1,387.67	1,091.97	1,001.63
Median	1,778	1,333	1,111	1,055
SD	286.336	305.336	515.293	436.160
Max	2,000	2,000	2,000	1,888
Min	1,000	889	0	111

The word distribution according to gender is illustrated in Figure 17. The spread in distribution shows a wider range of distribution for girls in the CB school, while this is much wider for boys in the FL school. In general, overall productive vocabulary measures are higher for both male and female students in the CB school.

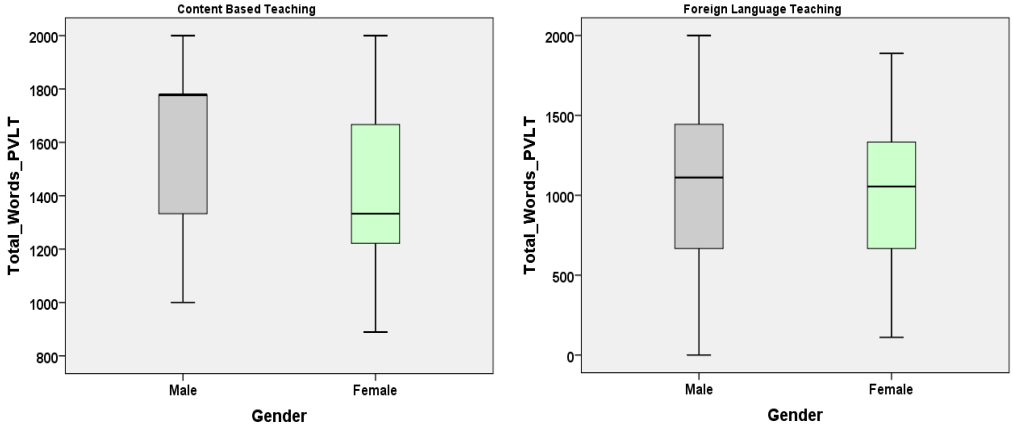


Figure 17. Box plot word distribution for gender effect in the PVLT.

In general, it could be concluded that the behavior, in terms of vocabulary outcomes, is quite different for each educational institution. While vocabulary counts seem to exhibit very similar behavior for both genders in the FL school, the differences between gender groups in the CB school are evident.

Table 36. Normality distribution for gender word counts in the PVLT

		Kolmogorov-Smirnov		
School	**Gender**	**Statistic**	**gl**	**Sig.**
Content Based	Masculine	.265	21	.000
	Feminine	.177	33	.010
Foreign Language	Masculine	.085	64	.200
	Feminine	.098	64	.200

Table 36 shows the results of a Kolmogorov-Smirnov test that was used on the data to establish normality assumptions for each gender group in each school. The test determines that the data for the CB school follows a non-normal distribution (masculine, p = .000; feminine, p = .010) while the data for the FL school follows a normal distribution (masculine, p = .200; feminine, p = .200). This calls for other statistical tests to explain the quality of the differences in each set of data.

Given that the data for the CB school follow a non-normal distribution, the Mann-Whitney U test was applied. The test produces a median rank of 34.05 that favors male students and a median rank of 23.33 for female students. The test results are present in Table 37. The test confirms that the distribution is not the same for the two genders and shows statistically significant results that favor male students in the PVLT in the CB school with a p value of .014.

Table 37. Mann-Whitney U test to determine statistical gender differences in the PVLT in the CB school.

2000 PVLT			
Mann-Whitney U	209.000	Min	889
Wilcoxon W	770.000	Max	2,000
Error	54.844	Mean	1,467. 33
Z	2.462	SD	312.07
P (two-tailed)	.014		

Since the FL school data follow a normal distribution, the Independent sample *t*-test is used to determine the nature of those differences. The test in Table 38 shows that there are no significant differences between male and female students in this sample population of the FL school.

Table 38. Use of *t*-Test to determine statistical gender differences in the PVLT for the FL school

		Levene's test for equality of variances		***t*-test for equality of means**		
		F	**Sig.**	***t***	**df**	**Sig. (2-tailed)**
SCORE_2K_PVLT	Equal variances assumed	1.96	.163	1.084	126	.280

The male student group (N = 64) obtained productive vocabulary level scores with a mean of 54.66 (SD 25.354), and the female sub-sample has a mean of 50.13 (SD 21.811). The results of the independent samples *t*-test are associated with a statistically non-significant effect: $t_{[126]} = 1.084$, $p = .280$. This means that the numerical differences observed in the data for total word counts do not translate into overall statistically significant differences for the gender variable at this school.

Additional tests were run, this time to compare gender results at inter-school level. As mentioned above, according to the Kolmogorov-Smirnov test, the distribution of the male CB school students follows a non-normal distribution ($p = .000$), while the FL male students' distribution follows a normal distribution ($p = .200$). Further examination using a Mann-Whitney U test shows that the numerically apparent differences between male students' scores are in fact statistically significant with a *p* value of .000 in favor of the CB male students. For the female students, the Kolmogorov-Smirnov test described the CB female distribution as non-normal ($p = .010$), whereas the FL female student group distribution was determined as normal ($p = .200$). The Mann-Whitney U test was implemented once again to determine the characteristics of the differences for total word counts in the PVLT for the female gender across schools. According to this test, statistically significant differences ($p = .000$) favor CB female students, when we compare the two schools in the study sample.

8.3.5. Regression analysis: VLT and PVLT

Using Lex30 and the yes/no test, Meara (2009) reports a "broadly proportionate" relation between the size of receptive vocabulary and the size of the productive vocabulary for many of the participants in his study. Laufer and Nation (1995) also maintain that to use a word correctly one must understand that word. That would mean that there must exist some kind of connection between the vocabulary that students know at the receptive level and what they use at the productive level. To test the possibility of an existing correlation between the scores obtained in the VLT and the results of the PVLT a regression analysis was conducted. Although the effect of the receptive vocabulary of students over their productive vocabulary is not as strong as one would have expected, we do find a positive correlation between these variables in both schools, albeit to a more limited extent in the FLS.

A regression analysis was applied to determine the relation between the effects that the receptive vocabulary knowledge of students may have on their productive vocabulary knowledge. In the case of the CBT, the assumption of residuals normally distributed around the mean is confirmed through a histogram in the output of the test. According to the results, the two variables are related with a moderately high Pearson correlation coefficient of .711, a positive correlation with significance .000. The ANOVA yields a significant regression equation (F (1, 52) = 53.216, p <.000) with an R^2 of 0.506 (N = 54). According to this R^2, 50.6% of the variance in the results of the PVLT can be accounted for by the results on the VLT. Based on the coefficient information, the participants' predicted score in the controlled PVLT is equal to -1110.712 + 1.437 (from the score in the VLT) for the productive score when the ability to use productive vocabulary is measured according to the score that students obtained in the VLT. The score for the PVLT test increased 1.437 for each word that students recognized in the VLT.

As for the FLT school, a residual normal distribution is also confirmed in the data. In this school, the correlation is also significant, (F (1, 126) = 6.396, p = <.013), with an R^2 of .048. The Pearson correlation shows a much lower relation of .220 for this school, although still positively significant p = .006. The students' productive vocabulary score is equal to 605.404 + .299 (VLT) for the productive vocabulary score if this is based on the students' scores obtained in the VLT. That means that for the FLT students, the productive vocabulary increased .299 for each word that students were able to recognize on the VLT. Based on the R^2 for this

school, 4.8% of the variance in the results of the PVLT could be determined through the VLT results.

8.3.6. Discussion

The first information to be drawn from the data analysis above, considering the 15-16 correct-item requirement for satisfactory mastery established by Laufer and Nation (1999, p. 41), is that neither of the groups in the present study can be said to master a productive vocabulary size of 2000 words. The CB school comes close at a mean of 13.19 but still does not make the cut. The FL school falls well below the requirement at a mean of only 9.30. While the differences between the results of the two schools prove to be statistically significant, these differences do not guarantee full knowledge of the 2000 word band for all students in the CB school group. These results, indeed, direct our attention to the FL school and its poor vocabulary attainment, given that the results are so much lower in this case.

Comparison of these results with those reported for previous studies in Table 31 reveals thought-provoking findings. If we compare our results with those in Laufer (1998), we discover that Laufer's (1998) groups reach much higher word counts despite having a relatively lower number of instruction hours. In her case, the 16-year-old group comes to 1,700 words (1,080 instruction hours), and the 17-year-old group reaches 2,550 (1,260 instruction hours). The same is the case for Laufer and Nation (1999). In this case, students reached 1,311 words (10^{th} grade, 1,080 instruction hours); 1,667 (11^{th} grade, 1,260 instruction hours); 1,800 words (12^{th} grade, 1,365 instruction hours), and 1,889 words (university students, 1,560 instruction hours). In the present study, the CB school reaches 1,467.33 words (1,368 instruction hours), and the FL school, 1,046.80 words (1,140 instruction hours).

With almost 1,400 hours of instruction, Laufer and Nation's (1999) group reaches 1,800 words; that is over 300 more words that the students in the CB group in the present study for a very similar number of hours of instruction. While not all of the groups in Laufer (1998) and Laufer and Nation (1999) reach the minimum requirement for mastery of the word band, all of the groups in those studies are still well above the mean of the present study. As clearly seen in the example mentioned above, the students who have attended a number of hours of instruction similar to

those received in the CB school show outcomes that demonstrate mastery of that word band.

The other two possibilities for comparison presented in Table 31 tell quite a different story. When compared to the results found by Moreno Espinoza (2010b) and Canga Alonso and Arribas García (2014), the students in the present study appear to exhibit a clear advantage. Moreno Espinoza (2010b) reports a word size of 645 while Canga Alonso and Arribas García (2014) speak of 644 words. For the latter study, both the age (15-16) and the total number of instruction hours (1,049) are a bit below those reported for the present study.

The question remains whether this seemingly small difference in age and hours of instruction could account for the well-defined difference in total word counts between the Costa Rican and the Spanish students. Both groups of Costa Rican students show higher word counts. In the case of the CB school, which is the one that resembles more closely the CLIL methodology followed in Spain, this difference is more than double in favor of the Costa Rican students. Although showing a more moderate difference, the FL, Costa Rican school also shows better results than those found in the Spanish studies.

In terms of the variable *gender* and returning to the data obtained in the present study, all other variables being equal, it would appear that the teaching method could have a different effect for male and female students in the CB school. The statistically significant vocabulary advantage of the male students could be due to how students are dealing with the material covered through English instruction. A similar logic could apply for general vocabulary attainment between both schools. Inter-school comparisons reveal a statistically significant advantage for both male and female students in the CB school. This suggests that students do benefit more from CB teaching when it comes to learning vocabulary. As mentioned above, although the results show that students do not yet have full command of the 2000 word band, CB students reflect a better mastery of this level.

As for the gender differences described in our study, these seem to add to the inconclusive results mentioned for gender studies and SLA. However, seen from a different angle, we could also say that these may contribute to the idea that the gender variable does not seem to be the determinant factor for variations in outcomes. As discussed in Chapter 5, differences may be regarded as mere results of the activity, task, background or personal conditions of the participants. Given the scarcity of studies, this *analysis* can only be compared to that of Canga Alonso and Arribas García (2014) and Moreno Espinoza (2010a). For those authors, the

male and female participants exhibit non-statistical differences in their productive vocabulary results, although in both cases female students' mean word is higher.

Contrariwise, in the present study, in both schools male students produced higher word means. Statistically significant differences favoring male students were found between gender groups for the CB school, while the differences are non-significant in the FL school. This hints once more at a greater effect caused by the teaching methodology rather than to a gender effect. It would appear that the treatment of vocabulary (or lack thereof) in the FL school produces the same effect for both genders. The gender similarity in productive ability in the test for this school seems to suggest that it does. In the case of the CBS, the results are not the same, and they would appear to indicate that male students are benefiting more from that approach. Further studies are required to determine the relevance of these conclusions.

Despite the evidence that points to a small correlation between the results of the VLT and the PVLT in the present sample, we still find that the correlation does exist and that it is becoming stronger as students use more vocabulary in production. This may be an indication that the VLT serves as a better predictor of productive proficiency at higher vocabulary levels. Similar results have been provided for tools such as the LFP (Laufer and Nation, 1995), which has been found to yield more stable results when measuring writing proficiency at levels beyond the 2000 word level.

Overall, these results call for reflection on the extent and quality of education that students are receiving in Costa Rica. It seems obvious that changes in the curriculum are necessary to provide students with better opportunities to produce language. Nation and Chung (2009), on the importance of providing students with the right opportunities for vocabulary learning, discuss the implementation of learning activities that rely on a set of equally balanced "four strands of meaning" (Nation 1996, 2007a). These four strands are *meaning-focused input*, *meaning-focused output*, *language-focused learning*, and *fluency development* (Nation 2007a). I would like to refer here to the *meaning-focused output* strand as one way to enable students to improve their productive vocabulary levels.

Nation and Chung (2009) describe this strand as achievable through speaking and writing activities, adding that if vocabulary were used to generate language creatively, this vocabulary would become strengthened in the students' memory. They suggest activities such as retelling, rewriting or group negotiating

tasks as prompters of this strand. While all four strands are significant, this one could be said to be likely absent from the classroom contexts under study.

Due to the constraints placed by the obligatory tests implemented by the Costa Rican Ministry of Education, the curriculum in Costa Rica relies heavily on reading. This is clearly to the detriment of students' productive abilities. Given the importance that knowing the first 2000 words exerts on language development in general, a more productive orientation of the curriculum would be called for (such as that provided by correct applications of CLIL, for example); that is, if we want our students to be able to use the language for real communication purposes.

Many reasons could be considered to explain the overall differences between the Spanish and the Costa Rican data. As mentioned in section 3.6, Costa Rica has constant contact with English-speaking tourists. There is also great contact with the English language through mass media. Cable television, music and Internet information in English are part of everyday life. The possibilities of having access to a job in a call center or transnational company are also appealing to many. All of these forms of contact have made Costa Ricans aware of the importance of learning English. These different interests may impact students' perceptions in regard to learning the language. It is possible that the Costa Rican students, immersed in a context where English proves itself to be a necessary tool, may be more interested in studying the language, while also benefitting from the exposure that they have to English through different channels.

However, the results of the present study indicate that the productive vocabulary profile of students in Costa Rica is limited and requires direct attention. Although neither school has mastered the 2000 word level, statistically significant results in productive vocabulary favor the CBT practice over the FLT methodology. As for gender differences, male students show significantly richer vocabulary in the CB school, while no gender differences are apparent in the FL school. We must analyze the incidence that these results have on much of the Costa Rican population, as the CBT is limited to certain private schools only. When learning a second language, students usually achieve a feeling of progress when they can find the words necessary to express their ideas. Having access to the language that helps them convey their thoughts certainly gives learners a more concrete substantiation of their growth in the language acquisition process.

While this idea of progression provides important motivation to actively continue in the process, the discouragement felt as a result of not knowing basic vocabulary can be detrimental. The study shows that, in their last year in high

school, the participants in the study have not reached the minimum vocabulary levels required to know the high frequency, basic, productive vocabulary of the 2000 band. The numbers are even more alarming for the FL school. They reflect a sample of the mainstream education that most of the Costa Rican teenage population receives. The educational system in Costa Rica requires attention and improvement. The evidence appears to suggest that while some efforts are being made to improve English language teaching in Costa Rica, they are not enough to conclude that we are actually preparing our students to use English in real communication evident in language production. If *controlled productive ability* is the preceding stage to free productive ability as Laufer and Nation (1999) suggest, we still have much work ahead to pave the way to effective communication.

CHAPTER 9

WRITING THEORY AND ANALYSIS

9.1. WRITING IN A SECOND LANGUAGE

Writing is a complex process that demands synchronization of many skills, from accessing information available in short and long-term memory, to attending demands at the structural and cognitive level for fine-tuned elements such as vocabulary, grammar and syntax. If we add doing so in a second language, the stakes are even higher. Laufer and Goldstein (2004) attest to the existence of a hierarchy of vocabulary skills and they maintain that *active recall*, the ability to come up with the target words when needed, is the most difficult task required of vocabulary knowledge. To produce a high-quality piece of writing, a great deal of organization is necessary. Ideas need to move from a mental conception to a written text through a process that requires brainstorming, planning, drafting and revision to eventually obtain the final product. If writers are familiar with the subject they are writing about (vocabulary being a key element of this knowledge) and have the appropriate writing techniques, they have the necessary tools to produce adequate pieces of writing.

As much as it has an effect on other language skills, vocabulary is one of the elements that lay the groundwork for writing in a non-native language. For Laufer (1994, p. 30), "the vocabulary quality of a piece of writing depends on the type of words used and also on the effective way of varying these words […]." We can infer that if learners lack the vocabulary they need, this will impede their ability to express themselves with ease. This difficulty will be reflected on poor writing quality that reveals limited language proficiency. Contrariwise, it has been noted that good writing offers various benefits in the second language learning process and it serves to demonstrate the writer's proficiency. Vocabulary is, thus, a key component of writing; and writing, in turn, has been associated with great gains when it comes to overall SLA. Pichette, de Serres and Lafontaine (2012) suggest that L2 writing offers advantages over reading due to the effect of "task overlap." They insist that given the fact that reading is necessarily involved during the writing process, this dual practice lends opportunity for deeper processing. The benefit of visualization that is associated with reading is also present when writing takes place. As discussed in section 1.1.1.4, the Output Hypothesis, discussed by Swain (1995, 2000a), is probably one of the strongest bases on which we can support and justify the objective of asking students to produce output through writing, as is the case for this study.

In addition, Swain argues that the additional mental effort that output demands leads students to put into practice all their knowledge to the maximum extent possible. This practical use of language is thus believed to have a tremendous impact on the process of second language acquisition. The benefits associated with writing are sought by the "meaning focus output strand," in Nations' (1996, 2007a), "Four Strands of Language Learning" theory (see section 4.4.4). Nation (2007a) maintains that Swain's (1995, 2000a, 2006) noticing/triggering function of output takes place in writing as students, in their attempt to produce language, identify the knowledge they lack. This theory can apply to the students' vocabulary. If the required knowledge has to do with vocabulary, these students would then find a way to solve their problem through searching and eventually learning the concepts they need. In addition, according to this theory, while writing, students also test different hypotheses and either maintain or modify their knowledge, even if the confirmation they need does not come immediately. Finally, the reflective function of the output hypothesis takes place during language use. Nation (2007a) maintains that blackboard or group composition produced in cooperation among students allows meaning-focused output and language-focused learning.

Furthermore, according to Cummins (2008), writing can be critical to consolidate knowledge that students have encountered through reading and it can help them consolidate their identities and expression of self. Production through writing can reflect the knowledge that students have acquired through practice, feedback, and correction, and serves as a cornerstone in the development of language proficiency. Whereas research in several areas involving vocabulary and second language acquisition is extensive, when we analyze research on vocabulary and writing, a number of researchers have argued that this area is indeed limited, as seen in Laufer (1994), Folse (2006), Lee and Munsie (2006), Hopewell and Escamilla (2014), and Harrison et al. (2016). Evidence suggests that vocabulary (productive vocabulary in particular) is among the most important factors for second language achievement. Surprisingly, to date, the connection between productive vocabulary knowledge and writing has not yet been extensively examined. This study seeks to contribute information to this area of study by providing an analysis of the students' written composition, and subsequently by establishing a connection between the quality of writing, in terms of vocabulary, and the productive vocabulary knowledge of students in the PVLT. Recognizing the many advantages associated with writing and aware of the demands and gains that writing brings, a solid basis is required to analyze writing samples provided by the participants.

9.1.1. Measuring vocabulary in writing

The analysis of the writing of students offers information about how much vocabulary they have acquired during their education and it also tells us how much of the vocabulary that students know is being used productively. While Laufer (1994, p. 21) acknowledges that several factors intervene in the writing process, she recognizes that vocabulary has a positive impact on writing and she adds that "[w]riting progress can be measured through lexical progress since lexical quality and writing quality are interconnected." Furthermore, Laufer and Nation (1995) note that "[o]ne of major determinants of the vocabulary used in written production is the vocabulary size of the writer, particularly if the writer is a second language learner with a relatively small vocabulary compared with native speakers" (307). For these authors, the intention of measuring vocabulary richness is to identify factors that affect writing quality and to determine the connection existing between vocabulary knowledge and its use. These authors further insist that effective use and richness of vocabulary, while representing only some of the factors affecting writing, are very important in this skill due to the impact they have on the readers.

The vocabulary used in written compositions can be measured in various ways. Laufer and Nation (1995) describe types of measures used more often for productive lexicon and indicate their drawbacks. First, "lexical originality" evaluates a writer's performance, how elaborate the writing is, as compared with that of the rest of participants in the group. As it depends on the environment, for Laufer and Nation (1995), the problem with this type of measure is that it is unstable and would change according to the group in which it is compared. Second, Laufer and Nation (1995) discuss "lexical density," the percentage of lexical words (nouns, verbs, adjectives, adverbs) in the total number of words in the text. For them, the disadvantage is that lexical words rely on the syntactic and cohesive features of a text, as well as on the connection between lexical and function words, rather than on lexis alone. The third type of measure for Laufer and Nation (1995) is "lexical sophistication," the percentage of *advanced* vocabulary in a text. However, *advanced* vocabulary may represent a subjective notion, as it depends on the researcher's own definition and changes from one learner to the next, making it an unstable measure. Fourth, in Laufer and Nation (1995), "lexical variation" is the ratio between the different types of words in a text and the total number of running words. This measure is also unstable in short texts. Furthermore, it depends on the definition given to "a word"; for example, whether derivatives are considered individual words or if the concept of family is used. It does not differentiate

between the types of words used in a composition in terms of how frequent or rare these words are. After analyzing the most common types of measures, the authors propose using the *Lexical Frequency Profile (LFP)*, which gives the percentage of words belonging to each vocabulary frequency level in the composition.

The LFP described in Laufer and Nation (1995) used the *VocabProfile* tool, an IBM compatible program that calculated the proportion of words per family at each level. This measure served as a basis for the creation of other vocabulary measure tools, such as the *ANTProfiler* (Anthony, 2014), the tool used in the present study. Here, as well as in Laufer and Nation's (1995), words are defined from the perspective of families, where a family is composed of a base word and its inflections, and its most important derivational forms. For example, *achieve, achieves, achieved, achieving, underachieve, achievers, (un)achievable* and *achievement* would all be in the same family. Thus, the present study measures the vocabulary of the written compositions according to their vocabulary frequency. This warrants each composition to be evaluated independently of the other compositions in the sample. Vocabulary is also analyzed in isolation of syntax and cohesiveness of the text. The subjectivity of sophistication is eliminated as vocabulary is measured according to the lexical band it belongs to. Finally, the notion of family as the basis for identifying the frequency of the words in the text eliminates the instability for shorter texts. All of these ideas are key, according to Laufer and Nation (1995).

9.1.2. Developing writing skills

Various authors discuss the benefits associated with writing practices in the SLA classroom. For example, in their study, Laufer and Nation (1995) worked with university students who were taking English for Academic Purposes in New Zealand (Group 1), and English Language and Literature in Israel (Groups 2 and 3). Students in Group 1 had the lowest proficiency level, while those in Group 3 had the highest of the three. The results of Laufer and Nation (1995) are shown in Table 39. For the two compositions that students wrote for their study, those with the lowest proficiency levels made more use of words at the 1000 level. Laufer and Nation (1995) conclude that as the language level of students progressed, the number of 1000 word level words was reduced and the words from the University Word List (now AWL) increased. That means, they say, that a richer vocabulary is characteristic of more advanced language learners.

Table 39. Laufer and Nation's (1995) results for word types in compositions

Word level	1000 level		2000 level		UWL		Not-in-the-list	
	Comp1	Comp2	Comp1	Comp2	Comp1	Comp2	Comp1	Comp2
Group 01	86.5	87.5	7.1	7.0	3.2	4.1	3.3	2.8
Group 02	79.7	79.4	6.7	6.8	8.1	7.8	5.6	6.6
Group 03	77.0	74.0	6.6	5.6	8.1	10.1	7.5	8.7

In a different line, Folse (2006) carried out an analysis to determine what type of writing exercise produced better results in L2 vocabulary retention. He maintains that given the low rates of vocabulary acquisition derived from reading, vocabulary learning connected to writing and to writing exercises in particular, deserves more attention. Folse's research involves 154 university level participants from various linguistic backgrounds, divided into three groups, who practiced target vocabulary in three types of exercise. The first group completed one fill-in-the-blank exercise; the second group completed three fill-in-the-blank exercises; and the third group wrote sentences containing the target words. The vocabulary Knowledge Scale (VKS), developed by Paribakht and Wesche (1979), was used as the pretest and posttest. The study concluded that the group of students that completed the three fill-in-the-blank exercises obtained significant differences (at the .0001 level) with a retention mean score (4.78) over students who wrote sentences (mean 2.39) and students who completed one fill-in-the-blank exercise (mean 2.18). Because the difference between the latter two exercises was not statistically significant, it cannot be determined which has a better effect on vocabulary learning.

Folse insists that the number of encounters and the time spent doing the exercises are key aspects of the three fill-in-the-blank exercises. First, the multiple encounters with the words in these exercises results in better processing derived from the multiplicity of retrievals involved. The time spent doing the three fill-in-the-blank exercises is about 50% longer than that spent on the other types of exercises. According to the author, fill-in-the-blank exercises demand deep processing and have proven to be very efficient in terms of student and teacher time required. Folse (2006) further maintains that while sentence writing is usually perceived as the best type of exercise due to the deep processing it requires, it may not be the most effective writing exercise for vocabulary learning. This type of practice, though, contributes to strengthening writing in general as well as syntax, spelling, and fluency improvement. In sum, the study points to the efficiency of fill-

in-the-blank exercises that provide opportunity for multiple encounters, hence multiple retrievals, for L2 vocabulary retention.

Lee and Muncie (2006) analyzed the written production of intermediate, secondary school, ESL students, ages 13-16, in Canada. These authors used the LFP (Laufer and Nation, 1995) which analyses and classifies written vocabulary and yields results of the percentages of the word families as belonging to the 1000, 2000, University Word List or *Not-in-the-list* levels. An analysis was made of three versions of compositions produced by the students in which students were expected to use vocabulary that had been encountered in readings and had received teacher explanation. Version 1 did not reflect productive use of the target vocabulary. After Version 2, students incorporated target vocabulary in their free style compositions. Version 3 showed target vocabulary use based on students' recall of this vocabulary; this indicated that the vocabulary was part of the productive store of the participants in a delayed test. The authors attribute the incorporation of target vocabulary to different factors. These factors may include the teacher's work on explaining the target words' meaning, multimodal exposure (through spelling, listening, reading, and writing) to the words, structuring of knowledge with the use of a frame for writing, the cumulative work on the progression of the three versions of the composition, and the specific instruction on including the target words in the written exercise.

The authors note that the increment in low-frequency words between versions 1 and versions 2 and 3 should be seen as a reflection of the benefits of reading, which in turn, should be considered a central classroom practice. According to these authors, it is essential to direct students' attention to vocabulary beyond that of the 2000 level (once this is mastered) with the intention of helping students' progress in the growth of SL vocabulary. This can be attained, to a large extent, through teachers' promotion of interactive practices in the classroom which help scaffold vocabulary knowledge to be used productively while improving the quality of the writing in the long run. The authors interpret that an increase in productive vocabulary implies an increase in passive vocabulary and is evidence that overall language proficiency is benefitted.

Research by Olinghouse and Wilson (2013) examined the role of vocabulary in three different genres (story, persuasive and informative). These written texts were based on the topic "outer space" and they were produced by elementary school children. These authors analyzed the vocabulary constructs of diversity, maturity, elaboration, academic words, content words, and register in the samples they studied. Their findings shed light on the importance of different uses

and various features of vocabulary for specific purposes in different genre types. After evaluation of factors affecting usage and the effect of vocabulary constructs on writing quality the authors discuss the following findings. The participants in this study adjust their vocabulary according to the type of genre they produce. Narrative or story texts have a richer diversity of vocabulary than informative texts and greater maturity than persuasive texts. Persuasive texts offer more diversity than informative texts and a higher register than narrative and informative texts. Finally, informative texts contain more content words and show more elaboration than stories and persuasive texts, and more maturity than persuasive texts.

Regarding the effects that vocabulary constructs exert on the quality of writing, Olinghouse and Wilson (2013) report the following findings. First, for story texts, the unique predictor construct of vocabulary diversity explains 8.4% of the variance in the quality of the written piece; maturity, on the other hand, explained only 1.1% of this variance. For persuasive texts the constructs of content words and register were both significant, unique predictors explaining 7.6% and 5% of the variance of writing quality respectively. Finally, for informative texts, the constructs of content words and maturity were unique predictors with a striking predictive value of 31% for content words and of 3.4% for maturity. In short, these results show that learners vary the type of vocabulary they use depending on the genre that they produce. The fact that the authors used the same topic across the different genre types helped them control for background knowledge. With this premise and the above results, these authors claim that both genre knowledge and topic knowledge serve as bases that students must access to come up with the vocabulary required to write a text. Furthermore, for these authors, the results indicate that the students' variation in the selection of vocabulary according to genre and the salience of certain constructs of vocabulary contribute in different ways to general writing quality. These authors further argue that these results may also indicate the participants' lack of knowledge of academic vocabulary and would call for explicit teaching and direct instruction to students in the use of this type of vocabulary.

In a different vein, Velasco and García (2014) offer a study that demonstrates the usefulness of translanguaging as self-regulated learning in the writing process. In their investigation, these authors found that students use metacognitive strategies such as self-regulation during translanguaging. In line with what was discussed in section 1.1.1.2, regarding the role of Cummins' (2000) Interdependence Hypothesis, Velasco and García suggest that even emergent bilinguals make use of their linguistic and experiential knowledge to attain

understanding and develop critical thinking. They suggest that the addition and integration of novel linguistic means should be anchored in students' previous knowledge. They point out that the translanguaging strategies that students implement reflect bilinguals' personal language practices. Moreover, the authors note that for translanguaging practices to be effective, they must be implemented in a meaningful context, which, like writing does, offers the opportunities for linguistic processing and writing development.

Velasco and García (2014) argue that learners, having specific goals in mind, implement specific practices to achieve these goals. Translanguaging is used as such a practice by students who rely on L1 knowledge to expedite the process of language learning. These authors argue that translanguaging practices are used at all levels of the writing process, during planning, editing and production. Velasco and García (2014) list 5 indicators of translanguaging use that were identified in the study. First, during planning, translanguaging is present in multimodal (using drawing and writing) and a multilingual repertoire, where both L1 and L2 languages are reflected. Second, again during planning, vocabulary expressed in the L1 was glossed with L2 glosses to be used later during writing. Third, during drafting, a multilingual repertoire, in the form of L2 internal dialogue, was present in a draft written in the students' L1. Fourth, also during drafting, word retrieval and transformation in the form of meaning change, moving from L1 to L2 to secure meaning, were identified. Translanguaging was found in the final product as a form of rhetorical engagement to establish a connection with the reader. The authors insist that translanguaging, sometimes serving as a scaffolding technique and sometimes found as an engagement strategy for effectiveness, is used by students with the aim of solving challenges in language comprehension and production in a unique and creative manner.

Harrison et al. (2016) analyzed the cognitive and linguistic components associated with spelling and writing in third grade, ESL and native-English speaking children. These authors' findings report ESL children in disadvantage in terms of their measures of vocabulary, syntactic awareness and decoding fluency. On the other hand, for factors such as rapid naming, phonological awareness, verbal short-term and working memory, reading fluency, single-word spelling, text spelling, handwriting fluency, paragraph writing fluency and writing quality, the participants show similarities in their results. These authors claim that phonological awareness serves as a predictor of single-word spelling for native students while phonological awareness and rapid naming predict this same trait for ESL learners. For ESL individuals, transcription was associated ($r = .35$), in moderation, with the

content and structure of these students' written paragraphs and with their total writing performance (r = .45). Rapid naming (r = .25) and syntactic awareness (r = .30) are predictors of the writing quality reflected in the content and structure of the ESL participants, accounting for 14.7% of the variance in their writing. According to Harrison et al. (2016), for the native speaker participants, syntactic awareness and oral vocabulary were identified as the predictors to account for 26.5% of the variance in writing quality. In the case of ESL participants, transcription was the predictor associated with overall writing achievement for ESL learners when transcription, vocabulary, and syntactic awareness all predicted overall writing ability for native speaking students.

Harrison et al. (2016) argue that these results suggest that ESL students resort to a different subset of skills to catch up with the level of writing reached by their native-speaking counterparts. They propose this explanation because the oral language skills of ESL students are not as developed as those of native-speaking participants, and yet they reach similar results in several areas analyzed and are thus undifferentiated from their native speaker counterparts. The authors find that ESL students are drawing on grammatical structure knowledge of the L2 as an important component for writing. They argue that, in line with Cummins (1980), although their syntactic knowledge is underdeveloped when compared with that of the native-speaker participants, this knowledge may have reached a threshold of linguistic competence (through rapid lexical retrieval) in L2 that enables their writing in the second language to access word-specific information and it thus develops despite their disadvantages in oral and syntactic knowledge.

As seen in section 2.3, the CLIL phenomenon in Europe has offered a new perspective on the teaching of second languages. In the areas of vocabulary and writing, several authors mention thought-provoking findings. Jexenicker and Dalton-Puffer (2010) discuss the results of a comparative study involving Austrian students studying under CLIL provision as compared to students in EFL teaching. These authors claim that the CLIL students show clear advantages in the areas of lexico-grammar, range and accuracy of vocabulary and orthographic correctness (as opposed to areas requiring improvement such as discourse competence and textual organization). As regards vocabulary, CLIL students are said to use fewer high frequency words, they present more lexical variation and use longer words. Jexenicker and Dalton-Puffer (2010, p. 181) claim "the greatest advantages of CLIL students in terms of writing skills result from their greater general language ability and also a greater awareness of the pragmatic demands of the task."

Along these lines, Ruiz de Zarobe (2010) carried out a longitudinal study involving students from CLIL and EFL programs. This author attributes CLIL students with an increasing advantage in the written production categories of content, organization, vocabulary, language use and mechanics. The differences favoring CLIL students are significant for content and vocabulary. She interprets these results as an indication of the advantages that a more extended amount of exposure has on the proficiency of writing in a foreign language, particularly when it was identified that this improvement increases as students advance in their school grade. The author further mentions that CLIL students are three years younger than their non-CLIL counterparts, noting that CLIL settings provide a faster rate of second language acquisition.

In another study, Gené-Gil, Juan-Garau and Salazar-Noguera (2015) investigated longitudinal progress in writing for CLIL students. They determined that at the level of language competence CLIL students demonstrate better complexity, accuracy and fluency. They explain how this reveals that these students' texts progressively become more complex in terms of syntax and lexicon, as reflected in better accuracy and fluency. From a more holistic perspective, the authors conclude that CLIL students also have significant differences at the levels of organization, vocabulary and language use.

We believe that these last three studies confirm the importance of classroom methodology and its role in having a direct impact on the development of writing skills in a second language in general, and especially on vocabulary gains. The fact that the vocabulary used in written production is identified as one of the aspects that is most positively affected by CLIL instruction sheds light on the importance of using the target language for meaningful communication. The advantages that this brings to overall improvement in second language proficiency, and to vocabulary development in particular are made evident in the previous studies. While Content Based Instruction is not considered as a CLIL practice *per se*, it does put into practice the notion of teaching content through the means of the target language. This entails greater exposure to the target language, which for authors such as Ruiz de Zarobe (2010) results in better content and vocabulary use in writing. It has previously been observed that methodological practices such as CLIL produce identifiable advantages in the student population. Part of the purpose of the present study is to determine whether these findings can be made extensive to the CBT institutions in the Costa Rican context.

9.2. THIS STUDY: WRITING RESULTS

This section aims at determining how much of the knowledge that students have acquired vocabulary-wise is brought into communicative use when students write. By measuring the quality of vocabulary, we attempt to determine the richness of the lexicon in their compositions.

9.2.1. Participants

The number of participants in this task is the lowest across the study. A total of 177 (52 from the CBS and 125 from the FLS) completed this writing activity. In the CBS group 20 were male and 32 were female. In the FLS group, there were 62 male and 63 female students. The missing data corresponds to students who were either absent on the day the test was administered or who did not complete the task.

9.2.2. Instrument and procedure

The students wrote a composition according to the following guidelines. They were given one hour to complete their composition. They were not asked to write a specific number of words, but were encouraged to write as much as they could on the topic of their choice. They could choose between three topics, all of which were general in nature and did not require any specialized expertise. They could use their general background knowledge and express their opinions or points of view. Having a choice of topic maximized the students' interest in the ideas they put into writing. The students could select one of the following topics:

Topic 1. What could a foreigner find and enjoy if he visited Costa Rica? Discuss the things and places that Costa Rica has to offer to visitors.

Topic 2. Should a government be allowed to limit the number of children a family can have? Discuss this idea considering basic human rights and the danger of population explosion (Laufer & Nation, 1995).

Topic 3. "A person cannot be poor and happy, because money is always needed to gain something that is important to that person." Argue for and against this idea (Laufer & Nation, 1995).

9.2.3. Data processing

Every composition was later digitized as a *Word* document to be analyzed using the program *AntWordProfiler*, version 1.4.1 (Anthony, 2014, available at http://www.laurenceanthony.net/software/antwordprofiler/). To be processed, the compositions were turned into *.txt* files and then processed in the *Vocabulary Profile Tool* of the *AntWordProfiler*. The *Vocabulary Profile Tool* provides vocabulary statistics and frequency information about each one of the compositions. In this tool, the compositions are compared against three vocabulary lists: the first thousand words, the second thousand words and the 570 words that make up the Academic Word List (Nation, 2013). Additional results are presented for words *Not-in-the-list*. The latter represent words belonging to higher categories of word lists, including proper names or non-English words. The tool provides percentages of words in each word list (as described by Nation, 1993, 2013; and Nation and Waring, 1997), and gives percentages of words as grouped according to word families (Bauer and Nation, 1993).

AntWordProfiler was built following the idea of Nation's Range Program (Nation, 2006a; available at <https://www.victoria.ac.nz/lals/about/staff/paul-nation>) designed to generate vocabulary statistics based on word lists and their frequency. It also contains features like those of the Lexical Frequency Profile (Laufer, 1994, Laufer and Nation, 1995). The *AntWordProfiler* 1.4.1 is a much more modern version of RANGE. It is a program that has undergone several revisions and improvements and contains many features that are not in RANGE. It does, however, use as a base the word lists and some of the basic features offered by Nation in RANGE; for example, it uses the word lists created by Nation for the RANGE tool. In *AntWordProfiler*, the content of the composition is divided depending on the different word levels that the words belong to. For our study, words will be divided as belonging to the first 1000, second 2000, Academic Word List and *Not-in-the-list* bands. In this analysis, the measure concentrates directly on lexis; and the analysis is almost entirely computerized. Except for the preparation of the students' compositions to be analyzed as computer input, the remainder of the analysis and the profiling of each case are done by the program.

In *AntWordProfiler*, the output of the analysis indicates the number of word types and their percentages in the composition. Word types are understood here as a unit of counting that represents a single word form; if the words "problem" and "problems" appear in the composition, they are counted as two types. A word

family, on the other hand, groups together words whose form and meaning are related. In that case, "problem" and "problems," for example, would be considered members of the same family (Bauer and Nation, 1993). Figure 18 presents an example of some of the results as shown in the output window of the *Vocabulary Profile Tool* in the *AntWordProfiler*. The statistical summary gives the total numbers and percentages of tokens, types and word families for each one of the word lists at the 1000, 2000, AWL levels as well as the *Not-in-the-list* items. These results are followed by an analysis of range and frequency of each one of the words in the composition. After the statistical summary, the program produces detailed lists of the actual words that belong to first 1000 words, the second 1000 words, the *Academic Word Lis*t and *Not-in-the-list* items in each one of the compositions. Once the compositions are analyzed in the *AntWordProfiler* program, descriptive statistics are calculated and differences between schools and gender groups are compared by using SPSS Version 20 (IBM Corp, 2011). An example of the output of *AntWordProfiler* is presented in Figure 18.

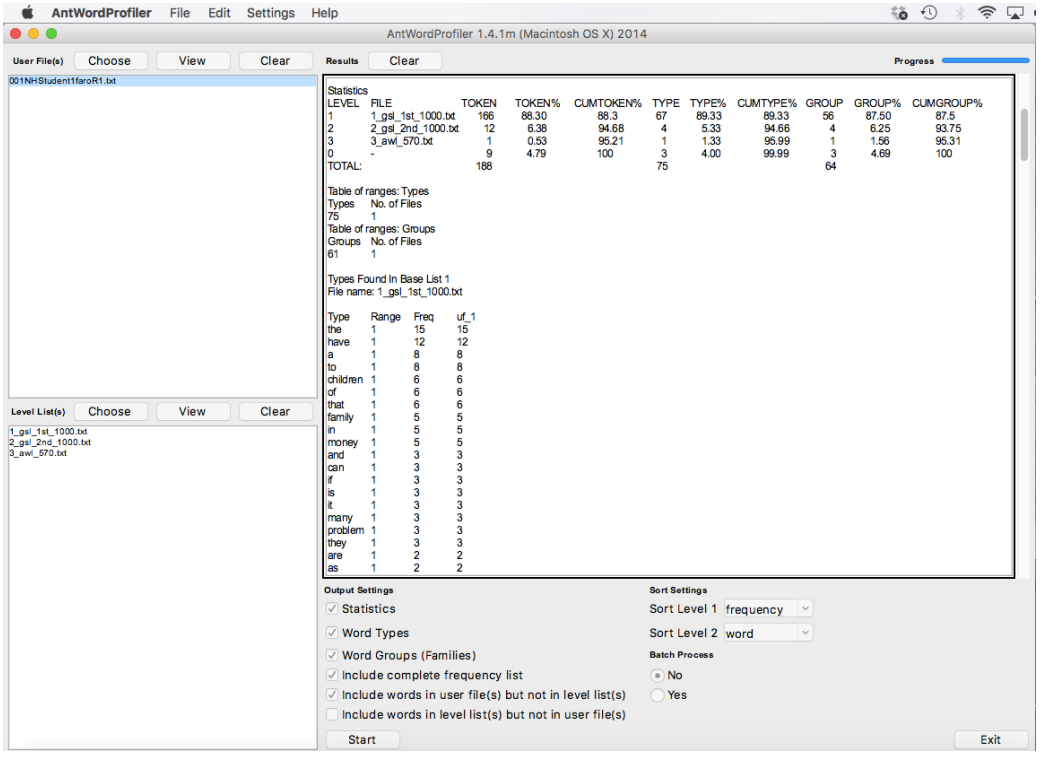


Figure 18. Sample output results from the *AntWordProfiler* program.

9.2.4. Results

In the following sections we will analyze different aspects of the written compositions of the sample population. These include topic, total word counts, word type and family, and gender results analysis.

9.2.4.1. Topic choice analysis

Regarding the students' choice of topics for the composition, the distribution is as follows: For the CBS, 50.9% chose Topic 1; 10.9%, Topic 2; and 32.7%, Topic 3. Three students did not work on the composition task; that makes up for the remaining 5.5%. In the FLS, 58% chose Topic 1; 7.6%, Topic 2; and 29.8%, Topic 3. In this case as well, 4.6% represents missing data. It follows from this distribution that most students felt more comfortable with Topic 1, next with Topic 3, and last, with Topic 2. Laufer and Nation (1995) argue that the topic is one intervening factor that may affect lexical richness in writing.

The percentage of participants following each topic in each school is also very similar. This distribution is illustrated in Figure 19.

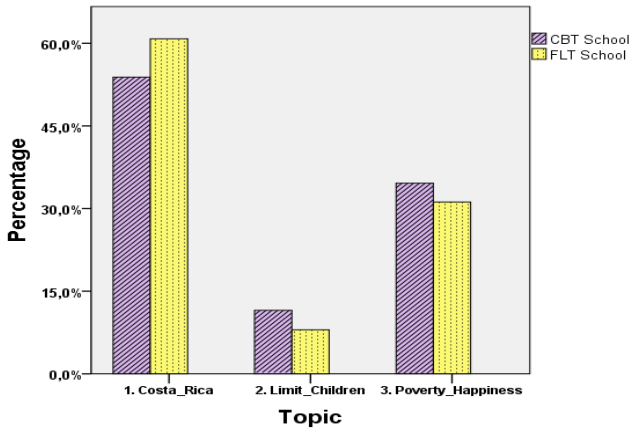


Figure 19. Topic choice per school in written composition.

As discussed above, it was important to give students a varied set of topics so that they would feel comfortable in making a choice and writing about something that felt familiar to them. It is reasonable for the students' main choice to have been topic 1, in which they were asked to describe their own country. Many of them resorted to describing places they had visited or activities that they had done in previous vacation periods, reflecting their personal experience. In general, the distribution of topics has a very similar division in both schools. The preference of topics chosen is the same: the first topic of choice is topic 1; the second, topic 3; and the last, topic 2.

9.2.4.2. Total word count analysis

Table 40 displays data on the total word count for each school, including the mean, median, and standard deviation (SD).

Table 40. Descriptive statistics for words in composition per school

	CB School	**FL School**
Median	248.50	189
Mean	260.77	195.64
Max words	473	467
Min words	70	27
SD	90.67	84.23

The CBS has a mean of 260.77 (SD 90.67) words while that of FLS is 195.64 (SD 84.23). The maximum number of words for the CBS is 473; for the FLS it is 467, only a slight difference. The minimums, however, are 70 words for the CBS and as low as 27 for the FLS. These results are illustrated in Figure 20.

The summary in Table 40 and the representation in Figure 20 show that students in the CBS use more words and have a higher minimum number of words in comparison with those of the FLS students. The gross data for the CBS seem to concentrate between 170 and 400 words. For the FLS, on the other hand, the data start to cluster at around 150 and up to around 275 words; in both sides of this cluster the numbers of words drop sharply. This spread of words and the fact that

the word minimum falls below 30 words causes the mean and general distribution of the data to fall below the figures for the CBS.

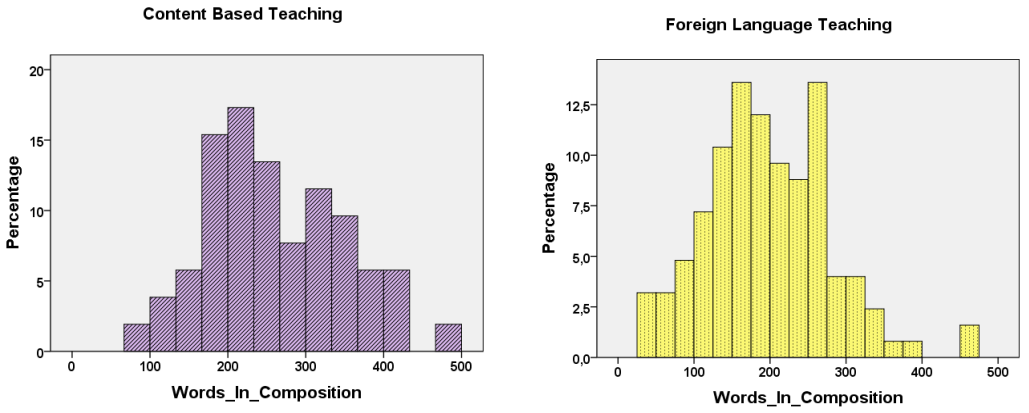


Figure 20. Distribution of words in composition per school.

The box diagram in Figure 21 offers additional information about these results.

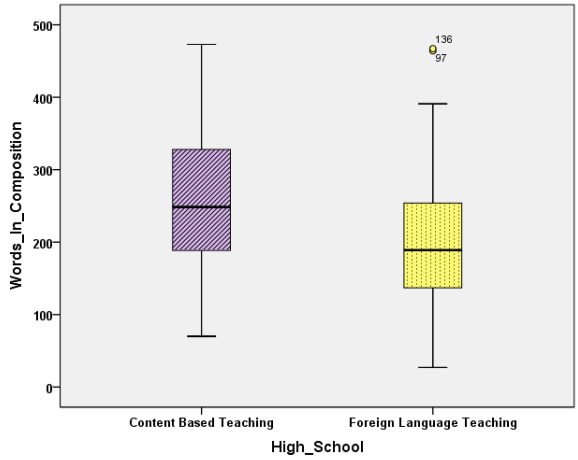


Figure 21. Box plot distribution for words in composition per school.

The median scores show a higher mean for the CBS, as described for the total word counts above. The distribution is slightly asymmetrical in relation to

their median. In both cases it seems to be slightly positively skewed with this slight tendency toward the upper levels of the scale. The whiskers are also slightly asymmetrical, with their distribution slightly pointing toward the upper levels of the scale. The two outliers in the FLS group are identified with their subject numbers 97 and 136.

Table 41. Test to determine normality distribution in words per composition

	High School	**Kolmogorov-Smirnov**		
		Statistic	**df**	**Sig.**
Words in Composition	CB School	.108	52	.187
	FL School	.049	125	.200

To determine normality assumptions, a Kolmogorov-Smirnov test was applied to the data. The results in Table 41 show that the data follow normal distribution for both schools.

Since the data show that both schools follow a normal distribution, the Independent sample *t*-test is used to determine the nature of those differences. The test shows statistically significant differences favoring the results of the CBS. This conclusion was reached through the analysis of the following results of the *t*-test. The CBS (n = 52) was described as having superior word counts with a mean of 260.77 (SD 90.67). By comparison, the FLS (n = 125) showed a lower numerical word count with a mean of 195.64 (SD 84.23).

To test the hypothesis that the CBS and FLS students actually showed statistically significant different word counts, a *t*-test was performed. The Kolmogorov-Smirnov test serves as an indication that the data follow a normal distribution. The CBS and FLS distributions were sufficiently normal for a *t*-test to be performed; statistical descriptors indicate that their skew is less than 2.0 and their kurtosis is less than 9.0. In addition, Table 42 shows how homogeneity assumptions were tested and satisfied via Levene's F test. Equal variances assumed, in this test $F_{[175]} = .97$, $p = .324$; where p is evidently greater than .05. The independent *t*-test is thus associated with a statistically significant effect, $t_{[175]} = 4.581$, $p = .000$. This means that the CBS word counts are indeed associated with a statistically-significant, larger, mean word count than those of the FLS.

Table 42. Independent *t*-test to determine characteristics of word counts in composition.

		Levene's Test for Equality of Variances		***t*-test for Equality of Means**				
		F	**Sig.**	***t***	**df**	**2-tailed Sig.**	**Mean difference**	**Std Error Difference**
Words in Composition	Equal variances assumed	.976	.324	4.581	175	.000	65.129	14.219

In sum, the statistical tests demonstrate that CBS students produce significantly more words per composition than FLS students do; the numerical advantages displayed in Table 40 represent statistically significant differences.

9.2.4.3. Word type and family analysis

Further analyses conducted through descriptive statistics shed light on the characteristics of the types of words (and the family groups they belong to) that the participants have used. The *AntWordProfiler* provided estimates of the results of each school in terms of the word-type percentages at the 1000, 2000, AWL and *Not-in-the-list* word levels, as well as the families for each of these words. Table 43 shows these results and Figure 22 demonstrates the distribution *per* word type and family. Results were further analyzed using SPSS 20.

Table 43. Independent *t*-test to determine characteristics of word counts in composition

Word Type		**School**		**Word Family**		**School**	
		CB	**FL**			**CB**	**FL**
Type-1000	Mean	75.60	73.59	Fam-1000	Mean	73.31	71.56
	SD	9.82	9.99		SD	10.02	10.28
Type-2000	Mean	6.24	5.88	Fam-2000	Mean	6.76	6.13
	SD	2.35	2.83		SD	2.52	2.83
Type-AWL	Mean	2.78	3.15	Fam-AWL	Mean	2.97	3.37
	SD	1.70	1.68		SD	1.75	1.79
Off-the-List	Mean	15.38	17.38	Off-the-List	Mean	16.95	18.94
	SD	9.45	9.57		SD	10.02	10.09

Initially we do not observe outstanding differences in the distribution of the types of words (and families) used by students from these schools. Both prefer words belonging to the first 1000 word type. The number of words belonging to the second 1000 and AWL word types is much more reduced. The *Off-the-list* or *Not-in-the-list* represents the second largest group of word types and families for both schools. This group includes proper names, non-English words and words that might be part of higher vocabulary lists. Figure 22 provides a graphic distribution of *types* and *family* words per school.

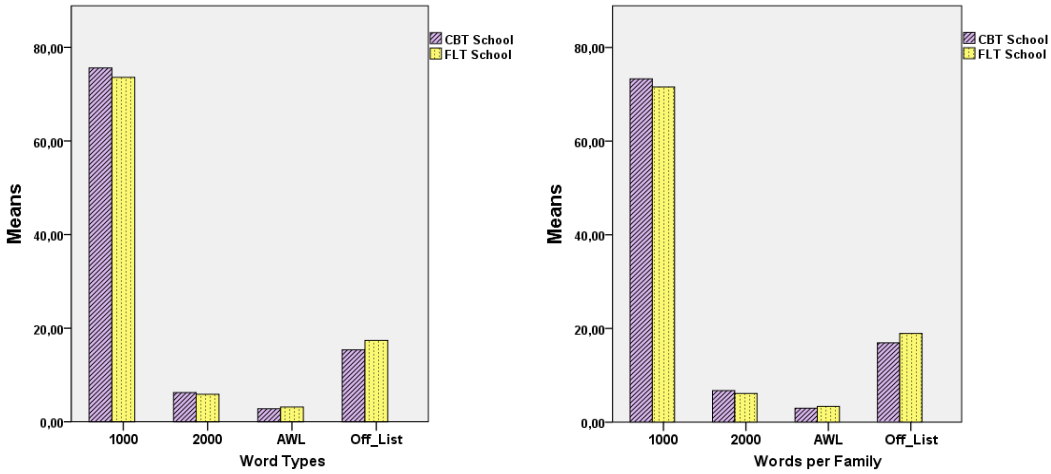


Figure 22. Word distribution according to Word Type and Word Family.

A Kolmogorov-Smirnov test was used to determine normality assumptions (see in Table 44). Here, word types 1000, 2000, and AWL follow a normal distribution; that is the case only for the FLS (not for the CBS) in the 1000 word band. The *Not-in-the-list* words, on the other hand exhibit a not-normal distribution.

Table 44. Parametric test to determine distribution of type of words

	Kolmogorov-Smirnov							
Descriptive Statistics	**Word_Type_1000**		**Word_Type_2000**		**Word_Type_AWL**		***Off-the-List***	
	CB	FL	CB	FL	CB	FL	CB	FL
Statistics	.182	.082	.120	.069	.117	.079	.160	.193
df	47	105	47	105	47	105	47	105
Sig.	.000	.078	.088	.200	.111	.099	.004	.025

Except for the results for *Not-in-the-list* words, the CBS shows superior numerical mean differences in all word types. With the following tests, we want to determine whether these seemingly superior numerical differences for the CBS are statistically significant. A Mann-Whitney U test was run to determine the types of difference between the use of words at the 1000 level between the two schools. The test shows that there are no statistically significant differences between the two schools at this word level. The summary of the results of this test is displayed in Table 45.

Table 45. Results of Mann-Whitney U test for the type 1000 words

N = 177	**Off-the-list-words**
Mann-Whitney U	3087.000
Wilcoxon W	10962.000
Error	310.509
Z	-.525
P (two-tailed)	.600

Given that at the word levels 2000 and AWL, the schools follow a normal distribution (p = >.05), a *t*-test was performed to determine the characteristics of the differences between these two schools. The homogeneity of variances is satisfied through a Levene's test. Table 46 shows the results of the Levene's test and the *t*-test for the word types 2000 and AWL. According to the *p* value of the Levene's test, we assume equal variances for the 2000 word types and for the AWL. The results of the *t*-test indicate no statistically significant differences in the production of these two types of words.

Table 46. Levene's and *t*-test for the 2000 and AWL word types

		Levene's test		***t*-test**		
		F	**Sig.**	**t**	**df**	**2-tailed Sig.**
Type_2000	Equal variances assumed	.376	.541	1.451	172	.149
Type_AWL	Equal variances assumed	.000	.995	- 1.397	152	.165

The 2000 word type was described as having a superior mean of 6.243 for the CBS. On the other hand, the FLS has a lower numerical mean of 2.352 (*SD* 2.831) for the 2000 level word type. On the contrary, the AWL showed a slight numerical superiority for the FLS at 3.148 (*SD* 1.685) and a mean of 2.773 (*SD* 1.702) for the CBS. To test the hypothesis that the 2000 and AWL word types

showed statistically significant different word counts, a *t*-test was performed. Table 46 shows how homogeneity assumptions were tested and satisfied via Levene's *F* test, first for the 2000 word type $F_{[172]} = .376$, $p = .541$ and, then, for the AWL, $F_{[152]} = .000$, $p = .995$; where p is evidently greater than .05 in both cases. The independent *t*-test is thus associated with non-significant statistical effects, $t_{[172]} = 1.451$, $p = .149$ for the 2000 word type; and $t_{[152]} = -1.397$ $p = .165$ for the AWL word type. This indicates that the word types at the 2000 and AWL are indeed associated with a non-statistically significant mean for the CB and FL schools. This indicates that the apparent numerical differences are not statistically significant.

The *Not-in-the-list* word type was also analyzed to determine the type of statistical differences that it represents. Table 47 shows the results for the Mann-Whitney U test that was applied to determine the nature of its characteristics. Like the other word types analyzed, the results show non-statistically significant differences for this word type.

Table 47. Results of Mann-Whitney U test for the Off-the-List-Words in composition

N = 177	**Off-the-list-words**
Mann-Whitney U	3517.500
Wilcoxon W	11392.500
Error	310.509
P (two-tailed)	.389

As can also be observed in Figure 22, *Word distribution according to Word Type and Word Family*, the behavior of word family distribution goes hand in hand with that of word type distribution. When statistical tests are performed to compare the results of word production according to family for each school group, we find that these do not have statistically significant differences either. In sum, there are no statistical significant differences in terms of the word types or word families used by the students sampled in this study.

9.2.4.4. Gender results

The outcomes are analyzed to determine the characteristics of the results from the gender perspective. Table 48 offers a summary of the results for gender

groups. We will begin by taking a look at the general features of these data from a gender distribution. The word count for each student group is analyzed first.

Table 48. Distribution based on words per composition per gender group

	CB School		FL School	
	Boys	**Girls**	**Boys**	**Girls**
N	20	32	62	63
Mean	219.55	286.53	199.90	191.44
Median	220.50	270.50	184.50	197.00
SD	68.915	93.996	86.429	82.501
Skew	-.351	.235	.856	.096
Kurtosis	-.245	-1.260	1.376	-.415

Based on the information in Table 48, word counts in relation to gender show that the girls' mean of 286.53 is higher than the boys' mean of 219.55 in the CBS. The SD for the boys is more concentrated (68.915) than that of the girls (93.996) in this school. On the contrary, the boys' mean is higher 199.90 than the girls' mean 191.44 at the FLS, although only slightly. This is also the case for the boys' SD, which is slightly more spread (86.429) than the SD for the girls (82.501) in this school. At first glance, from this preliminary analysis we would not be able to anticipate a pattern in connection to the effect of the gender variable, as different genders present greater advantages in each school. The distribution of word counts and their spread are illustrated in Figure 23. It can be seen that the FLS has a couple of male outliers. Also, in the case of the CBS girls, but also for the boys, the CBS students' median is well above that of FLS students.

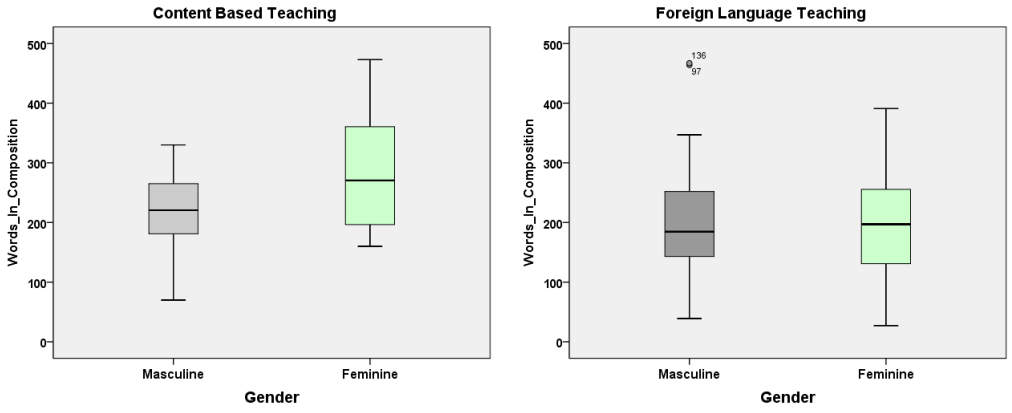


Figure 23. Word distribution per gender group in each school

For each of the schools, we will analyze the results according to gender groups with the purpose of describing any possible identifiable pattern that may be associated with this variable. The results are first analyzed at the intraschool level and then at the interschool level. A Kolmogorov-Smirnov test was implemented to establish the normality of data distribution for each gender and school. This test reflects normal distribution for both gender groups in both schools as observed in Table 49.

Table 49. Normality tests words per school per gender

		Kolmogorov-Smirnov		
School	**Gender**	**Statistic**	**df**	**Sig.**
Content Based	Masculine	.094	20	.200
	Feminine	.148	32	.073
Foreign Language	Masculine	.096	62	.200
	Feminine	.059	63	.200

The normality of the distribution in the data calls for a parametric test analysis. The *t*-test was used to determine the type of differences in the data. According to this test, certain statistically significant differences favor female students over male students in the CBS. No statistically significant differences are found among gender groups for the FLS. The summary of the results for this test is presented in Table 50.

Table 50. Levene's and *t*-test for word counts according to gender in each school

			Levene's test		***t*-test**		
			F	**Sig.**	***t***	**df**	**2-tailed Sig.**
Words in composition	CBS	Equal variances not assumed	5.931	.018	-2.956	48.600	.005
	FLS	Equal variances assumed	.047	.829	.560	123	.577

In the case of the CB school, female students obtained a higher word count means in their composition 286.53 (SD 93.996) in comparison to boys who obtained a mean of 219.55 (SD 68.915). Homogeneity in variances is assumed and tested with Levene's *F* test, $F_{[48.600]} = 5.931$, $p = .018$. In this case, as p is not greater than .05, we can conclude that there are statistically significant differences favoring

female students in the CBS. For the FLS, male students produced a slightly higher word count means in their composition 199.90 (SD 86.429) in comparison to girls who obtained a mean of 191.44 (SD 82.501). In addition, for this school homogeneity in variances is assumed and tested with Levene's *F* test: $F_{[123]} = .047$, $p = .577$. For this school setting, given than *p* is greater than .05, we assume that there are no statistically significant differences in the data. In sum, the tests applied to the data to determine whether certain differences may be associated with the variable *gender* show that statistically significant differences indicate that female students outscore male students in the CBS, whereas no gender-related differences are found in the data analyzed for the FLS.

Further studies were conducted to examine the possibility of gender differences influencing results across schools. To do this, the results at interschool level were analyzed. The descriptive statistics associated with student genders and schools are reported in Table 48 above. It can be seen that the lowest word count in the composition was associated with the numerically smallest mean for girls (M = 191.44), followed by boys (M = 199.90) in the FLS. The highest word production was associated with the numerically highest word count in the composition for girls (M = 286.53) and then boys (M = 219.55) in the CBS. To test the hypothesis that the students' gender affected their performance in the composition for each school, a Two-Way ANOVA was performed.

Prior to conducting the ANOVA, we evaluated whether the different assumptions were satisfied. First, the independent variable has two or more categories. Because we are testing two genders in each of the two schools, this assumption is met. We also satisfy the requirement that the dependent variable (number of words in composition) is measured at the scale level. We further need the distribution to be at least approximately normal; the Kolmogorov-Smirnov test indicates that this is so. The assumption of normality was satisfied as all four gender groups' distribution have an absolute skew value < 2.0, and an absolute kurtosis value < 9.0. Furthermore, the assumption of homogeneity of variances is satisfied based on Levene's *F* test, $F_{[3, 173]} = 1.602$, $p = .191$. Having satisfied all of the requirements, we proceeded with the ANOVA test (Table 51).

The two-way ANOVA with between-subjects interaction yielded statistically significant effects that can be observed in Table 51.

Table 51. ANOVA for words in composition in connection to gender and school

	Type III Sum of squares	df	Mean square	F	Sig.	Partial η^2
Model	**213227.067**	**3**	**71075.689**	**9.902**	**.000**	**.147**
Intercept	7111382.789	1	7111382.789	990.719	.000	.851
Gender	30241.213	1	30241.213	4.123	.042	.024
High School	116234.429	1	1162234.492	16.193	.000	.086
Gender*High School	50252.498	1	50252.498	7.001	.009	.039
Error	1241793.894	173	7177.999			
Total	9619655.000	177				
Total corrected	1455020.960	176				

Both the model and the intercept factors demonstrate the robustness of the test. The results show statistically significant effects in terms of gender, $F_{[3, 173]} = 4.213$, $p = .042$, $\eta^2 = 0.24$; for high school, $F_{[3, 173]} = 16.193$, $p = .000$, $\eta^2 = .086$, and for the interaction gender and high school $F_{[3, 173]} = 7.001$, $p = .009$, $\eta^2 = .039$. This indicates that the variable gender has the smallest effect, although still significant, for word production (higher for female students). The variable High School (favoring the CBS) has the strongest effect on the number of words that students produced followed by the interaction from the combination of the variables of gender and school (higher for female CB students). Thus a null hypothesis of no difference between the means was rejected, and this model can account for a variance of 14.7% of the total word counts for a composition showing that both gender differences and gender and school differences favor the CBS related to the total number of words produced in the composition. Figure 24 provides a visual depiction of the means in relation to gender and school distribution.

The figure clearly depicts the marked difference between the gender groups in the sample. From an intra-school comparison, the difference is not as marked between male and female students in the FLS as it is between male and female students in the CBS. We note once again that while female students produce more in the CBS, it is male students who produce more in the FLS. From an interschool perspective, a striking difference is evident between female students in each school. The difference in the number of words produced by female-CBS students is much more noticeable in this figure. There is also a clear difference between male participants between the FLS and the CBS, although this is not as marked as that observed between female students across schools.

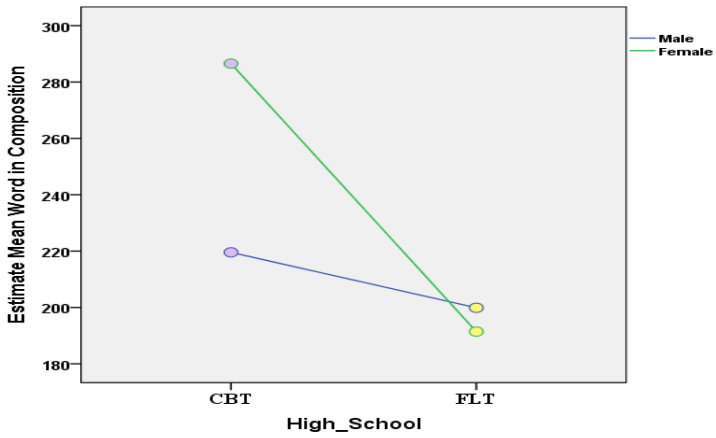


Figure 24. Relation between mean word count and gender per school

As mentioned above, however, all differences between gender groups and school groups are statistically significant, with higher word counts for female students in the CBS. Overall, it has to be noticed, however, that the variable *school* is the one that exerts the strongest influence in the results. While it is true that female CB students do exhibit a statistically significant advantage in terms of word counts, this advantage may as well be boosted by the school context where these students are enrolled. It seems to follow that female students are benefitting the most out of the context of CBS.

9.2.5. Discussion

The total number of words produced by students in the compositions already serves as an indicator of the degree of variation within schools in the sample. While some students produce close to 500 words in their compositions, others produce as few as 70 or 27 words. In their study, Laufer and Nation (1995) required between 300 and 350 words from their students and allotted the same amount of time as we did in the present research (one hour). In their analysis, they indicate they used the first 300 words per composition and left out the rest of words because they considered that this amount of words would yield stable results. The students in the present study were not asked to produce a specific number of words. That decision is justified by the notion that both instructors and learners admittedly recognized that the writing skill receives very limited attention in the curriculum and it is thus

limited. Instead, the students were encouraged to write as much as possible during the time that they were given, with the intention of not discouraging those students who would not have been able to reach the 300-350 mark, had that been required.

An analysis of the mean for each school shows that many of these students, indeed, would not have been able to reach the 300-350 word mark as for both schools the means fall below the lower figure of this range: 260.77 words for the CBS and 195.64 words for the FLS. With these results, the lower score is still close the 200-word mark that Laufer and Nation (1995) identify as conducive to a stable profile. We have to keep in mind however, that that mean score indicates that a good number of students are well below producing that number of words. Once we continue the analysis of the results, we see that not only are students from the CBS writing more words, but also their spread of word distribution is much more concentrated than that of the FLS, thus indicating greater word count variability in the FLS sample. The statistical *t*-test confirms that the numerical difference in the amount of words produced by each group of students is, indeed, statistically significant, favoring the writing performance of students in the CBS.

A closer analysis of the types of words that students are producing, however, yields noteworthy results. According to Laufer (1994), as students' progress in the development of their writing skills, the percentages of the first and second 1000 words decrease and the percentages of low frequency vocabulary (AWL and upper level vocabulary words) increase. Based on the results in the present study, we confirm that these groups of learners are in the earlier stages of vocabulary development. The results of their writing show that most of their vocabulary is still concentrated within the most frequent 2000 words (81.84% for the CBS, 79.47% for the FLS) and even more particularly so on the first 1000 words of English (75.60% for the CBS, 73.59% for the FLS). The percentages of academic words are still low: 2.78% for the CBS and 3.15% for the FLS. It would be expected, then, that as students continue progressing in the development of their writing skills, the pattern would be inverted, and the high frequency, percentages of basic words would decrease as the low frequency and AWL would increase. The results indicate, though, that these learners have a long way ahead of them in this process.

One interesting finding is based on the *type* and *family* of the words used by the students, which are not statistically significantly different for the schools in this population sample. That is, the advantage observed in the production of the students from the CB school is limited to producing more words from the same word band that students from the FLS are using. We could conclude, then, that students from the CBS do not possess higher levels of vocabulary than students from the FLS.

They do, however, use the low-level vocabulary bands to a greater extent than the FLS students do. It is again interesting to observe that the FLS students have produced slightly more words from the AWL than the CBS students have: 3.15% and 2.78% respectively. Although non-statistically significant, these results also hint at the possible influence exerted by the methodology that each school implements. While neither school gives specific emphasis to the writing skill, the CBS students carried out this activity more easily. They seem to be making a more profitable use, in terms of production, of the primarily receptive training that they receive. As noted above, while the CBS students do not seem to have advanced to acquire higher levels of vocabulary, they do use the high frequency levels that they know for writing purposes. This advantageous use of vocabulary in writing associated with the CBS students is limited to producing greater numbers of words but still from the same word band (1000 and 2000 levels) that the FLS students are using in writing. A similar trend is found in CLIL contexts. In such contexts, where fluency is emphasized over accuracy, students tend to produce more than in other contexts (Ruiz de Zarobe, 2010; Gené-Gil and Salazar-Noguera, 2015).

From a gender perspective, the results in this study, in line with general findings in the literature, show slightly mixed results. Whereas regarding word counts female students produce more total words in the CBS, it is male students who produce a rather higher number of words in the FLS. While there are no gender-based statistically significant differences among students in the FLS, female students outperform male students in the CBS with statistically significant differences. At interschool level, female students show this statistically significant advantage only at the CBS, and female students at the FLS do not enjoy this superiority when compared to either one of the other groups in the sample. Once more, although the test shows statistically significant differences in favor of female students, this is the case only in one of the school settings. These results suggest that the female gender cannot be identified as the sole predictor of better written production in this sample.

Overall, the results indicate that greater attention should be given to the teaching and learning of vocabulary for students to be able to use the target language for communicative purposes. Whereas students in the sample appear to use words from the first 1000 words to similar extents, they certainly do not handle the second 1000 word level well enough to employ this knowledge for effective communication, especially students from the FLS. Laufer (1994) contends that vocabulary should be an important component of curricula. Laufer (1994) recommends the following:

> 1. If the course has no lexical syllabus, such a syllabus should be provided. 2. If it exists, it is not enough to encounter the syllabus words in reading. They should also be practiced for productive purposes. 3. Learners should be taught that effective lexical use implies not only the use of sophisticated vocabulary, but also effective variation of words. This can be achieved by systematic practice on lexical reformulation. (p. 32)

The results of this study are evidence that this needs to become a reality in the Costa Rican context. At the end of high school education, the students in the sample do not handle the basic 2000 words of English, nor are they prepared to use the first 1000 words fully when producing written texts; that is particularly the case of the FLS students. The Costa Rican educational system must strengthen the bases for solid second language development, and a good way to begin could be by strengthening the students' bases for vocabulary development.

Those who have undergone the process of SLA know that producing language is much more demanding than processing the language that we receive through input. As mentioned above, Laufer and Goldstein (2004) maintain that vocabulary production lies at the top of the hierarchy of vocabulary knowledge as one of the most difficult skills to achieve. When producing language, we must integrate many aspects into the vocabulary item that we select. Phonological, grammatical, syntactical and contextual aspects are all encoded into the vocabulary that we chose to be part of a written text. If vocabulary teaching and learning take place in the classroom, the students' task may be at least partially aided.

Multiple studies speak of the incremental development of vocabulary knowledge (Schmitt and Meara, 1997; Schmitt, 2007, 2010a; Laufer and Goldstein, 2004; Doczi and Kormos, 2016). The present study also serves as evidence for this progressive development of vocabulary as represented by the percentages of the different types of words found in students' compositions. Therefore, the bases appear to be there, but we still seem to require greater emphasis on vocabulary in the curricula. Both professors and students need to become aware of the importance of vocabulary for general language proficiency and for writing skill development in particular. Hopewell and Escamilla (2014) recommend considering writing as a key aspect of literacy, an honor, they say, usually associated to reading. These authors maintain that biliterate writing must be understood as a critical factor in the process of bilingualism and biliteracy. It is hoped that the present study will contribute to achieving this in the Costa Rican context in the near future.

There is a myriad of activities that instructors can implement to ensure vocabulary learning. Specialists tell us to help students notice vocabulary (Schmidt, 2010); to present them with the process of need, search, and evaluation (Laufer and Hulstijn (2001); to help them practice using experience tasks to put into practice the ability that comes from life experience (Nation, 2007b); and to use spacing techniques to facilitate vocabulary recall (Schuetze, 2015; Nakata and Webb, 2016). Writing offers the "push" that Nation (1996) describes as needed to boost vocabulary knowledge development.

When students realize that they lack words, and they find them and use them in writing, this lends itself to vocabulary learning very much in line with Swain's (2000, 2006) output and languaging hypotheses. Writing activities, in general, contribute greatly to the fourth strand of *Four Strands of Language Learning* (Nation, 2007a), fluency development. This strand, deals with the idea of students' use of the language that they know. For fluency to be achieved students need to have many chances to practice, in a free context, the words that they have acquired when concentrating more on form. Multiple opportunities for vocabulary practice were also described in section 4.4. where different learning vocabulary strategies and the instructor's role in vocabulary teaching were discussed. Whereas students in both schools seem on the right track for productive vocabulary development, those of the CB school are using the vocabulary that they know more effectively.

In terms of the effect of the variable gender, this section of the study also shows that this variable seems to have a limited effect on SLA. It seems that the effect of the type of instruction plays a more significant role in terms of effective use of vocabulary than gender does.

9.3. VOCABULARY LEVELS AND WRITING

In the following paragraphs we describe studies that examine the connection existing between vocabulary knowledge and the writing skill. As mentioned above, the dearth of studies analyzing the connection between productive vocabulary measures and writing quality is overwhelming. The analysis conducted aims at offering additional information in this area. We will later compare the results obtained by the participants here in the PVLT with those shown in the writing task with the purpose of determining how much of the writing ability of students can be predicted on the bases of the PVLT.

Laufer and Nation (1995) investigated whether the written production of students correlated with the students' scores in an active version of the VLT. The study involved different backgrounds, different levels (presumably equivalent to novice, low intermediate and high intermediate), and English learners in New Zealand (n = 22) and Israel (n = 43). These students wrote two compositions and were also tested in the active version of the VLT at the levels of 2000, 3000, 5000, the University Word List (UWL; now called the Academic Word List or AWL), and the 10,000-word band. Laufer and Nation (1995) digitized the compositions to be analyzed by the LFP tool, an earlier version of a tool with similar characteristics and functions as those of the *AntWordProfiler*. For the first composition, the study determined that the lower the level of the students the higher the presence of high frequency words and *vice versa*. This feature was maintained in subsequent stages. The authors found that groups were not statistically significantly different in terms of the 2000 level words used. In regard to the UWL, the least proficient group used fewer words from this level than did the other groups. For the second composition, the three groups were more clearly differentiated in the use of the UWL and *Not-in-the-list* words. Once more, the lowest proficiency group used fewer words from the UWL than did the high intermediate group. The general trend was for less proficient participants to use more words from the second 1000 level.

With respect to the correlation between the Active Levels Test and the written composition, Laufer and Nation (1995) conclude that students with higher scores in the Levels Test used more vocabulary from the UWL and from the *Not-in-the-list* words, that is upper level vocabulary words. They further report a negative correlation between the first 1000 words and the Levels Test because the more words participants knew the fewer words from the 1000, high frequency band they would use. In addition, there was no correlation between the Levels Test and the second 1000 level words found in the participants' compositions. Laufer and Nation (1995) conclude that vocabulary from the 2000 word band "was used as frequently by learners with low vocabulary size as by those with larger vocabulary size" (p. 317). The authors note that the LFP may be a more stable tool when it measures lexical richness above the 2000 level. In their study, when they calculated results including the 1000 level words, the results yielded several statistical differences across compositions. However, when the 1000 level words were not part of the profile (including only 2000 level, AWL, and *Not-in-the-list* words), the LFP tool produced very stable results across the two compositions analyzed, thus proving to be highly reliable at these levels. They further conclude that a productive

use of language can be taken as an adequate reflection of students' vocabulary size as measured by a vocabulary test.

Along these lines, these authors maintain that students exhibiting a high frequency profile usually obtain high scores also in vocabulary level tests such as the VLT. For Laufer and Nation (1995, p. 313), "one cannot use a word correctly without understanding it, or without being able to provide it in a word-elicitation task." This justifies why several studies attempt to establish a correlation between written production and the VLT. For instance, Stæhr (2008) investigates the connection between learners' vocabulary size and their lexical sophistication in writing. This study used the VLT and compared the results in this test to those obtained by the subjects on a writing task. Stæhr does acknowledge that a productive vocabulary size measure could have provided higher correlations. Stæhr's students were evaluated at the 2000, 3000, 5000 and 10,000 levels of the VLT (Schmitt et al., 2001). For the writing test, students wrote a letter to a job agency applying for a job. They were asked to write between 350 and 450 words. Those subjects had dictionary access and were allowed one and a half hours to complete the task. This study reports that only 20 out of the 88 participating teenagers mastered the 2000 VLT. As for the distribution along the different levels assessed in the test, twelve subjects mastered only the 2000 level; three participants mastered the 2000 and 3000 word levels; and five participants mastered the 2000, 3000, and 5000 word levels; none reached the 10,000 word level. A correlation analysis and a regression analysis are also reported for these tests.

The results show a high, statistically significant correlation between vocabulary size and the quality of the students' writing skill (coefficient 0.73). In addition, 52% of the variance in the written compositions could be explained through the results in the VLT (R^2 1 /4 0.524). Stæhr notes that this correlation of 0.73 is notoriously high, considering that the VLT measures vocabulary size at the receptive level. He suggests that using the PVLT could have yielded an even higher correlation. Stæhr further argues that knowledge of the 2000 words resulted in students performing above average in the writing task. In addition, he maintains that the results hint at the role played by other factors that also contribute to generate a good piece of writing and the opportunity that students had in accessing a dictionary. Once again, Stæhr concludes that knowledge of the 2000 words has a positive effect on the students' performance in the writing task. This renders knowledge of this word band a crucial aspect, especially for low-level EFL learners.

In another study, Agustín Llach and Terrazas Gallegos (2009) offer more evidence in relation to receptive vocabulary levels and writing quality for 274

Spanish, primary students, age 12.39, with 629 hours of instruction. The 1000 and 2000 Receptive Vocabulary Levels Tests were used to evaluate receptive vocabulary measures. A letter containing personal information, hobbies and interests, as well as a hometown description was the written task to evaluate writing quality. Agustín Llach and Terrazas Gallegos (2009) describe their subjects as knowing 1,106 words (1K = 713 words, and 2K = 393 words). They assume that those numbers indicate that some words of the 2000 are part of these students' vocabulary, though it could not be claimed that they have mastered this level. In calculating the correlation between receptive vocabulary and quality of written production through a Spearman test, these authors find a significant correlation (p = .01) of r = .542 and r = 503 for the 1000 and 2000 word levels respectively. They further state that the 1000 word level vocabulary accounts for 29% of the variance in writing quality and the 2000 word vocabulary accounts for 25% of the variance in the quality of their students' compositions (p. 139).

Moreover, the authors note that students seem to use words from the 1000 words more easily and accurately. They maintain that the moderate correlations in connection to the 2000 word level words may be interpreted as an indication of these students' use of "other frequency levels […] apart from those of high frequency" (p. 140-141). We should consider this explanation very carefully since many studies have argued that vocabulary develops progressively and incrementally (Nation, 1983, 1990, 2006b, 2006c; Schmitt et al., 2001; Schmitt, 2010a), thus making this explanation highly unlikely. We expect students to master the lower vocabulary levels before mastering, first receptively and then productively, higher vocabulary bands. Agustín Llach and Terrazas Gallegos conclude that the more words the students know, the better the quality of their writing. These authors interpret the weak correlation between vocabulary and writing as an indication of the influence of other factors such as accuracy, fluency, syntax, mechanics and content on the overall quality of writing.

9.4. THIS STUDY: WRITING AND PVLT REGRESSION ANALYSIS

When students are asked to apply their knowledge, according to Laufer and Nation (1995, p. 308), we "would expect to see a relationship between direct measures of learners' vocabulary size and the richness of vocabulary in their language production." We attempt to establish this relationship between the PVLT and the written composition of these students. A Simple Linear Regression was

calculated to find the possible relation between a controlled productive activity and a free productive one. We wanted to determine how much of the students' written production could be predicted by their achievement in the PVLT. We have analyzed the case of the FLS first. Once the assumption of normally distributed residuals around the mean is satisfied, the Pearson correlation indicates a relation of .357, a small, but positive correlation with significance of .000. The test shows a significant regression (by the equation $F_{[1, 123]} = 17.984$, $p < .000$), with the R^2 of .128. Based on this R^2, 12.8% of the variance in the number of words in the composition could be predicted based on the PVLT. The participants' predicted word count is equal to 129.848 + .063 (words), when the words in their compositions are measured based on the PVLT. The number of words increased 0.63 for each word in the PVLT.

For the CBS, no significant regression is found, despite satisfaction of residuals normally distributed around their mean. The Pearson correlation indicates no relation between the two variables with $p = .483$. When applying the model $F_{[1, 49]} = 002$, $p < .066$, with a R^2 of .000, that indicates that according to R^2, we cannot predict the students' production in their composition, based on the results of the PVLT. Here, if we apply a formula using this model, the possibility of prediction is very limited. The students' predicted word count is equal to 256.302 + .002 (words), when the words in their compositions are measured based on the PVLT. That is, the median in the composition is not related to the number of words they recognized in the PVLT; it remains independent from the results in this test. The number of words in the composition increased by 0.2% for each word in the PVLT.

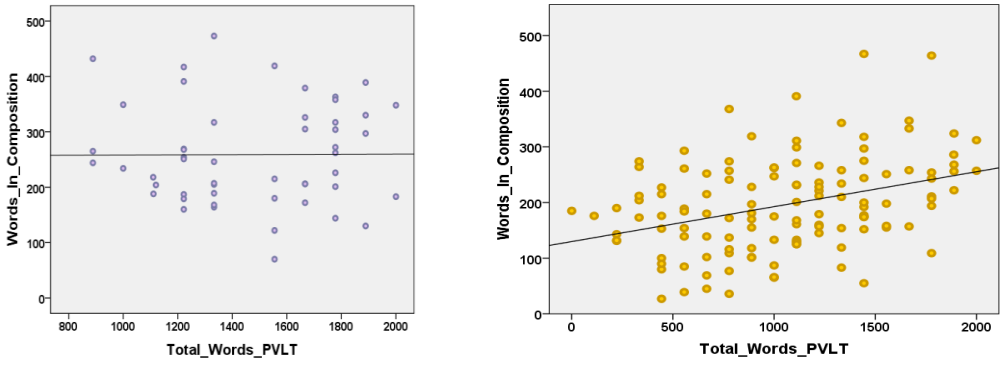


Figure 25. Regression analysis for words in the PVLT and words in the composition.

In sum, production is better explained through the PVLT in the FLT school than in the CBS, but still to a limited extent. In the FLT school, 12.8% of the variance in the number of words in the composition could be predicted based on the PVLT, whereas in the CBT school, there is no relation between students' scores in the PVLT and the number of words that they produced in the composition. These results are represented in the scatter plot in Figure 25.

A further regression analysis was performed to determine whether there was a correlation between the results obtained by students in the VLT and the performance of students in the composition. Authors such as Stæhr (2008) and Agustín Llach and Terrazas Gallegos (2009) have been able to identify correlations with these measures. For the CBS, no significant correlation is found, despite satisfaction of residuals normally distributed around their mean. The Pearson correlation indicates no relation between the two variables with $p = .161$. When applying the model $F_{[1, 49]} = 1.309$, $p < .258$, with a R^2 of .026. That suggests that according to R^2, we cannot predict the students' production in their composition, at large, based on the results in the VLT.

After application of a formula based on this model, the possibility of prediction is very limited. The students' predicted word count is equal to 83.187 + .098 (words), when the words in their compositions are measured based on the VLT. That is, the median in the composition is not related to the number of words they recognized in the VLT; it is independent from the results in this test. The number of words in the composition increased by .161 for each word in the VLT. Only 2.6% of the variance in the number of words in the composition could be predicted based on the VLT.

We now proceed to analyze the case of the FLT school. Having satisfied the assumption of normally distributed residuals around the mean, the Pearson correlation indicates a relation of .087, a very small correlation which is not statistically significant with $p < .167$. The test shows a non-significant regression with the equation $F_{[1, 123]} = .945$, $p < .333$), with a R^2 of .008. Based on this R^2, the variance in the number of words in the composition could not be predicted based on the VLT. The participants' predicted word count is equal to 164.300 + .021 (words), when the words in their compositions are measured based on the VLT. The number of words in the composition increased .21 for each word in the VLT. In sum, for the present study, we find no statistically significant correlation between the words students identified receptively in the VLT and the number of words that were produced by the subjects in the composition.

9.4.1. Discussion

Assuming that a good quality vocabulary reflects solid language knowledge, which in turn reflects adequate language proficiency, then we would have to interpret the poor results in writing and in particular the poor connection between the PVLT and the written composition as a sign of a very limited language proficiency for the participants. According to Laufer and Nation (1995), it is expected that the vocabulary of students, as measured by a vocabulary test, should be reflected in the learners' productive use of these words. They clarify that "[w]here this does not happen, we need to look with concern at the opportunities that the learner has to bring language knowledge into use and adjust the language program accordingly" (p. 319). The results for the CBS, in terms of correlation between the active version of the vocabulary test and the composition, are in line with those of Laufer and Nation (1995). In that research the authors also tried to establish a correlation between the results in the active version of the vocabulary levels test and the words in the compositions. As was the case here for the CBS, Laufer and Nation (1995) found no correlation between the results obtained in the active version of the VLT and the written composition. For the FLS, however, a small but significant correlation is found.

In the case of the CBS, one possible explanation for the lack of correlation between the results in the PVLT and the writing texts produced by the students is the actual level of words that students used the most. While the PVLT includes words from the first 1000 words, many of the words assessed in this test are part of the second 1000 word list. Going back to the results of the types of words that students use, we recall that around 75% of the words in the texts produced by students in CBS belonged to the first 1000 words of the language. Thus the disparity between the words assessed in the 2000 PVLT and the lower level words produced by students in the composition may explain the lack of correlation between the two tasks. This does not explain why there is a small, significant correlation between the PVLT and the composition of the students in the FLS. In this school the 1000 word list vocabulary is slightly smaller (73.59) than that of the CBS and yet for the FLS a correlation is present.

A second possible explanation, related to the previous one, is that students clearly do not handle the 2000 word band. The lack of knowledge of words from this word band makes it impossible for students to produce these words in a written text. Thus, even if some students were able to produce words in this band for the

PVLT, other factors might have influenced that production. Perhaps students were guided by the content of the sentence that the word was immersed in or maybe because the word in the PVLT was associated to a topic the students were more familiar with. Neither explanation would guarantee that students would use words in this band in their own free production. And this was, indeed, the case for most students only produced between 6.24% (CBS) and 5.88% (FLS) of words belonging to the 2000 word band. Again, the percentage is slightly lower for the FLS.

Another possible explanation for this limited correlation in the present study could be the uneven number of total words (very reduced in some cases) produced here. Laufer and Nation (1995) found that short compositions (less than 200 words) were not stable enough when analyzed by the LFP, a former version of the *AntWordProfiler*, used in the present research. Laufer and Nation (1995) also argue that because the first 1000 words are made up mainly of basic and function words, these words may not serve as an indication of a highly developed lexicon, saying: "true lexical quality of a piece of writing is determined by the proportion of all the other words at the more advanced frequency levels" (p. 317). They further note that the tool they used, the LFP, may have been more useful when measuring "lexical richness in general or that it is better for the advanced students" (p. 318). Thus, the reduced number of words produced in the composition, along with the fact that these words were mostly associated with the first 1000 word band, may have contributed to this lack of correlation.

One final reflection that can be drawn based on the existing correlation between the PVLT and the composition is that students from the FLS seem to be better suited to use the knowledge that they have acquired in free production. Although the results of the FLS in the PVLT show lower productive levels than those of the CBS, the production of the former in the composition serves as a better reflection of this knowledge. That is, the knowledge that they showed in the controlled production test (the PVLT) and how they used this knowledge in free production (the composition) represents a more efficient use for the FLS than for the CBS. Since students from the CBS attained much better results and seemed to possess a larger number of words from the 2000 band in the PVLT, we expected this knowledge to be reflected in their composition. This, however, did not occur. As mentioned in the initial paragraph of this section, this calls for reflection regarding the teaching and learning practices taking place in each of the settings involved in this study. Knowing that the written production of the students serves as evidence to the limited vocabulary that students have acquired should raise flags

and direct instructors toward focusing more attention on the development and practice of vocabulary both inside and outside of the classroom. Writing provides an ideal context to review, reinforce and provide space for general vocabulary practice to be implemented according to the students' level.

Regarding the VLT and its possible effect on the students' writing skills, we found that no correlation can be identified between the two. We had already anticipated that, knowing that the VLT measures vocabulary at the receptive level while writing is clearly a productive skill. Nonetheless, we wanted to explore this to determine whether we could find some type of correlation such as those identified by other authors. Stæhr (2008), for example, found a significant correlation of .73, where 52% of the variance in writing could be predicted based on the VLT. Agustín Llach and Terrazas Gallego (2009) speak of a correlation of .54 and .50 for the 1000 and 2000 word levels, where the former could explain 29% of the variance, and the latter would explain 25%. Such a correlation is non-existent in this study. Some of the explanations given above for the PVLT could apply here. In particular, the reduced and varied total number of words produced by students in the compositions may be affecting the stability of the data in connection to other measures. Future studies could help us determine whether those explanations are correct. For now, the data does show that the limited vocabulary sizes attained by students in their VLT and PLVT are also made evident in the limited production in their composition. Action should be taken to allow for greater learning and use of vocabulary within the population analyzed.

Limitations have also been identified in this part of the study: One of the most evident is related to the fact that students do not receive specific instruction directed toward writing techniques. When they do write, they do so for a reading activity they are working on. This turns composition writing into an unfamiliar classroom activity for most students, and affects the final students' products. Second, the results analyzed in this section correspond to a one-time collection of data. A longitudinal study in which several collections of data could be made during a semester would provide a more detailed description of the students' writing level. Finally, it would be ideal to select compositions that are over 200 words long. This could not be done here because most compositions did not reach this word mark; and eliminating shorter compositions would not have offered a general perspective on the real situation of the sample. Future research could take this into consideration. Finally, this analysis involves only two high schools from one specific area of the country. To draw more general conclusions, a larger sample including more institutions and a wider geographical area are necessary.

CHAPTER 10

EFFECT OF MOTIVATION RESULTS

For greater detail on the differences between the results of the two schools participating in the tests, students from both institutions also completed a motivation survey, adapted from Gardner's (1985, 2004) Attitude/Motivation Test Battery. Although his entire instrument contains over 100 items, it was reduced in the present study to 28, to concentrate on specific clusters of the survey, and reduce the time required to complete the task. The analysis of the data obtained is presented below. This survey contains 5 scales or clusters of items that are intended to gather information on the following affective aspects: *Motivation intensity, Desire to learn English, Integrative orientation, Instrumental orientation,* and *Attitude toward learning English*.

10.1. MOTIVATION RESULTS

A Cronbach's alpha test was used to analyze coefficients for the whole test and for each cluster in each school. The reliability results for the entire test and for each of the different scales were examined to provide an estimate of internal consistency. This information is summarized in Table 52.

Table 52. Reliability measures results for motivation survey

	School	**α value**
Entire test	CBT (N= 52)	0.70
	FLT (N= 130)	0.90
Alphas per cluster		
Motivation intensity	CBT	0.45
Items: 5, 11, 18, 23, 1, 8, 13, 20	FLT	0.66
Desire to learn English	CBT	0.63
Items: 4, 15, 19, 26, 7, 17, 22, 24	FLT	0.80
Integrative orientation	CBT	0.64
Items: 3, 10, 14, 25	FLT	0.70
Instrumental orientation	CBT	0.03
Items: 6, 12, 16, 27	FLT	0.61
Attitude toward learning English	CBT	0.48
Items: 2, 9, 21, 28	FLT	0.73

As observed in Table 52, the CB school obtains quite low reliability on three of the scales; namely, Motivation intensity, Instrumental orientation and Attitude toward learning English. The rest of the scales in the study exhibit acceptable internal consistency for the CB school. All the clusters show moderately high

internal consistency for the FLS. For the scales with limited internal consistency for the CBS, different interpretations could be given. On the one hand, they can be taken to indicate that these scales, as such, do not represent the individual characteristics of that population sample for that specific construct. On the other, given the nature of the variable "motivation," it is still worth exploring the data collected by the survey as the lack of identification of the students with any given scale is in itself representative of the nature of the features in the sample. In addition, the overall consistency of the test, the whole test alpha, is reliable for both schools. While all of the clusters will continue to be analyzed in the following paragraphs, our main attention will be given to the clusters Desire to learn English and Integrative orientation, both of which have adequately high internal consistency in each of the schools. The notion of integrativeness (Gardner, 1985, 2007), as well as its connection with the L2 Motivational Self-System (Dörnyei, 2005, 2009a, 2014), can be explained in part through these two clusters.

To obtain as much information as possible, the data was analyzed to determine more specific characteristics of the population. Table 53 depicts descriptive information as well as inferential statistical results for each affective cluster per school. The results of the non-parametric and parametric tests used to determine statistical differences are also included in Table 53.

Table 53. Descriptive statistics results on 5 clusters for CBT and FLT schools

		Median	**Mean**	**SD**	**Sig. 2-tailed**
Motivation intensity	CBS	4.00	3.97	.064	.056
	FLS	3.87	3.80	.056	
Desire to learn English	CBS	4.87	4.69	.051	.040
	FLS	4.87	4.69	.049	
Integrative orientation	CBS	4.75	4.73	.051	.001
	FLS	4.50	4.42	.060	
Instrumental orientation	CBS	4.25	4.28	.059	.012
	FLS	4.25	4.01	.062	
Attitude toward learning English	CBS	5.00	4.67	.061	.254
	FLS	4.75	4.55	.054	

According to the results of the Kolmogorov Smirnov test, all clusters (except for that of Motivation intensity) follow a normal distribution. For that reason, a Mann-Whitney parametric test was used to determine the nature of statistical differences for all clusters, except for the Motivation intensity cluster for which a *t*-test was applied. The results of these tests are also presented in the right-

most column of Table 53. Two of the scales show non-statistically significant differences among schools; this is the case of Motivation intensity (barely showing non-statistically significant effects) and Attitude toward learning English. In contrast, Desire to learn English, Integrative orientation, and Instrumental orientation do show statistically significant differences that favor the CBS. These have ranks of 100.74 (as opposed to 83.51), 109.16 (as opposed to 81.38), and 105.23 (as opposed to 83.93) respectively for each of these scales for the CBS as compared to the FLS.

More tests were needed to find the statistical differences between gender groups at intraschool level, and then for the other school population. The results for the effect of gender analysis are given in Table 54, for the Mann-Whitney U test used to establish the type of difference for Desire to learn English, Integrative orientation, Instrumental orientation and Attitude toward learning English.

Table 54. Gender effect for motivation inside each school

	Desire to learn English	Integrative orientation	Instrumental orientation	Attitude toward English
Content Base School (20 male, 32 female)				
Rank Masc/Fem	27.25/25.19	27.23/26.05	25.15/26.55	26.93/26.23
Mann-Whitney U	285.500	305.500	293.000	311.500
Wilcoxon W	781.000	833.500	503.000	839.500
Z	-.491	-.293	-.336	-.170
P (two-tailed)	.623	.769	.737	.865
Foreign Language School (64 male, 64 female)				
Rank Masc/Fem	63.70/62.27	59.35/67.52	65.36/63.64	62.63/65.39
Mann-Whitney U	1907.500	1727.000	1907.500	1928.500
Wilcoxon W	3798.500	3680.000	3798.500	4008.500
Z	-.222	-1.285	-.222	-.437
P (two-tailed)	.825	.199	.825	.662

Table 55 shows the results of Levene's test and the *t*-test applied to the data for the Motivation intensity cluster for each gender group. For that test, we must consider that the median ranks correspond to 33.30 and 30.87 for male and female students in the CBS, and 30.25 and 30.49 for male and female students in the FLS. For the effect of gender on motivation in each school, no statistically significant differences were identified, except for the Motivation intensity cluster in the CBS. In that case, statistical significant differences ($p = 0.018$) favor male students in the CB school. Keeping in mind the low internal consistency for this cluster in each school (see Table 52), these results should be interpreted with care.

Table 55. Gender effect for motivation inside each school

		Levene's Test		*t*-test for equality of means		
		F	**Sig.**	***t***	**df**	**Sig. (2-tailed)**
Motivation intensity CBT	Equal variances assumed	.049	.826	2.441	49	0.018
Motivation intensity FLT	Equal variances assumed	.309	.579	-.266	119	0.791

Another set of tests was required to establish the type of differences between gender groups across schools. A Mann-Whitney U test was applied to the clusters Desire to learn English, Integrative orientation, Instrumental orientation and Attitude toward learning English; a *t*-test was used for the cluster Motivation intensity. For this last test, we have to consider that the median rank is 33.30 and 30.25 for male students in the CBS and FLS respectively, and 30.87 and 30.49 for female students in the CBS and FLS. The statistical results are presented in Tables 56 and 57. In this analysis, the results favor male students in the CBS for the Integrative orientation cluster, but females in the CBS show statistically significant differences for Instrumental orientation and Motivation intensity.

Table 56. Mann-Whitney U test for gender effect on motivation across schools sampled

Male (CBS, FLS)				
	Desire to learn English	Integrative orientation	Instrumental orientation	Attitude toward English
Rank CBS/FLS	50.13/40.12	53.55/37.61	49.13/40.43	46.15/41.36
Mann-Whitney U	487.500	379.000	507.500	567.000
Wilcoxon W	2567.500	2332.000	2587.500	2647.000
Z	-1.619	-2.679	-1.414	-.794
P (two-tailed)	.105	.007	.157	.427
Female (CBS, FLS)				
Rank CBS/FLS	51.63/43.89	55.83/44.84	56.08/44.09	50.92/46.52
Mann-Whitney U	786.500	789.500	741.500	914.500
Wilcoxon W	2677.500	2869.500	2821.500	2930.500
Z	-1.327	-1.897	-2.018	-.770
P (two-tailed)	.185	.058	.044	.441

Table 57 shows the results of Levene's test and the *t*-test applied to the data for the Motivation intensity cluster for each gender group. For the effect of gender on motivation, in general (not by school), statistically significant differences were identified for females, but not for males.

Table 57. Use of a *t*-test to determine gender effect on motivation across schools in the sample

		Levene's Test		*t*-test for equality of means		
		F	**Sig.**	***t***	**Df**	**Sig. (2-tailed)**
Motivation intensity Female	Equal variances assumed	1.432	.235	2.695	76	.009
Motivation intensity Male	Equal variances assumed	2.745	.101	92	92	.706

10.1.1. Discussion

Before beginning this analysis, we must recall that motivation is difficult to measure because it is subject to constant changes and it may vary dramatically even within short periods of time. Thus, while the results obtained here may be valid for the situation at the moment the survey was conducted, changes may have taken place at different points in time during the students' learning process. With that in mind, every result involving the construct of motivation must be approached with caution.

The analysis of the different clusters in the survey on motivation do shed light on the issue of motivation. In particular, we found that in the cases of Desire to learn English, Integrative orientation and Instrumental orientation, the statistical results suggest that pupils from the CBS outperform their FLS counterparts. The interpretation of these clusters in regard to students' motivation becomes especially important in light of theories such as Gardner's (1985, 2007) integrativeness theory and Dörnyei's (2005) L2 Motivational Self-System. No differences were perceived for Motivation intensity and Attitude toward learning English. However, these results should be analyzed with care considering that the internal consistency for the CBS in these clusters was rather reduced.

Interesting findings can also be observed when the analysis focuses on the effect of gender in each of the schools. That analysis shows that, for the most part, there are no statistically significant differences associated with gender groups in either school. The only difference is in Motivation intensity in the CB where males outperformed female students. Once again, given that that cluster yielded low internal consistency for the CBS, these results should be considered cautiously. Greater differences are found when we compare gender groups across the two school settings; more differences surface, although all of them favor students from the CBS. For Integrative orientation, male students from the CB school appear to

be more motivated than male students from the FLS. Female students in the CBS also show higher Motivation intensity and Instrumental Motivation than the female students in the FLS. In sum, when differences across gender do appear, they constantly favor students enrolled in the CBS.

Of particular interest is the fact that male students in the CBS show a stronger Integrative orientation to learn English that their peers in the FLS. Also interesting is the case of female students in the CBS setting whose Instrumental orientation is higher than that of their peers in the FLS. A look back at each of these types of motivation helps us recall their characteristics. The integrative part of motivation (associated with the construct of Ideal L2 Self) has to do more with intrinsic motivation. Based on this, male students in the CBS would seem to be more internally motivated than students from the FLS. Interestingly, the instrumental orientation of motivation (Dörnyei's Ought-to Self) is traditionally associated with extrinsic aspects of motivation. It has to do more with what students believe they should do to comply with requirements of society. Female students from the CB school appear to be more concerned with their supposed responsibilities, and personal and social duties that they associate with their L2 learning, than are female students in the FLS.

While the previous paragraphs discussed the differences found in terms of the effect of gender groups, the remainder of this section will address the more general effect of the Desire to learn English, Integrative orientation and Instrumental orientation clusters for which the CBS showed higher statistically significant differences over the FLS. Through consideration of Gardner's (1985, 2007) and Dörnyei's (2005) L2 Motivational Self System, we attempt to provide additional explanation for the higher scores obtained by the CBS across the different tests of the present study. Dörnyei (2010) establishes a parallel between Gardner's integrativeness motive and Dörnyei's Ideal L2 Self. According to Gardner (1985, 2007), based on the integrativeness motive theory, we can explain the role played by factors pertaining to individual differences and to the learning context where students are immersed. The focal point of integrativeness is on the individual interest in learning the language to use it for interaction in the future. In other sections, it has been seen that the CB instructional context seems to positively influence the results obtained by students across the tests. In the final part of the analysis, we find that the results of the CBS are also supported on the basis of stronger levels of integrative and instrumental motivation as well as by a stronger desire to learn the language.

Gardner (2007) assigns a fundamental role to the *educational context* in the students' general concept of motivation, the type of educational system that students are part of contributes to forming the expectations that students set and shapes the interests that students develop in connection to their learning process. The second fundamental role in Gardner's theory is held by the *cultural context* in the sense that it guides the social variables that lead to integrativeness and cultural identification that mirror the individual's attributes in the language-learning context. This is associated with the instrumental orientation of learning a language. Both factors (integrative and instrumental) are what, in the end, shape students' motivation and serve as the moving force that guides their SLA process.

Mention was made above of how the cultural influence of the U.S., the constant presence of English language in the media, and tourism itself exert a great impact on the students' intention to learn English. The socio-educational context discussed by Gardner (2007) is indeed what conforms the dynamics between the individual, the classroom setting and the cultural context that cause individuals to be interested in learning English. In general, the students' integrative motive requires both an educational and a cultural context that provide them with the support they need for SLA. Gardner (2007) affirms that integrativeness consists of having the motivation to learn the language, having a real interest in using that language for interaction, and having a positive attitude toward the learning situation. Although in the present study the results are good for both schools when compared to other educational contexts outside Costa Rica, the CBS seems to offer students a better combination that results in a greater interest in learning the language.

For Dörnyei's L2 Motivational Self System and in particular the Ideal L2 Self, on the other hand, the emphasis is set on the self-image that the L2 learner has of herself and the goals that are in mind for the future. For Dörnyei (2014), the actions that learners' implement are a reflection of the individuals' motivation shaped through the cognitive and emotional factors. Dörnyei (1998) insists that motivation is the force that provides initial stimulus to learn an L2 and it is also the force that maintains students trying through the difficult process. The Ideal L2 Self, the Ought-to Self and the L2 Learning experience all reflect how learners envision themselves in the future. In the first, the Ideal L2 Self refers to the individual's personal hopes and dreams. In the Costa Rican context, learners are aware, very early in their lives, that English may open many doors for them professionally and occupationally. They are aware of the possibility of traveling, accessing entertainment opportunities and having the chance of getting to know people from

other countries in Costa Rica as well. These opportunities may serve as an invitation for considering English as a tool to achieve what they want in their future lives. This construct is very linked to the idea of integrative motivation that is part of Gardner's theory and it is a lot more intrinsic in nature.

The Ought-to Self, more related to duties and responsibilities set by society, is likely to be associated with the idea of future job opportunities, thus offering a more extrinsic perspective. Costa Ricans are aware of the job-related possibilities existing within their context and outside the country. This awareness may push students to develop an interest in learning English if they want to be part of a work force that stands a better chance of getting a more profitable job if another language is available to them. The last construct, the L2 Learning Experience, is more closely associated with the students' experience in the language classroom, or in general, in language learning. This aspect, as well as those mentioned above, seems to have played a determinant role in the results of this investigation. It sets the context's specific conditions related to immediate learning settings. This study shows that more positive results are obtained in the CBS and the results of the motivation survey also suggest that students in this setting are more motivated as well.

Intricately related to the concept of Ideal L2 Self is Dörnyei's (2017) idea of narrative identity. Serving as a link between the process of learning an L2 and the use of the knowledge of this L2 in real life experience, the narrative identity helps leaners have a complete perspective of their personal experience in terms of the real use of the L2 either in the school or social context. Students, in their personal experience and implementation of the Ideal L2 Self and narrative identity, or through contact with others (family or acquaintances) with experiences for whom English has played a key role in their everyday lives, may be moved to learning English and may feel more motivated to do so at a younger age. This is added to a well-known interest on the part of parents who look for ways to help their children come into contact with English and especially with English instruction since early stages of learning. In general, for students in Costa Rican schools, it would be fair to say that they are exposed to a number of aspects that create an awareness of the benefits that the English language may bring to their lives.

The concepts of Ideal L2 Self and narrative identity might as well serve as an explanation for the differences in receptive and productive vocabulary profiles of learners who have participated in studies in the Spanish context and the Costa Rican students in the present study when the results are based on the VLT and the PVLT. In all the studies discussed in the VLT and PVLT sections of the present study, the Spanish students obtain lower vocabulary word counts than the Costa

Rican students in the present sample. We could hypothesize that, in the case of Costa Rica, aspects such as being a much smaller country (resulting in better chances of contact with target language speakers) and being more exposed and allegedly more in contact with an English-speaking culture (that of the U.S.) inadvertently have had an effect on students' predisposition, interest and motivation toward learning English. This, together with the central role that English has for these students' future job markets and the fundamental role of English in connection to the economy in general, assign English a more critical role for secondary students.

The idea of the demand for English in Costa Rica has been documented through a study where instructors, graduate students, management representatives from several national and international companies, and university program administrators have discussed the need for a labor force that is proficient in, at least, writing, reading, and speaking skills (Hernández Herrero, 2008). It may be that in Spain although there is interest in EFL instruction and attention is given to forms of language instruction, especially in terms of CLIL, the future challenges for students might not be made to be so intricately connected to English in particular. In Spain, with other options for SLs inside the country and with contact with so many other languages in the European zone, the focus of attention, for secondary students at least, might not be so constantly placed specifically on English as a second language. Further studies dealing with motivation in both of these populations would help shed light in this area. For now, the differences identified in this study have led to hypotheses on possible explanations for this diversity.

The above analysis points to a general degree of motivation among Costa Rican students which, when combined with the correct implementation of EFL instruction, is what could be enabling students to advance in the SLA process. It also would appear that the context of Content Based Instruction is yielding better results for the population in the study sample both in terms of word count and score results and in respect to motivation. Whereas the students in the CB school still must improve their language levels, students in the traditional Foreign Language Instruction system show results well below the measures produced by the CB school, and coincidentally, lower motivation levels. While both schools' results still have room for improvement, the CBS seems to offer better conditions for EFL instruction and higher degrees of motivation than the FLS setting does.

Nevertheless, the fact that the survey was severely reduced (due mainly to time constraints and the need to narrow down the topics of interest) may have affected the results obtained in the data. The reliability of tests tends to be higher

when they include more items. This reliability may have been hindered especially where there were only four items in a cluster. Additionally, certain students appeared to be confused with wording when the statements were presented in the negative form. In future studies, these negative statements should be limited to where there are sufficient items to show the tendency of the results. In general, more items in the instrument as well as more time to carry out a more extensive survey are highly recommended. It would also be essential to conduct surveys with larger population samples and with longitudinal studies. Although additional studies could corroborate the information discussed in this section, for now the analyses of the results suggest that both data and motivation support the idea of CBI yielding greater benefits to the student population. The value of these preliminary results resides in the fact that it has been shown that CBI is worthy of further attention in the case of language teaching in Costa Rica.

CONCLUSIONS

This study is based primarily on the undeniable need for scientific evidence substantiating the state of EFL instruction in the Costa Rican context. While the field of EFL instruction in this context could be approached from many different standpoints, our research addresses development in its essential role in the process of second language acquisition. With the purpose of determining receptive and productive vocabulary knowledge of students in different instructional settings, several tests were applied. The analysis of the results obtained through the collection of varied data yielded a wealth of information in an area of research that has not been sufficiently explored in the past, especially when it comes to SLA. Many conclusions can be drawn in connection with the concepts and analyses discussed here. These findings are examined in detail. First, we reflect on a number of conclusions derived directly from different tests that were given to the student population. With this purpose, the specific conclusions stemming from each of the instruments are focused on one by one. Then attention is given to more general conclusions drawn from the application of the theoretical concepts applied to the Costa Rican context throughout the thesis.

ANALYSIS OF SPECIFIC FINDINGS

1. Controlled receptive vocabulary: The VLT

The receptive vocabulary profile of students as measured by the VLT is set at 1,793.94 words for students in Content Based Instruction and at 1,475.85 words for students taught with more traditional Foreign Language approaches. Following the specification of Schmitt et al. (2001), students from the CBS master this word level (albeit still showing some limitations) while students from the FLS do not. Statistical tests show that the word count difference is statistically significant, favoring the CBI students. The knowledge of this word band is not only higher for the CBS but also more cohesive as word counts are more concentrated above the 1400 word range. This can be taken as an indication of more uniform vocabulary knowledge in the CB school student group than in the FLS population, whose vocabulary knowledge appears more spread along the word band continuum with some subjects having high word counts while others show very limited vocabulary knowledge. Together with the degree of motivation, these results point to greater

benefits derived from the type of methodology implementing the teaching of content using English as a medium of instruction.

The results also indicate that specific practices are required to remedy vocabulary deficiencies, especially in the FLT context where students in the sample do not master this basic word band at the end of their secondary education. Some students in the CBS still need to improve their vocabulary counts as well. In the latter setting, while the results show that the scores are sufficient to reach the minimum required to master this word band, the margin is only barely sufficient. It is imperative for students to reach this word band in the early years of EFL instruction, and certainly before leaving secondary schooling, for the following reasons: 1. Vocabulary growth develops faster at initial language levels (Dóczi and Kormos, 2016). 2. Students require frequent, repeated exposures for vocabulary to be learned (Dóczi and Kormos, 2016; Pigada and Schmitt, 2006; Nation, 1993). 3. The 2000 word band covers over 80% of the vocabulary appearing in everyday texts, conversations and novels and thus knowing this word band brings great benefits to students (Nation, 2013). 4. Students must know vocabulary of upper level bands to be able to use that of lower level bands, thus lower word bands serve as a bases for development of other word bands (Laufer and Nation, 2001). 5. Vocabulary develops incrementally (Read, 1998; Schmitt et al., 2001; Schmitt, 2007, Dóczi and Kormos, 2016), so this is the first word band students must learn to ensure subsequent development of vocabulary knowledge. The benefits derived from knowledge of this word band are essential for the adequate development of major language skills such as reading and writing. Thus, mastering this word band should become of central importance in the future in the Costa Rican language context.

2. Receptive vocabulary in context: Reading comprehension test

When vocabulary knowledge is evaluated in context through reading comprehension, students receiving CB instruction have higher scores and statistically significant differences in comparison to peers receiving FL instruction. The knowledge of the CB students—as reflected by the spread of the distribution of their grades—also appears more unified. That is, there is greater cohesion in terms of the knowledge that students display in reading comprehension in the CB school, as opposed to a greater dispersion of levels of reading comprehension skills found in the FLS.

In general terms, the results of this study indicate that students develop an "adequate" level of reading comprehension when measured in connection with the ability required to read simplified texts. Texts similar to those used in this study are also found in the compulsory test required by the Costa Rican Board of Education at the end of high school (samples could be found at the official website of the National Board of Education: http://www.mep.go.cr/). The texts in the reading comprehension test selected for this study contained vocabulary belonging mainly to the 1000 and 2000 word bands. The students in the CBS and FLS obtained scores of 86.13 and 74.18 respectively. This suggests that they still have limitations with texts containing high frequency vocabulary. Their scores, however, are still sufficient to pass the test prepared by the Board of Education. Whereas there is room for improvement in both school settings, the overall grades indicate that students develop an adequate level of reading comprehension if the aim is to develop limited reading ability. If the aim is independent, autonomous reading ability involving texts containing vocabulary above the 2000 word level, the students do not appear fully prepared to face such a task.

The data analyzed provides further evidence of the positive correlation between receptive vocabulary size and reading comprehension. In both schools, a positive correlation was identified between the number of words in the students' receptive vocabulary (based on the VLT) and the predictive influence that that has on the students' reading abilities. These results are in line with studies such as those of Qian (2002), Jiménez Catalán and Terrazas Gallego (2005), Stæhrt (2008), Lervåg and Aukrust (2010). Therefore, we can also conclude that larger vocabulary sizes appear to serve as better predictors because a larger correlation is found when receptive vocabulary measures are higher. For the FLS, given its lower receptive vocabulary size, the predictive correlation that can be established with the reading skills is still positive but lower than that of the CBS.

Limited vocabulary measures at the 2000 word level are an indication that the participating students have not reached the vocabulary threshold established for reading. This threshold, the minimum number of words readers must recognize without outside help, has been fixed at 3000 words (Laufer, 1992). For Laufer, this allows access to reading strategies that students implement in their L1 reading, whereas less than 3000 can interfere with students' accessing these strategies easily. Inasmuch as the students in the present study have not reached this threshold, they cannot benefit from these strategies. While some students in the CBS showed a mastery of the 2000 word band, others still have limitations in the knowledge of that band. In the FLS, where the level of mastery was insufficient for the 2000 level,

the limitations are even greater and lead to disadvantages in development of the reading skill and the basic vocabulary threshold for reading.

The contribution of vocabulary is essential for effective reading skills. However, other factors also contribute because the reader already has a substantial amount of information when approaching a reading task (Nation, 1993). Schmitt et al. (2011) speak of a linear relationship between vocabulary coverage and reading comprehension, in which as vocabulary size increases, reading comprehension also increases. When we apply this theory to the present study, we can confirm that higher vocabulary counts in the CBS coincide with higher scores on the reading comprehension test, whereas for the FLS, the lower word counts in the VLT coincide with lower reading scores when compared to those of the CBS.

In the Costa Rican context there is a clear emphasis on input-related practices over output-focused activities. The results of reading practice commonly found in language classrooms place students in beginning levels in the development of reading comprehension skills. We find evidence for this in the "acceptable" results obtained in the reading comprehension test. This can be explained by Nation's (2007a) meaning focus input strand of learning. For this strand to have positive results, students must be exposed to interesting, familiar input information, or whose meaning can be accessed with contextual clues or background knowledge. Assuming that this is possible in the Costa Rican context, we must determine whether the amount of input is adequate. The right amount of input will strengthen the results of the meaning focus input strand; students also need language-focused learning. This concentration on *form* might be what students require to boost their vocabulary learning. Although communicative approaches to teaching are often preferred, the effectiveness of focusing on form is valid as it leads to more permanent vocabulary retention and recall.

3. Controlled productive vocabulary: PVLT

For controlled productive vocabulary, the findings include the following. The sample population that participated in the present study, whose productive vocabulary was evaluated through the PVLT (Laufer and Nation, 1999), has not yet mastered the 2000 word band for productive vocabulary. While students enrolled in CBI show statistically significantly better results than those of students in traditional FLI, they are not enough to be interpreted as pertaining to a full

command of this word band level in either school setting. While the CB context prepares learners to a better extent in the knowledge of these vocabulary forms, their word band knowledge is insufficient, and students in this setting still have not demonstrated knowledge of these vocabulary forms. This knowledge is even more limited in the FL school setting. The productive-vocabulary measures are also lower than the outcomes obtained for the receptive levels in both school settings: 1,793.94 versus 1,467.33 in the CBS, and 1,475.85 versus 1,046.80 in the FLS for receptive and productive measures respectively.

This difference can be explained from a theoretical basis inasmuch as specialists have suggested that the process of vocabulary learning moves from receptive to productive knowledge and that receptive vocabulary tends to be larger than productive vocabulary (Laufer, 1998; Nation, 2013). It has also been claimed that learners need to develop receptive knowledge of vocabulary before productive ability is attained (Webb and Nation, 2017), especially due to the more demanding cognitive load established by knowledge of form, meaning and use aspects necessary to use vocabulary productively. To these types of knowledge, we should add that to produce vocabulary, learners also must master knowledge associated with word recognition and word recall (Nation, 2013). Given that we have already seen that the receptive knowledge of the 2000 word band is barely attained in the CBS and that students in the FLS setting have not yet mastered the 2000 word band at the receptive level, it follows that the productive levels must lag behind. That is exactly what is observed in the results examined here; they provide further evidence for the theoretical assertions addressed above.

Based on the PVLT, the controlled productive ability of the participants in the present study also serves to predict the results in terms of what to expect in the free productive task discussed below. Laufer and Nation (1999) have claimed that students must show mastery of productive vocabulary use in controlled settings (such as that provided by the PVLT) before they can do so in free writing tasks. Hence, based on the premise that the expertise of the productive vocabulary knowledge at the 2000 word band is *limited to severely limited* in a controlled task, we could foresee and explain the outcomes observed in the free production task.

In comparing the Costa Rican results to those of studies carried out with Castilian Spanish students, the Costa Rican students, both under CBI and FLI, largely outperform the Castilian Spanish students (Moreno Espinoza, 2010b; Canga Alonso and Arribas García, 2014). Given the similarities in terms of L1, hours of instruction and classroom methodology (at least for the CB setting in the present

study), these differences are worth taking note of. In the Motivation section below, we have explored possible reasons for these differences.

4. Productive vocabulary in free contexts: Writing task

The results of written production in a free context, as examined in a free composition, show that while students following CBI produce a larger mean in terms of total word counts than students from the FLS, their production is similar to that of students in the FLS in the sense that both groups produce the same *type* of words. The apparent advantage observed in the number of words produced by students from the CBS does not mean that these students possess a vocabulary belonging to higher word bands, but that they use more words in the same bands than students in the FLS do (mainly the 1000 word band and to a more limited extent the 2000 and AWL word bands).

The writing results also provide evidence that students are placed at the very beginning stages of development of their writing skills (Laufer, 1994). Higher concentration of percentages at the 2000, but especially at the 1000 level, indicate that students' writing ability is still limited, and that they indeed have not reached the 2000 productive word band. This conclusion is based on the notion that to use one word-band productively, students require knowledge of higher word-band levels (Laufer and Nation, 2001). That is, according to Laufer and Nation (2001), automatic use of word-band knowledge, such as that required for writing, demands knowledge of vocabulary levels (at least receptively) at higher bands than those used in writing. The present results confirm that; given that the CB students' receptive vocabulary levels barely reach the 2000 word band, and the FLS does not reach that level at all, it is reasonable that most of their production contains words belonging to the 1000 word band. Because they are only now beginning to master the 2000 word band receptively, they have a productive vocabulary knowledge limited to the preceding vocabulary band, the 1000 word level. Once again, this supports the idea of incremental knowledge of vocabulary that goes from basic to more specialized levels (Schmitt and Meara, 1997; Schmitt, 2007, 2010a; Laufer and Goldstein, 2004; Doczi and Kormos, 2016), and that the vocabulary knowledge grows at different rates for distinct types of knowledge (Schmitt, 2007).

It was not possible to establish a clear statistical correlation between the PVLT and VLT and written production for the CB school. For the FLS, a small but

positive correlation (predictive value of 12.8%) is found between the PVLT and writing; again, no correlation is found for the VLT. The varied range of total word counts contained in the compositions as well as the low word levels used by this population have probably impeded establishing an adequate correlation between the tests. In other studies (Laufer and Nation, 1995), it has been shown that word levels beyond the 2000 word band yield more stable results when subjected to analyses like those found here. Students from the FLS appear to show greater consistency between the results obtained on the PVLT and the types and numbers of words that they produced in the free composition. For the CBS, we do not find a clear correlation between the number of words that they have mastered, based on the PVLT and the types and number of words identified in their compositions. We would have expected more words at the 2000 level in this group; their absence indicates that this word level is, indeed, not yet fully mastered.

Vocabulary knowledge is essential for language use in general and depth of vocabulary is imperative for language production (Schmitt, 2008). As discussed above, Laufer and Nation (2001), also established that to improve fluency at a given level—such as writing that uses words at the 2000 level—demands students to have acquired vocabulary at levels beyond the given level, 2000 in this case. In addition, Laufer and Nation (2001) saw that to exercise the use of vocabulary with automaticity learners need to have a strong vocabulary size. Schmitt (1998) also points to the importance of using vocabulary in production tasks, because this ensures that this vocabulary will remain more permanently available in the students' memory.

The results of the present study support the notion that vocabulary knowledge develops progressively and incrementally. This is supported primarily through word knowledge development observed in the writing task where the percentages of words are higher at the basic levels of 1000 words, followed by the 2000 words and AWL level words. This serves as support for the concept of incremental and progressive word knowledge development put forward by authors such as Read, (1998); Schmitt et al. (2001); Schmitt, (2007), Dóczi and Kormos (2016), who maintain that word knowledge progresses from the simplest, most common words toward the least frequent concepts. The evidence found in the present study also lends support to Schmitt's (2007) notion that this incremental knowledge development affects the knowledge of words at different levels. As confirmed by this study, it evolves faster at the receptive level than at the productive level, indicating that while some types of words knowledge are acquired at one

point in time, others take longer; i.e., receptive knowledge precedes productive knowledge.

It has also been confirmed in this study that students' receptive vocabulary is more developed than their productive vocabulary. Theory has demonstrated that receptive vocabulary measures are generally higher than productive vocabulary measures (Nation, 2013; Laufer and Goldstein, 2004). Laufer and Goldstein (2004) attribute the higher counts of passive over active vocabulary to the fact that language production requires *active recall* which occupies the highest rank in their hierarchy for knowledge in the form-meaning relationship. It is followed by *active recognition*; *passive recognition* occupies the easiest, most accessible position in this rank, and is preceded by *active recognition*, which thus implies a little more difficulty. A similar notion is proposed by Dóczi and Kormos (2016), who describe a possible hierarchy of word knowledge for which students in the present study are indeed located at the very initial stages of word knowledge development. In fact, they are just beginning to develop the abilities required to master word knowledge at the level of "part of speech" and "written form," the basic stage of the hierarchy. The limited knowledge in these basic levels of word knowledge limits the production of vocabulary both in the PVLT and in the written task. This suggests that students still have a long way ahead to reach the final stages of word knowledge.

5. Gender

In the analysis of the effect of gender on receptive and productive vocabulary knowledge, the present study sheds light on the subject and points to the type of task and the type of learning context as factors that exert a stronger effect on the outcome than gender does. On the one hand, for the receptive tasks, neither school showed a statistically significant difference in the out-of-context task, the VLT, or in the in-context task, the reading test. In the case of the VLT, both in the CBT and FLT settings, the male students outperformed girls in terms of total word counts, but these differences were not statistically significant in the end. Whereas boys obtained higher scores on reading in the CBS, girls obtained slightly higher scores in the FLS. Overall, at the receptive level, we can safely conclude that gender does not exert an effect on the number of words that students recognize in tasks involving in-context or out-of-context receptive vocabulary.

On the other hand, regarding productive vocabulary, the FLS shows no gender effect either in the controlled or in the free productive vocabulary tasks. In both types of tasks, the analysis yields no statistically significant results favoring either one of the gender groups. The situation changes considerably in the CBS setting. In that setting, the male students obtained higher, statistically significant results in the controlled productive vocabulary task, the PVLT. Female students, on the other hand, produce statistically significant more words in the free writing task. While the only study that analyzes gender in connection with the PVLT found no statistically significant gender differences (Canga Alonso and Arribas García, 2014), the present study does record them in favor or male students. The fact that girls produce more words has been found in several studies analyzing written production; however, the statistically significant differences do not always appear (Ojeda Alba and Jiménez Catalán, 2007; Agustín Llach (2010b), van der Slik et al., 2015; Song et al., 2015).

The present research supports the idea of gender not exerting a critical influence over vocabulary learning, or SLA in general. The fact that production yields better results in one task for one gender group and in a different task for the other gender group points to the notion of the type of task, rather than gender itself, as the triggering factor in the different outcomes produced by each gender group. In this study, girls produce more items when the task allows them to do so through free writing. Males, however, make use of their productive vocabulary to a better extent when the task is controlled and thus they seem to make better use of the context in which the vocabulary item is immersed. More importantly, aside from the tasks themselves, what we find is that the learning setting itself plays a more fundamental role in the outcomes of vocabulary production, as these gender differences are not matched in the other school setting in the study. What this points to, therefore, is that the language-learning environment plays a key role in the use that students give to the vocabulary they have learned. The type of instruction, rather than the gender, exerts a difference in the results. Across the different tests, the CBS gender groups outperformed their counterparts in the FLS in the results of all the tests. This suggests that the school's methodology has a much stronger effect on the test results than gender does.

In line with Sunderland (2010), we can identify tendencies in the data, but these tendencies change as the task changes. More importantly, when these tendencies did appear, they came up only at the productive level and only in the CBS. The tendency that appears more frequently is that which shows no statistically significant differences in the sample across different tasks, especially in the FLS.

This situation suggests that other factors play a more fundamental role in SLA and in vocabulary learning, in particular, than gender does. Along those lines, the present study shows that aspects such as teaching methodology, motivation or affective factors are indeed more crucial than gender is in the process of SLA. In conclusion, gender may intervene in the results of certain tasks, but when overall competence in SLA is analyzed, the effects of gender do not appear to play a fundamental role in test results (Lin and Wu, 2003; Lynn et al., 2005; Law et al., 2013).

6. Motivation

The overall results of motivation show that the CBS students exhibit statistically significant higher degrees of motivation in terms of Desire to learn English, Integrative orientation and Instrumental orientation. Given the low internal consistency of the Instrumental orientation cluster for the CBS, the conclusions related to this cluster ought to be interpreted with utmost care. In light of theories such as Gardner's (1985, 2007) *integrativeness* theory and Dörnyei's (2006) *L2 Motivational Self System*, the results lead us to the following conclusions. Given the connection that has been established between the *Ideal L2 Self* and the concept of integrativeness, we can conclude that students from the CBS have shown a greater awareness of what Dörnyei (2006) calls their Ideal L2 Self—what they hope and dream to have in the future. This is associated with what Gardner (1985, 2007) identifies as a more *intrinsic* orientation to learning an L2. Students from the CBS also show greater *instrumental* orientation, associated with Gardner's idea of *extrinsic* motives and Dörnyei's Ought-to Self. In both cases, this motivation is derived from characteristics the individual assumes as future duties and responsibilities established primarily by the society and the surrounding context. The CBS students also exhibit a stronger desire to learn the L2. The fact that CBS students have greater motivation in these three key areas of motivation is critical in explaining the results of the present study.

In general, the concepts of integrativeness (Gardner, 1985, 2007) and the L2 Motivational Self System (Dörnyei, 2005) provide support for the students' learning experience as related to the educational context. Both authors emphasize the effect of the instructional context on the degree of motivation that students develop and maintain over the course of learning a second language. Here we have found that the CBS represents a better setting for establishing, developing and

maintaining levels of motivation that, according to the results of the different tests in this analysis, are more likely to nurture the process of SLA and vocabulary learning in particular. Motivation, as a fluctuating phenomenon, should always be interpreted with the utmost caution. However, thus far, the present study shows favorable results for CBI as a setting that offers both better outcomes in terms of vocabulary learning and higher instrumental and integrative motivation, as well as a greater desire to learn English as a second language.

GENERAL CONCLUSIONS

7. Vocabulary knowledge

Among the results of the present study, there is evidence that students' knowledge of words is mostly limited to the *form-meaning* link that they can assign to the different words and that the type of knowledge associated with the *use* of the words appears to be even more reduced at the productive level than at the receptive level. The results indicate that students' knowledge of the *form-meaning* connection is still in the early stages of development. Authors such as Schmitt, (2010a, 2010b) insist that the *form-meaning* knowledge represents a basic specification of word knowledge. This can also be explained through Nation's (2013) theory regarding what is involved to *know* a word at the receptive and productive level.

At the receptive level, the VLT and the reading test evaluated the students' ability to recognize the given words when they were encountered in written form. They evaluated whether these words were being used with specific meanings in context. At this level, students obtained relatively good results in the CBS while the FLS showed more limited knowledge. Even more important is the fact that this type of knowledge was measured through written channels. Had spoken forms been used, the results would have been lower as these forms require a greater command of the language. More advanced degrees of word knowledge such as meaning association, collocational use, or register use were not evaluated in this study, but given the limited degrees of knowledge at this basic level, the *form-meaning* knowledge, we could expect these degrees of knowledge to be out of reach for most students in the sample.

At the productive level, the results are even more challenging. The fact that the receptive knowledge of *form-meaning* link is greater than it is at the productive level is made evident by the means of words produced by students in the PVLT and in the writing task. Productively speaking, at the *form* level of knowledge, some students still have problems writing words with the correct spelling. This was evident in the PVLT results but even more so in the writing task, where many students' production was very limited. At the *meaning* level of knowledge, many students failed to provide the words that expressed the meaning that was necessary for both writing tasks. Finally, at the *use* level of knowledge, measured mainly in the free writing task, the total word counts were even more reduced and the correlation with the words produced in the PVLT was limited or non-existent.

This suggests that students are not fully capable of using the words correctly when they are reproducing their original thoughts and ideas in a free writing task. Given that the *form-meaning* connection is a basic step in the journey of vocabulary knowledge, attention should be given to these aspects in the classroom context. Only a strong basis of form-meaning knowledge of words can lead students to correct word usage, in both reception and production. This basic knowledge also precedes more in-depth types of word knowledge required for the successful use of language.

8. Bilingualism and bilingual education

Costa Rica, like many other countries around the globe, is making efforts to provide younger generations with a more functional use of foreign languages, English in particular. This pursuit of SLA is in line with the notion of *additive* bilingualism (Lambert, 1977, 1981) in the sense that Costa Rican citizens are making efforts to acquire a second language while consciously acknowledging the importance that knowing Spanish, their native tongue, entails. While granting importance to Spanish, Costa Ricans recognize the value of the English language and the many associated benefits that it brings to a society that maintains regional, social, and especially economic ties with the United States and several European countries with which communication is carried out mainly in English.

In this context, the learning of English as a foreign language finds the right environment for language development to take place. Contrary to the situation of the native aboriginal languages that are still today fighting for survival, English and

Spanish seem to be necessary and to some extent coexist in some contexts that allow for their presence in different stages of society. From means of communication to the educational sphere, or from the central fruit market to the many multinational companies and call centers, it is not uncommon to be exposed to the English language in use. Bilingual education in Costa Rica is, thus, guided by the economy and the politics of the country, as pointed out by Baker (2012), and discussed in section 1.2.1. Although more effective implementation in the form of a more effective language policy in the classroom is required, the English language seems to have found a niche in the Costa Rican society where its development finds support, at least on paper, from people in charge of policy making in education, as well as from those on the streets and in the classrooms.

Close inspection of the objectives put forward by the National Board of Education, however, shows that while the conditions seem to provide an adequate environment which should favor a type of EFL instruction that yields proficient second language speakers, reality is different and does not always seem to support this premise. The objectives (as stated by *Ministerio de Educación*, 2016, 2017) seem to target a more passive variety of bilingualism rather than its more active form. There is a palpable inclination toward more advanced development of passive skills such as reading over speaking or writing skills. This is made evident even in the choice of words contained in the objectives of reading that target students' ability at a more *complex* level in comparison to more *basic* capacities in speaking, for example. This is seen in practice where we encounter students who have different levels of a passive knowledge of English and very limited or lacking active skills in the use of the language. While this is possibly justified on the basis of preparing students for the compulsory tests that students take at the end of secondary education, it fails to serve a long-term purpose and function. Even as a pre-service teacher preparing to work in the system, I myself experienced a greater degree of training along the lines of the importance of input (Krashen, 1982, 2013; Krashen and Bland, 2014), rather than the key role played by output (Swain, 1995, 2000a).

Along these same lines, following Cummins' (1979, 1980, 2008) distinction between BICS and CALP, the Costa Rican educational system can be more readily identifiable through the less cognitively demanding BICS, through which students can deal with information relying mostly on contextual clues. The EFL classrooms are in apparent need of more challenging practices commonly associated with CALP tasks. While the context outside of class offers more demanding challenges for understanding and using the language, the classroom, especially the public EFL

system is not meeting the needs of the student population. The notion of CALP can be put into practice within the implementation of Content Based Teaching, the ability of understanding academically oriented information would have a great impact on students' development of L2 abilities. CALP can take many forms, some of which can be within the reach of the system and may remain cost effective. For example, at the moment, revision of the curricular content of subjects such as math, social studies or science in the English class could bring a double benefit to students. During English classes, students could review the material they are learning in other specific subjects. This would offer the opportunity to learn new concepts in English while at the same time learning about the subject through the English language. This practice would bring benefits to students as it allows for use of background knowledge in the L1 to learn new concepts in the L2, all the while putting Cummins' Interlanguage Hypothesis to work. This, of course, would pose greater demands on instructors, something the Board of Education and the universities that prepare pre-service teachers would have to anticipate.

Furthermore, although progressive implementation of CBT can bring great benefits to learners, there are other ways of aiding students to advance in the SLA process while strengthening their first language knowledge. Cummins' (2005, 2009) Interdependence Hypothesis accounts by this idea. It states that the levels of L2 attainment are intrinsically related to the students' knowledge of their L1 at the moment of exposure to L2 material. In this sense, the work on language development must begin in elementary education and it can be done through establishing strong knowledge bases of language knowledge of Spanish, the L1 of the students. Evidence from other studies points to Costa Rican students' limitations in the use of Spanish (Murillo Rojas, 2009; Araya Ramírez, 2012; Madrigal and Vargas, 2016).

This has a significant effect on how much students can benefit in the SL classroom. More effort should be made for academic development to take place in Spanish so that transfer of knowledge and skills can take place between the two languages at the level of language in general, or specifically in terms of skills and conceptual terms, which according to Cummins (2005) are all possible, given the right conditions. If a *Common Underlying Proficiency* system (Cummins, 2007) is strongly developed, the cognitive and academic benefits observed in academic performance which are associated with learning a second language, will take place. Advantage should be taken from *languaging* (Swain, 2006) and *translanguaging* (García, 2009; García and Wei, 2014) practices as well. In all of the cases

mentioned above, the L1 plays a fundamental role; hence the importance of strengthening Spanish knowledge alongside English knowledge.

Regarding the *Threshold Hypothesis* (Cummins, 1986, 2014a, 2014b), a few students in the present research could be classified as *less balanced* bilinguals but most are in the first threshold level (*limited* bilinguals) of linguistic competence. These levels are below the category of *balanced* bilinguals, the ultimate target of SLA (Baker, 2014). For that scenario, Cummins argues that students still portray cognitive disadvantages in comparison to balanced bilinguals in their age group and cannot enjoy the cognitive advantages of bilingualism because their overall functioning is similar to that of monolingual speakers. Explicit efforts to increase the language abilities would further the students' overall knowledge of the language and ensure the benefits associated with bilingualism.

In a different vein, subjecting the characteristics identified through this investigation to some of the tenets discussed by specialists such as Baker (2011), the individual bilingualism of the Costa Rican sample can be described as follows. The receptive *ability* of the sample has been identified as slightly more developed than its productive ability, particularly when comparing the reading ability of the subjects to their writing ability. The *use* of language for the present sample is mostly limited to the school context. Although many opportunities exist for other types of use outside the classroom setting for entertainment, travel or work, not everyone takes advantage of those opportunities.

As for the *balance* between languages, as the only official language of the country, Spanish is unmistakably more developed than English, especially in terms of its active uses. In regard to *age*, the EFL educational system attempts to form *sequential* bilinguals given that students already speak Spanish when they enter the educational system, when many are first exposed to English. In Costa Rica, the *development* of English is incipient in comparison to Spanish, inasmuch as students have a limited command of English while the development of Spanish is much higher, in line with its status as the mother tongue. The situation is slightly different in terms of the *cultural* aspect of bilingualism. Geographical closeness facilitates frequent trips to the U.S.A. for many Costa Ricans. A permanent influx of information through media allows for exposure to everyday events taking place there. This, along with the constant contact with foreigners, especially U.S tourists, has resulted in many Costa Rican citizens becoming highly bicultural when it comes to knowledge of the Costa Rican and the U.S. cultures. As a consequence, many students are particularly familiar with multiple aspects of that target culture.

The notion of *context*, as described by Baker (2011), is also framed by the situation described above. While English is not the official language of the country, there are enough opportunities to be in contact with the language in the surrounding context, which offers opportunities for additive bilingualism to develop. The final dimension in Baker's analysis calls for a combination of his tenets in the context under study; it has to do with the possibility of choosing whether to become bilingual or not. On the one hand, the students' context seems to guide learners toward *electing* English bilingualism in contrast to other possibilities given its presence and demand in the society. Additionally, this context seems to require learners to choose bilingualism in response to the *circumstantial* characteristics of the future job market and the needs of learners in their immediate future, either as a tool to travel abroad, to create connections with foreigners in the country, or to apply for a job.

In light of the detailed discussion contained in Chapter 2, dealing with strong forms of bilingual education, and aware of the pressing need for effective bilingual education, we could conclude that the Costa Rican setting could benefit greatly from a type of instruction resembling CLIL practices. Although Immersion Education and Dual Language Education are excellent programs of bilingual education, Costa Rica does not offer the ideal context for implementation of several of their basic principles. For example, while Immersion Education has proven to be extremely successful, immersion in the target language is not possible in the Costa Rican context to the extent that it occurs in settings such as that of Canada. Along these lines, Dual Language Instruction would not be possible either because we do not have access to large numbers of speakers of a second language to match the basic requirements for this type of program and which are common in contexts such as those of the U.S. where the program has produced positive results. CLIL, however, is implemented in Europe in contexts where a prestigious language (mostly English) is taught in a context where another prestigious language (Spanish in our context) is present at large in the society. While Costa Rica might not have the economic resources that Europeans countries have at their disposal, there are very good chances or implementing teaching practices where language learning is integrated into content learning. These are practices that the Ministry of Education and the different organizations and institutions involved in SLA policy, planning, and implementation should be aiming at.

In general, when applying Grosjean's (2013) definition of bilingualism and what being bilingual entails, we must conclude that the Costa Rican society is not a bilingual society, nor are the students in the present study bilingual, because they

do not use English in their everyday lives. As mentioned above, many of these students are likely to become bilingual in the process of finding a bilingual position in the work force. The type of bilingualism found in the Costa Rican context does not usually reach equal levels of proficiency in Spanish and English, as proposed by Skutnabb-Kangas and McCarthy (2008). On the contrary, it often results in non-fluent speakers of the L2, as described by Edwards (2006). While today's definitions of the concept are much more flexible than those found several decades ago, we still must use the concepts of bilingualism and bilingual speakers with care, when conveying the idea that speakers have the ability to use the language to communicate effectively.

The notion of bilingual education as a reflection of *politics* and its association with the *economic perspective of education*, as discussed in Baker (2012), seems to find a place in the Costa Rican context. Through this perspective, we could explain the interest in the proposal for the Multilingual Costa Rica project (*Costa Rica* Multilingüe, Arias-Sánchez et al., 2008). Although theoretically worthwhile, it has lacked effective implementation in the classroom. The problem may spring from the wrong conceptualization that appears to be present in the mindset of most, by which many continue to conceive traditional language education as a form of bilingual education when that is clearly not the case (García, 2009; Bialystok, 2016). The dearth of programs using the target language as a medium of instruction is undeniable, especially in the public sector in which most students are enrolled. The system still supports a mixture of approaches. Particular emphasis is said to be given to the communicative approach, but there is still strong presence of a combination of methods such as the Grammar Translation, Task-Based Teaching or Audiolingual methods, included within what Baker and Wright (2017) define as Mainstream Education with Foreign Language Teaching, and describe as a weak form of bilingual education. The students in the context analyzed are facing challenges that demand more effective forms of bilingual education, one where English has to be used as a medium of instruction.

9. Implications and closing remarks

The importance of the relationship existing between vocabulary knowledge in connection with language skills, reading and writing in particular, has been established in this study. Both school settings have shown themselves to require improvement in a number of aspects related to vocabulary learning and SLA in

general. Overall, however, the results of the present study show that CBI offers better solutions for the current, pressing need of preparing individuals for a very competitive society that requires individuals to be functional users of second languages, English in particular. The mainstream Foreign Language setting, must identify effective ways of instructing learners for a more active use of the language for communication. The population represented here for this type of instruction lags behind in terms of both receptive and productive vocabulary measures. The Costa Rican EFL system requires revision and implementation of methods that respond to the challenges posed by today's society. Using English as medium of instruction is an adequate way to face this challenge. It would represent a true form of bilingual education, a type of language instruction that yields the results sought after when we embark on the journey of SLA.

The cognitive advantages associated with bilingualism are multiple. While the present study did not engage in the task of identifying these cognitive advantages, based on Cummin's (2005, 2007) theories, we could conclude that students in our sample do not meet the language proficiency required to have access to these benefits. Greater efforts are required on the part of policy makers, universities in charge of preparing pre-service teachers, instructors, students and society, in general, to provide future generations with the tools required to become functional bilinguals. More importantly, bilingualism should be seen through the lens of the dynamic phenomenon that García and Wei (2014) describe: a single system containing features of the two languages that make up the bilingual speaker, represented through the linguistic systems speakers use. By a better grasp of this complex phenomenon, we will begin to enjoy the language, along with the cognitive and social advantages associated with bilingualism. This can only be achieved through strong forms of bilingual education.

It is evident from the perspective of the vocabulary counts in the present study that the treatment given to vocabulary in the Costa Rican EFL context should be revised. Schmitt (2008) describes the importance and effectiveness of explicit vocabulary teaching especially in initial stages of second language acquisition. Dóczi and Kormos (2016) suggest that explicit instruction produces better results when it comes to vocabulary learning. While many may dislike this more direct form of instruction, evidence supports its effectiveness. Nation and Meara (2010) argue that instructors must assess their students' needs in terms of vocabulary learning, as a basis to direct students to learn the most useful vocabulary. In addition, File and Adams (2010) assure that any kind of instruction is better than

no instruction at all because of its impact on retention and general vocabulary learning.

The ultimate challenge of implementing vocabulary learning strategies and different teaching practices is to help students find ways to add more elements to the items that conform the semantic fields in their memories so that their semantic network can, in turn, be strengthened. Semantic networks play a key role in the growth of the mental lexicon. The aim should be directed to making new concepts part of long-term memory for them to be readily available to students for their use of vocabulary in real language: reading and writing as they were analyzed here.

In addition to the more specific findings outlined above in the field of vocabulary acquisition as a whole, this study has provided a unique design for a framework of aspects to consider especially in contexts with little prior work in the area. This basis brings together diverse theoretical concepts and results from previous experimentation for immediate use in the analysis of the specific studies to be carried out in the target area. Thus, this design may also serve as a guide for future research in other similar settings and conditions. This study is unique in Central America and in Costa Rica in particular. This in itself is an essential contribution to the field of vocabulary acquisition, over and beyond other findings discussed above. Moreover, it has provided new results for the growing body of data accumulating around the concept of content-based instruction such as CLIL, in comparison to other approaches and in different contexts.

BIBLIOGRAPHY

Ackerman, P. L., & Ellingsen, V. J. (2014). Vocabulary overclaiming—A complete approach: Ability, personality, self-concept correlates, and gender differences. *Intelligence*, *46*, 216–227. doi:dx.doi.org/10.1016/j.intell.2014.07.003

Adams, J. N., & Swain, S. (2010). Introduction. In J. N. Adams, M. Janse, & S. Swain (Eds.), *Bilingualism in ancient society: Language contact and the written text* (pp. 1–20). Oxford: Oxford University Press.

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, *24*(4), 425–438. doi:doi.org/10.1093/applin/24.4.425

Agustín Llach, M. P. (2009). The role of gender in lexical transfer in EFL written compositions across grades: A preliminary study. Presented at the 19th EUROSLA Conference, University College Cork, Ireland.

Agustín-Llach, M. P. (2010a). An overview of variables affecting lexical transfer in writing: A review study. *International Journal of Linguistics*, *2*(1:E2), 1–17. doi:https://doi.org/10.5296/ijl.v2i1.445

Agustín-Llach, M. P. (2010b). Exploring the role of gender in lexical creations. In R. M. Jiménez Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 74–92). London: Palgrave Macmillan. doi:10.1111/j.1473-4192.2011.00286.x

Agustín-Llach, M. P. (2015). The effects of the CLIL approach in young foreign language learners' lexical profiles. *International Journal of Bilingual Education and Bilingualism*, 5, 1–17. doi:doi.org/10.1080/13670050.2015.110 3208

Agustín-Llach, M. P., & Canga-Alonso, A. (2016). Vocabulary growth in young CLIL and traditional EFL learners: Evidence from research and implications for education. *International Journal of Applied Linguistics*, *26*(2), 211–227. doi: 10.1111/ijal.12090

Agustín-Llach, M. P., & Fernández Fontecha, A. (2014). Lexical variation in learners' responses to cue words: The effect of gender. In R. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 69–81). London: Springer. doi: https://doi.org/10.1007/978-94-007-7158-1_5

Agustín-Llach, M. P., & Terrazas-Gallego, M. (2009). Examining the relationship between receptive vocabulary size and written skills of primary school learners. *Atlantis*, *31*(1), 129–147. Retrieved from http:/www.jstor.org/stable/41055350

Agustín-Llach, M. P., & Terrazas-Gallego, M. (2012). Vocabulary knowledge development and gender differences in a second language. *Estudios de Lingüística Inglesa Aplicada*, *12*, 45–75. Retrieved from http://revistas.uned.es/index.php/ELIA/article/view/18027

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. Oxford: Wiley-Blackwell.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Anthony, L. (2014). AntWordProfiler (Version 14.0). Tokyo: Waseda University. Retrieved from www.laurenceanthony.net/software/antwordprofiler/

Araya Ramírez, J. (2012). La competencia de la expresión oral de niños escolares en Costa Rica: El componente léxico. *Káñina, Revista de Artes y Letras*, *XXXVI*(1), 169–183. Retrieved from https://revistas.ucr.ac.cr/index.php/kanina/article/view/File/1231/1294

Arias Sánchez, O., Arias Sánchez, R., Mata Segrega, A., & Ruiz Gutiérrez, M. V. *La creación del ente rector del Plan Nacional de Inglés*, Pub. L. No. D34535-48705, N° 34535-MP-MEP-COMEX 26 (2008). Retrieved from www.gaceta.go.cr/pub/2008/05/30/COMP_30_05_2008.pdf

Baddeley, A. (2015). Working memory in second language learning. In Z. E. Wen, M. Borges Mota, & A. McNeill (Eds.), *Working memory in second language acquisition and processing* (pp. 17–28). Bristol, UK: Multilingual Matters.

Baetens Beardsmore, H. (2008). Multilingualism, cognition and creativity. *International CLIL Research Journal*, *1*(1), 4–19.

Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed.). Bristol, UK: Multilingual Matters.

Baker, C. (2012). Bilingual education. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 294-304). Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780195384253.013.0020

Baker, C. (2014). *A parents' and teachers' guide to bilingualism* (4th ed.). Bristol, UK: Multilingual Matters.

Baker, C., & Wright, W. E. (2017). *Foundations of bilingual education and bilingualism* (6th ed.). Bristol, UK: Multilingual Matters.

Baker, S. C., & MacIntyre, P. D. (2003). The role of gender and immersion in communication and second language orientations. *Language Learning*, *53*(S1), 65–96. doi:10.1111/0023-8333.00119

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, *73*, 116–130. doi:10.1016/j.jml.2014.03.002

Banegas, D. L. (2016). Teachers develop CLIL materials in Argentina: A workshop experience. *Latin American Journal of Content and Language Integrated Learning*, *9*(1), 17–36. doi:10.5294/laclil.2016.9.1.2

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279. Retrieved from https:/doi.org/10.1093/ijl/6.4.253

Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and university level vocabulary tests. *Language Testing*, *16*(2), 131–162. doi:10.1177/026553229901600202

Bialystok, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology*, *24*(4), 560–567. doi:dx.doi.org/10.1037/0012-1649.24.4560

Bialystok, E. (2005). Consequences of bilingualism for cognitive development. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 417–432). New York, Oxford University Press.

Bialystok, E. (2016). Bilingual education for bilingual children: Review of the effects and consequences. *International Journal of Bilingual Education and Bilingualism*. (pp. 1–14) doi.org/10.1080/13670050.2016.1203859

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, *122*, 67–73. doi:10.1016//j.cognition.2011.08.003

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics*, *19*(1), 69–85. doi:doi.org/10.1017/S0142716400010584

Bialystok, E., Majumder, S., & Martin, M. M. (2003). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics*, *24*(1), 27–44. doi:10.1017.S014271640300002X

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*(1), 103–119. doi:doi.org/10.1177/13670069050090010701

Bialystok, E., Peets, K. F., & Moreno, S. (2014). Producing bilinguals through immersion education: Development of metalinguistic awareness. *Applied Psycholinguistics*, *35*, 177–191. doi:10.1017/S0142716412000288

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, *112*, 494–500. doi:10.1016/j.cognition.2009.06.014

Bingham Wesche, M. (2002). Early French immersion: How has the original Canadian model stood the test of time? In P. Burmeister, T. Piske, & A. Rohde (Eds.). *An integrated view of language development. Papers in honor of Henning Wode* (pp. 357–379). Wissenschaftlicher: Trier.

Bingham Wesche, M., & Paribakht, T. S. (2000). Reading-based exercises in second language vocabulary learning: An introspective study. *The Modern Language Journal*, *84*(2), 196–213. doi:10.1111/0026-7902.00062

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Aktürk Ari, B. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*, 1461–1476. doi:10.1017/S0142716416000060

Bloom, B. S. (1956). *Taxonomy of educational objectives* (Vol. 1: Cognitive domain). New York: McKay.

Bloomfield, L. (1933). *Language*. London: George Allen & Unwin.

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic structure in vocabulary knowledge interacts with lexical and sentence processing in infancy. *Child Development*, 87(6), 1893–1908. doi:10.1111/cdev.12.554

Bourgoin, R. (2014). The predictive effects of L1 and L2 early literacy indicators on reading in French immersion. *The Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes*, *70*(3), 355–380. doi:10.3138/cmlr.2346

Brantmeier, C. (2010). More than words: Inferential and incorrect units recalled. In R. M. Jiménez Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 23–43). London: Palgrave Macmillan. doi:10.1111/j.1473-4192.2011.00286.x

Bruner, J. (1983). *Child's talk: Learning to use language*. New York: W. W. Norton.

Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, *39*, 523–532. doi:10.1016/j.system.2011.08.002

Bruton, A. (2013). CLIL: Some of the reasons why and why not. *System*, *41*, 587–597. doi:10.1016/j.system.2013.07.001

Bruton, A. (2015). CLIL: Detail matters in the whole picture. More than a reply to J. Hüttner and U. Smit (2014). *System*, *53*, 119–128. doi:10.1016/j.system.2015.07.005

Calvo, A., & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition*, 130, 278–288. doi:10.1016/j.cognition.2013.11.015

Cambridge English Language Assessment. (2011). *Cambridge English ESOL skills for life: Entry 3*. Cambridge: Cambridge University Press. Retrieved from http:// www.cambridgeenglish.org/exams-and-tests/skills-for-life/

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research, 6*(2), 145–173. doi:10.1191/1362168802lr103oa

Canga-Alonso, A. (2013a). The receptive vocabulary of Spanish 6th-grade primary-school students in CLIL instruction: A preliminary study. *Latin American Journal of Content and Language Integrated Learning*, *6*(2), 22–41. doi:10.5294/ laclil.2013.6.2.2

Canga-Alonso, A. (2013b). Receptive vocabulary size of secondary Spanish EFL learners. *Revista de Lingüística y Lenguas Aplicadas*, *8*, 66–75. doi:10.4995/ rlyla.2013.1180

Canga-Alonso, A., & Arribas-García, M. (2014). Productive vocabulary knowledge of Spanish EFL learners. *Revista Electrónica de Lingüística Aplicada*, *1*, 39–56. Retrieved from dialnet.unirioja.es/servlet/articulo?codigo=5031517

Castro, D. C., Páez, M. M., Dickinson, D. K., & Frede, E. (2011). Promoting language and literacy in young dual language learners: Research, practice, and policy. *Child Development Perspectives, 5*(1), 15–21. doi:10.1111/j.1750-8606.2010. 00142.x

Castro-García, D. (2014). Spaced learning: Its implications in the language classroom. *Revista de Lenguas Modernas*, 20, 241–257.

Castro-García, D. (2017). Are we preparing secondary students for a productive use of vocabulary in English as their second language? *Porta Linguarum*, 28, 141-155.

Castro-García, D. (2017). Bilingualism, CLIL, and immersion programs: The case of Costa Rica. In G. Nieto Caballero (Ed.), *Nuevas aportaciones al estudio de la enseñanza y aprendizaje de lenguas* (pp. 123–136). Cáceres: Universidad de Extremadura.

Castro-García, D. (2017). Receptive vocabulary measures for EFL Costa Rica high school students. IJES, 17(2), 81–99. doi:10.6018/ijes/2017/2/265681

Castro-García, D. (2017). *Three forms of bilingual education: Immersion, Dual Language, and CLIL*. Heredia: Ediciones Escuela de Literatura y Ciencias del Lenguaje.

Cerdas E., D. (2017, April 24). Colegiales manejan un promedio de tan solo 2.200 palabras. *La Nación*. Retrieved from http://www.nacion.com/nacional/educacion/ Colegiales-manejan-promedio-solo-palabras_0_1629637119.html

Chung, T.M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, *15*(2), 103–116. Retrieved from http://nflrc.hawaii.edu/rfl/ October2003/chung/chung.pdf

Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System*, *32*, 251–263. doi:https://doi.org/10.1016/j.system.2003.11.008

Clément, R. (1980). Ethnicity, contact and communicative competence in a second language. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Language: Social psychological perspectives* (pp. 147–154). Oxford: Pergamon.

Cobb, C. (2015). Is French immersion a special education loophole? And does it intensify issues of accessibility and exclusion? *International Journal of Bilingual Education and Bilingualism*, *18*(2), 170–187. doi:10.1080/13670050.2014. 887052

Coombe, C. (2011). Assessing vocabulary in the language classroom. *Malaysian English Language Teaching Association*, 111–124.

Costa Rica, INEC. (2017). Censo 2011. Indicadores demográficos y sociales según provincia, cantón y distrito. [Government Page].

Council of the European Communities, & Commission of the European Communities. (1992). Maastricht Treaty on European Union. Office for Official Publications of the European Communities.

Coyle, D. (1999). Supporting students in content and language integrated learning contexts: Planning for effective classrooms. In J. Masih (Ed.), *Learning through a foreign language: Models, methods and outcomes* (pp. 46–69). London, UK: Centre for Information on Language Teaching and Research.

Coyle, D. (2006). Content and language integrated learning: Motivating learners and teachers. *Scottish Languages Review, 13*, 1–18.

Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism*, *10*(5), 543–562. doi:10.2167/beb459.0

Coyle, D., Hood, P., & Marsh, D. (2010). *Content and language integrated learning*. Edinburgh, UK: Cambridge University Press.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222–251. doi:10.3102/00346543049002222

Cummins, J. (1980). The construct of language proficiency in bilingual education. In J. E. Alatis (Ed.), *Current issues in bilingual education* (pp. 81–103). Washington DC: Georgetown University Press.

Cummins, J. (1984). Bilingualism and minority-language children. In J. Cummins, S. Lapkin, & M. Swain (Eds.), *Language and literacy*. Ontario: The Ontario Institute for Studies in Education.

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review, 56*(1), 18–36. doi:10.17763/haer.56.1.b327234461607787

Cummins, J. (2000). Interdependence of first and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70–89). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511620652

Cummins, J. (2001). *Negotiating identities: Education for empowerment in a diverse society.* California: Association for Bilingual Education.

Cummins, J. (2005). Teaching for cross-language transfer in dual language education: Possibilities and pitfalls. In *TESOL symposium on dual language education: Teaching and learning two languages in the EFL setting*. Istanbul, Turkey.

Cummins, J. (2007). Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Journal of Applied Linguistics, 10*(2), 221–240.

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 2: Literacy, pp. 71–83). New York: Springer Science and Business Media. doi:10.1007/978-3-319-02252-9_6

Cummins, J. (2009). Fundamental psychological and sociological principles underlying educational success for linguistic minority students. In A.K. Mohanty, M. Panda,

R. Phillipson & T. Skutnabb-Kangas (Eds.), *Multilingual education for social justice: Globalizing the local* (19–35). New Delhi: Orient BlackSwan.

Cummins, J. (2014a). Rethinking pedagogical assumptions in Canadian French immersion programs. *Journal of Immersion and Content-Based Language Education*, *2*(1), 3–22. doi:10.1075/jicb.2.1.01cum

Cummins, J. (2014b). To what extent are Canadian second language policies evidence-based? Reflections on the intersections of research and policy. *Frontiers in Psychology*, *5*, 1–10. doi:10.3389/fpsyg.2014.00358

Cummins, J., & Swain, M. (1986). Bilingualism in education. In C. N. Candlin (Ed.), *Applied Linguistics and Language Study*. New York: Longman Group Limited.

Cummins, J., & Swain, M. (2014). *Bilingualism in education: Aspects of theory, research and practice*. Oxon: Routledge.

Cummins, J., Bismilla, V., Chow, P., Cohen, S., Giampapa, F., Leoni, L., & Sastri, P. (2005). Affirming identity in multilingual classrooms. *Educational Leadership*, *63*(1), 38–43.

Curtis, A. (2012). Colombian teachers' questions about CLIL: What can teachers' questions tell us? (Part II). *Latin American Journal of Content and Language Integrated Learning*, *5*(2), 1–12. doi:10.5294/laclil.2012.5.2.6

Dalton-Puffer, C. (2008). Outcomes and processes in content and language integrated learning (CLIL): Current research from Europe. In W. Delanoy & L. Volkmann (Eds.), *Future perspectives for English language teaching* (pp. 1–19). Heidelberg: Carl Winter.

Dalton-Puffer, C. (2011). Content and language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, *31*, 182–204. doi:10.3726/978-3-0351-0171-3

Dalton-Puffer, C., Nikula, T., & Smit, U. (2010a). Charting policies, premises and research on content and language integrated learning. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL Classrooms* (pp. 1–18). Amsterdam: John Benjamins. doi:10.1075/aals.7.01dal

Dalton-Puffer, C., Nikula, T., & Smit, U. (2010b). Language use and language learning in CLIL: Current findings and contentious issues. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 279–291). Amsterdam: John Benjamins. doi:10.1075/aals.7.14dal

Dalton-Puffer, C., & Smit, U. (2013). Content and language integrated learning: A research agenda. *Language Teaching*, *46*(4), 545–559. doi:10.1017/s0261444813000256

Daniel, S. M., & Conlin, L. (2015). Shifting attention back to students within the sheltered instruction observation protocol. *TESOL Quarterly*, *49*(1), 169–187. doi:10.1002/tesq.213

de Jong, E. J. (2014). Program design and two-way immersion programs. *Journal of Immersion and Content-Based Language Education*, *2*(2), 241–256. doi:10.1075/jicb.2.2.06jon

Díez Prado, M. (2010). Gender and L1 influence on EFL learners' lexicon. In R. M. Jiménez Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 44–73). London: Palgrave Macmillan. doi:10.1111/j.1473-4192.2011.00286.x

Dóczi, B., & Kormos, J. (2016). *Longitudinal development in vocabulary knowledge and lexical organization*. Oxford: Oxford University Press.

Dörnyei, Z. (1994). Understanding L2 motivation: On with the challenge! *The Modern Language Journal*, 78(4), 515–523. doi:10.2307/328590

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching*, *31*, 117–135. doi:10.1017/S0261444800001315x

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah: Lawrence Erlbaum.

Dörnyei, Z. (2009a). Motivation and the vision of knowing a second language. In B. Beaven (Ed.), *IATEFL 2008: Exeter conference selections* (pp. 16–22). Canterbury: IATEFL.

Dörnyei, Z. (2009b). The L2 motivation self system. In Z. Dörnyei & E. Ushiola (Eds.), *Motivation, language identity and the L2 self* (pp. 9–42). Bristol, UK: Multilingual Matters.

Dörnyei, Z. (2010). Researching motivation: from integrativeness to the ideal L2 self. In S. Hunston & D. Oakey (Eds.), *Introducing applied linguistics: Concepts and skills* (pp. 74–86). New York: Routledge.

Dörnyei, Z. (2014). Motivation in second language learning. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (4th ed., pp. 518–531). Boston: National Geographic Learning/Cengage Learning.

Dörnyei, Z. (2017). Conceptualizing learner characteristics in a complex, dynamic world. In L. Ortega & Z. H. Han (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman* (pp. 79–96). Amsterdam: John Benjamins.

Dörnyei, Z., & Chan, L. (2013). Motivation and vision: An analysis of future L2 self images, sensory styles, and imagery capacity across two target languages. *Language Learning*, *63*(3), 437–462. doi:10.1111/lang.l2005

Dupuy, B. C. (2000). Content-based instruction: Can it help ease the transition from beginning to advanced foreign language classes? *Foreign Language Annals*, *33*(2), 205–223. doi:https://doi.org/10.1111/j.1944-9720.2000.tb00913.x

Echevarria, J., Vogt, M., & Short, D. J. (2008). *Making content comprehensible for English learners: The SIOP model* (3rd ed.). Boston, MA: Pearson. doi:10.5617/adno.1092

Edwards, J. (2006). Foundations of bilingualism. In T. K. Bathia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 7–31). Oxford: Blackwell.

Elgort, I., & Warren, P. (2014). L2 Vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*(2), 365–414. doi:10.1111/lang.12052

Ellis, A. W., Burani, C., Izura, C., Bromiley, A., & Venneri, A. (2006). Traces of vocabulary acquisition in the brain: Evidence from covert object naming. *NeuroImage*, *33*, 958–968. doi:10.1016/j.neuroimage.2006.07.040

European Commission. (1995). White paper on education and training. Commission of the European Communities.

European Commission. (1996). Green paper: Promoting the learning mobility of young people. Commission of the European Communities.

Feinauer, E., & Howard, E. R. (2014). Attending to the third goal: Cross-cultural competence and identity development in two-way immersion programs. *Journal of Immersion and Content-Based Language Education*, *2*(2), 257–272. doi:10.1075/jicb.2.2.07fei

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology*, *25*(10), 1261–1278.doi:10.1080/02687038.2011.606974

Fernández Fontecha, A. (2010). Gender and motivation in EFL vocabulary production. In R. M. Jiménez Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 93–116). London: Palgrave Macmillan. doi:10.1111/j.1473-4192.2011.00286.x

Fernández Fontecha, A. (2014). Receptive vocabulary knowledge and motivation in CLIL and EFL. *Revista de Lingüística y Lenguas Aplicadas*, *9*, 23–32. doi:10.4995/rlyla.2014.2077

Fernández Fontecha, A., & Canga-Alonso, A. (2014). A preliminary study on motivation and gender in CLIL and non- CLIL types of instruction. *IJES*, *14*(1), 21–36. Retrieved from http://revistas.um.es/ijes

Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias-Blanco, J. M. (2017). Analysing students' content-learning in science in CLIL vs. non-CLIL programmes: Empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism*. doi:10.1080/13670050.2017.1294142

File, K. A., & Adams, R. (2010). Should vocabulary instruction be integrated or isolated? *TESOL Quarterly*, *44*(2), 222–249. doi:10.5054/tq.2010.219943

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, *40*(2), 273–293. doi:https://doi.org/10.2307/40264523

Freeman, R. (2000). Contextual challenges to dual-language education: A case study of a developing middle school program. *Anthropology & Education Quarterly*, *31*(2), 202–229. doi:10.1525/aeq.2000.31.2.202

Gabillon, Z., & Rodica, A. (2015). Content and language integrated learning: In search of a coherent conceptual framework. In *Official Conference Proceedings* (pp. 311–324). Brighton: The International Academic Forum (IAFOR).

García, O. (2009). *Bilingual education in the 21st century: A global perspective*. West Sussex: Wiley-Blackwell.

García, O., & Lin, A. (2016). *Translanguaging in bilingual education*. Dordretch: Springer.

García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Basingstoke: Palgrave Macmillan.

Gardner, R. C. (1985). *Social psychology and second language learning: The roles of attitudes and motivation*. London: Edward Arnold.

Gardner, R. (2004). *The Attitude/Motivation Test Battery (Technical report)*. Ontario: University of Western Ontario.

Gardner, R. C. (2007). Motivation and second language acquisition. *Porta Linguarum*, *8*, 9–20. Retrieved from http://hdl.handle.net/10481/31616

Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second language acquisition. *Canadian Journal of Psychology*, *13*, 191–197. Retrieved from https://eric.ed.gov/?id=ED031968

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.

Gardner, R. C., & MacIntyre, P. D. (1992). A student's contributions to second language learning. Part I: Cognitive variables. *Language Teaching*, *25*, 211–220. doi:10.1017/S026144480000700X

Gardner, R. C., & MacIntyre, P. D. (1993). A student's contributions to second-language learning. Part II: Affective variables. *Language Teaching*, *26*, 1–11. doi:10.1017/S0261444800000045

Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2015). Writing development under CLIL provision. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual environments* (pp. 139–161). New York: Springer. doi:10.1007/978-3-319-11496-5_9

Genesee, F., & Jared, D. (2008). Literacy development in early French immersion programs. *Canadian Psychology*, *49*(2), 140–147. doi:10.1037/0708-5591.49.2.140

Gibson, J., & Roy, S. (2015). Canadian parents for French: How a grassroots organization has contributed to the advancement of Canada's official languages policy. *Journal of Immersion and Content-Based Language Education*, *3*(2), 218–240. doi:10.1075/jicb.3.2.03gib

Gorjian, B., & Javadifar, M. (2013). Effects of gender and passage content on multiple-choice reading comprehension test. *Procedia: Social and Behavioral Sciences*, (84), 723–727. doi:10.1016/j.sbspro.2013.06.634

Granados Sirias, M., & Chavez Soto, Y. (2018, March 21). Fwd: Info LEB's: Lineamientos para el desarrollo y regulación de los liceos bilingües públicos de Costa Rica.

Granados Sirias, M. (2016, January 20). Fwd Info LEBS.

Grosjean, F. (1997). The bilingual individual. *Interpreting*, *2*(1/2), 163-187. doi:10.1075/intp.2.1-2.07gro

Grosjean, F. (2010). *Bilingual: Life and reality*. Cambridge, MA: Harvard University Press. doi:10.1558/sols.v6i3.595

Grosjean, F. (2013). Bilingual and monolingual language modes. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. New Jersey: Blackwell. doi:10.1002/9781405198431.wbeal0090

Harrison, G. L., Goegan, L. D., Jalbert, R., McManus, K., Sinclair, K., & Spurling, J. (2016). Predictors of spelling and writing skills in first- and second-language learners. *Reading and Writing*, *29*(1), 69–89. doi:doi.org/10.1007/s11145-015-9580-1

Harrop, E. (2012). Content and Language Integrated Learning (CLIL): Limitations and possibilities. *Encuentros*, *21*, 57–70.

Harvey, T., Tihinen, T., Määttä, K., & Uusiautti, S. (2013). Content and Language Integrated Learning in practice: Comparison of three cases. *International Journal of Academic Research in Progressive Education and Development*, *2*(2), 86–98. Retrieved from http://www.hrmars.com/admin/pics/1775.pdf

Heaton, J. B. (1990). *Classroom testing*. Essex: Longman.

Henderson, L., Weighall, A., & Gaskell, G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of Experimental Child Psychology*, *116*, 572–592. doi:10.1016/j.jecp.2013.07.004

Heras, A., & Lasagabaster, D. (2015). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*, *19*(1), 70–88. doi:10.1177/1362168814541736

Hernández, A. M. (2015). Language status in two-way bilingual immersion: The dynamics between English and Spanish in peer interaction. *Journal of Immersion and Content-Based Language Education*, *3*(1), 102–126. doi:10.1075/jicb.3.1.05her

Hernández-Herrero, A. (2008). El inglés en Costa Rica: requisito indispensable en un mundo globalizado. *Actualidades Investigativas en Educación*, *8*(2), 1–23. https://doi.org/10.15517/aie.v8i2.9332

Hipfner-Boucher, K., Pasquarella, A., Chen, X., & Deacon, S. H. (2016). Cognate awareness in French immersion students: Contributions to grade 2 reading comprehension. *Scientific Studies of Reading*, *20*(5), 389–400. doi:10.1080/10888438.2016.1213265

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, *8*(2), 689–696. Retrieved from https://eric.ed.gov/?id=EJ474555

Hoffmann, C. (1991). *An introduction to bilingualism* (Vol. 2). New York: Longman. doi.org/10.1093/applin/15.3.351

Hopewell, S., & Escamilla, K. (2014). Biliteracy development in immersion contexts. *Journal of Immersion and Content-Based Language Education*, *2*(2), 181–195. doi:10.1075/jicb.2.2.02hop

Hosoda, C., Tanaka, K., Nariai, T., Honda, M., & Hanakawa, T. (2013). Dynamic neural network reorganization associated with second language vocabulary acquisition: A multimodal imaging study. *The Journal of Neuroscience*, 33(34), 13663–13672. doi:10.1523/JNEUROSCI.0410-13.2013

Howard, E. R., Sugarman, J., Christian, D., Lindholm-Leary, K. J., & Rogers, D. (2007). *Guiding principles for dual language education* (2nd ed.). Washington, DC: Center for Applied Linguistics.

Howe, E. R. (2014). A narrative of teacher education in Canada: Multiculturalism, technology, bridging theory and practice. *Journal of Education for Teaching*, *40*(5), 588–599. doi:10.1080/02607476.2014.956540

Hu, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430. Retrieved from https://eric.ed.gov/?id=EJ626518

Hummel, K. M. (2014). *Introducing second language acquisition: Perspectives and practices*. Oxford, UK: John Wiley & Sons.

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language*, *17*(1), 23–59. Retrieved from https://eric.ed.gov/?id=EJ689121

Hüttner, J., & Smit, U. (2014). CLIL (Content and Language Integrated Learning): The bigger picture. A response to: A. Bruton. 2013. CLIL: Some of the reasons why... and why not. *System*, *41* (2013): 587-597. *System*, *44*, (2014): 160–167. doi:10.1016/j.system.2014.03.001

IBM Corp. (2011). IBM SPSS Statistics for Windows (Version 20.0). Armonk, NY: IBM Corp.

Jexenicker, S., & Dalton-Puffer, C. (2010). The CLIL differential: Comparing the writing of CLIL and non-CLIL students in higher colleges of technology. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (Vol. 7, pp. 169–189). Amsterdam: John Benjamins.

Jiménez-Catalán, R. M. (2003). Sex differences in L2 vocabulary learning strategies. *International Journal of Bilingual Education and Bilingualism*, 13(1), 54–77. doi:10.1111/1473-4192.00037

Jiménez Catalán, R. M. (2010). Gender tendencies in EFL across vocabulary tests. In R. M. Jiménez Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 117–138). London: Palgrave Macmillan. doi:10.1111/j.1473-4192.2011.00286.x

Jiménez Catalán, R. M., Ruiz de Zarobe, Y., & Cenoz Inagui, J. (2006). Vocabulary profiles of English foreign language learners in English as a subject and as a vehicular language. *Vienna English Working Papers*, *15*(3), 23–27.

Jiménez Catalán, R. M., & Terrazas Gallego, M. (2005). The receptive vocabulary of English foreign language young learners. *Journal of English Studies*, *5-6*, 173–191. Retrieved from https://publicaciones.unirioja.es/ojs/index.php/jes/article/viewFile/127/107

Juan-Garau, M., & Salazar-Noguera, J. (Eds.). (2015a). *Content-based language learning in multilingual educational environments*. New York: Springer. doi:10.1007/978-3-319-11496-5

Juan-Garau, M., & Salazar-Noguera, J. (2015b). Introduction: The relevance of CLIL education in achieving multilingualism on the global stage. In M. Juan-Garau & J. Salazar-Noguera (Ed.), *Content-based language learning in multilingual environments* (pp. 1–10). Springer International, Switzerland. doi:10.1007/978-3-319-11496-5_1

Knouzi, I., Swain, M., Lapkin, S., & Brooks, L. (2010). Self-scaffolding mediated by languaging: Microgenetic analysis of high and low performers. *International Journal of Applied Linguistics*, *20*(1), 23–49. doi:10.1111/j.1473-4192.2009.00227.x

Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the Input Hypothesis. *The Modern Language Journal*, 73(4), 440–464. doi:10.2307/326879

Krashen, S. (1991). The input hypothesis: An update. In J. E. Alatis (Ed.) *Georgetown University Round Table on Languages and Linguistics 1991*. Washington, D.C.: Georgetown University Press. 409-431.

Krashen, S. (2003). *Explorations in language acquisition and use*. Portsmouth: Heinemann.

Krashen, S. (2013). The case for non-targeted, comprehensible input. *Journal of Bilingual Education Research & Instruction*, *15*(1), 102–110. Retrieved from http://www.sdkrashen.com/content/articles/nontargeted_input.pdf

Krashen, S., & Bland, J. (2014). Compelling comprehensible input, academic language and school libraries. *Children's Literature in English Language Education Journal*, *2*(2), 1–12. Retrieved from https://zenodo.org/record/322586#.WuPUQNPOVn0

Lambert, W. E. (1977). The effects of bilingualism on the individual: cognitive and sociocultural consequences. In P. A. Hornby (Ed.), *Bilingualism: Psychological, social and educational implications* (pp. 15–27). New York: Academic Press.

Lambert, W. E. (1981). Bilingualism and language acquisition. *Annals, New York Academy of Sciences*, 9–22. doi:10.1111/j.1749-6632.1981.tb41993.x

Lapkin, S., Swain, M., & Psyllakis, P. (2010). The role of languaging in creating zones of proximal development (ZPDs): A long-term care resident interacts with a researcher. *Canadian Journal on Aging*, *29*(4), 477–490. doi:10.1017/S0714980810000644

Lappin-Fortin, K. (2014). Comparing written competency in core French and French immersion graduates. *Canadian Journal of Applied Linguistics*, *17*(2), 91–112.

Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, *1*, 30–41. doi:10.2174/1874913500801010030

Lasagabaster, D. (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in Language and Teaching*, *5*(1), 3–18. doi:10.1080/17501229.2010.519030

Lasagabaster, D. (2016). The relationship between motivation, gender, L1 and possible selves in English-medium instruction. *International Journal of Multilingualism*, *13*(3), 315–332. doi:10.1080/14790718.2015.1105806

Lasagabaster, D., & Sierra, J. M. (2010). Immersion and CLIL in English: More differences than similarities. *ELT Journal*, *64*(4), 367–375. doi:10.1093/elt/ccp082

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Bristol, UK: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). London: Macmillan.

Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal*, *25*, 21–33. doi/10.1177/003368829402500202

Laufer, B. (1996). The lexical threshold of L2 reading: where it is and how it relates to L1 reading ability. In K. Sajaara & C. Fairweather (Eds.), *Approaches to second language acquisition* (pp. 55–62). Jyväskylä: University of Jyväskylä.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, *19*(2), 255–271. doi:10.1093/applin/19.2.255

Laufer, B. (2001). Reading, word-focused activities and incidental vocabulary acquisition in a second language. *Prospect*, *16*(3), 44–54. Retrieved from http://www.ameprc.mq.edu.au/docs/prospect_journal/volume_16_no_3/Prospect_16,3_Article_4.pdf

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size strength and computer adaptiveness. *Language Learning*, *53*(3), 399–436. doi:10.1111/j.0023-8333.2004.00260.x

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26. doi:10.1093/applin/22.1.1

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–321. doi:10.1093/applin/16.3.307

Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. doi:10.1177/026553229901600103

Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA Yearbook*, *1*, 7–28. doi:10.1075/eurosla.1.05lau

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30. Retrieved from https://files.eric.ed.gov/fulltext/EJ887873.pdf

Law, J., Rush, R., Parsons, S., & Schoon, I. (2013). The relationship between gender, receptive vocabulary, and literacy from school entry through to adulthood. *International Journal of Speech-Language Pathology*, *15*(4), 407–415. doi:10.3109/17549507.2012.721897

Lazaruk, W. A. (2007). Linguistic, academic, and cognitive benefits of French immersion. *The Canadian Modern Language Review*, *63*(5), 605–628. doi:10.3138/cmlr.63.5.605

Lee, S. (2007). Vocabulary learning strategies of Korean university students: Strategy use, vocabulary size, and gender. *English Teaching*, *62*(1), 149–168. Retrieved from http://kate.bada.cc/wp-content/uploads/2015/02/kate_62_1_7.pdf

Lee, S. H., & Muncie, J. (2006). From receptive to productive: Improving ESL learners' use of vocabulary in a post-reading composition task. *TESOL Quarter*ly, 40(2), 295–320. doi:10.2307/40264524

Lervåg, A., & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry*, 51(5), 612–620. doi:10.1111/j.1469-7610.2009.02185.x

Lin, J., & Wu, F. (2003). *Differential performance by gender in foreign language testing* (Annual Meeting of the National Center of Measurement and Evaluation) (pp. 1–18). Chicago: The Centre for Research in Applied Measurement and Evaluation.

Lindholm-Leary, K. (2005). *Review of research and best practices on effective features of dual language education programs*. Washington, DC: Center for Applied Linguistics and the National Clearinghouse for English Language Acquisition.

Lindholm-Leary, K. (2012). Success and challenges in dual language education, theory into practice. *Theory into Practice*, *51*(4), 256–262. doi:10.1080/00405841.2012.726053

Lindholm-Leary, K., & Block, N. (2010). Achievement in predominantly low SES/Hispanic dual language schools. *International Journal of Bilingual Education and Bilingualism*, *13*(1), 43–60. doi:10.1080/13670050902777546

Lindholm-Leary, K., & Genesee, F. (2014). Student outcomes in one-way, two-way, and indigenous language immersion education. *Journal of Immersion and Content-Based Language Education*, *2*(2), 165–180. doi:10.1075/jicb.2.2.01lin

López Mezquita, M. (2005). La evaluación de la competencia léxica: tests de vocabulario. Su fiabilidad y validez. Unpublished Doctoral Dissertation, University of Granada, Spain.

Lorenzo, F., Casal, S., & Moore, P. (2009). The effects of content and language integrated learning in European education: Key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics*, *31*(3), 418–442. doi:10.1093/applin/amp041

Lorenzo, F., & Moore, P. (2010). On the natural emergence of language structures in CLIL: Towards a theory of European educational bilingualism. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 23–38). Amsterdam: John Benjamins. doi:10.1075/aals.7.02loe

Lynn, R., Ferguson, D. M., & Horwood, L. J. (2005). Sex differences on the WISC-R in New Zealand. *Personality and Individual Differences*, *39*, 103–114. doi:10.1016/j.paid.2004.12.009

Lyster, R., & Tedick, D. J. (2014). Research perspectives on immersion pedagogy: Looking back and looking forward. *Journal of Immersion and Content-Based Language Education*, *2*(2), 210–224. doi:10.1075/jicb.2.2.04lys

Macedonia, M. (2015). Learning styles and vocabulary acquisition in second language: How the brain learns. *Frontiers in Psychology*, *6*, 1–4. doi:10.3389/fpsyg.2015.01800

Mackey, W. F. (2005). The description of bilingualism. In L. Wei (Ed.), *The bilingualism reader* (pp. 22–52). London: Routledge, Taylor & Francis e-Library.

Madrigal, M., & Vargas, E. (2016). Índice de la riqueza léxica en producciones escritas por estudiantes universitarios. *Káñina, Revista de Artes y Letras*, *XL*, 139–147. doi:10.15517/rk.v40i3.29260

Mady, C. (2015). Immigrants outperform Canadian-born groups in French immersion: Examining factors that influence their achievement. *International Journal of Multilingualism*, *12*(3), 298–311. doi:10.1080/14790718.2014.967252

Maillat, D. (2010). The pragmatics of L2 in CLIL. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 29–58). Amsterdam: John Benjamins. doi:10.1075/aals.7.03mai

Makropoulos, J. (2009). Gaining access to late French-immersion programs: Class-based perspectives of Canadian students in an Ottawa high school. *Bilingual Research Journal*, *32*(3), 317–330. doi:10.1080/15235880903378941

Makropoulos, J. (2010). Students' attitudes to the secondary French immersion curriculum in a Canadian context. *Language, Culture and Curriculum*, *23*(1), 1–13. doi:10.1080/07908310903494525

Marian, V., Shook, A., & Schroeder, S. R. (2013). Bilingual two-way immersion programs benefit academic achievement. *Bilingual Research Journal*, *36*, 167–186. doi:10.1080/15235882.2013.818075

Marín Arroyo, E. (2012). Cronología de la enseñanza del inglés en Costa Rica durante el siglo XX. *Revista Comunicación*, *21*(1), 4–15.

Marsh, D. (2012). *Content and Language Integrated Learning (CLIL): A development trajectory*. Córdoba: Servicio de Publicaciones de la Universidad de Córdoba.

Marsh, D., Maljers, A., & Hartiala, A.-K. (2001). *Profiling European CLIL classrooms: Languages open doors*. The Netherlands, Finland: University of Jyväskylä, Finland & European Platform for Dutch Education.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*(1), 81–93. doi:10.1017/S1366728907003227

Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Philadelphia: John Benjamins.

Meara, P. (2010). *EFL Vocabulary Tests* (2nd Edition). Swansea: Lognostics.

Meara, P. and Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2), 142-151. https://doi.org/10.1177/026553228700400202

Meara, P. and Jones, G. (1990). *Eurocentres Vocabulary Size Test*. 10KA. Zurich: Eurocentres.

Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL: Content and language integrated learning in bilingual and multilingual education*. Oxford: McMillan Education.

Meyer, O. (2010). Towards quality-CLIL: Successful planning and teaching strategies. *Pulso*, *33*, 11–29.

Ministerio de Educación Pública. (2013). Lineamientos sobre horarios para las diferentes ramas, niveles y ciclos del sistema educativo costarricense. Ministerio de Educación Pública, Costa Rica. Retrieved from http://www.drea.co.cr/sites/default/files/Contenido/lineamientos-de-horarios-2016%20-version-final.pdf

Ministerio de Educación Pública. (2016). Programas de estudio de inglés: Tercer ciclo de la Educación General Básica y Educación Diversificada. Ministerio de Educación Pública, República de Costa Rica.

Ministerio de Educación Pública. (2017). Programa de estudio de inglés tercer ciclo: Liceos Experimentales Bilingües y Secciones Bilingües. Ministerio de Educación Pública, República de Costa Rica.

Moreno Espinoza, S. (2010a). Boys' and girls' L2 word associations. In R. M. Jiménez Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 139–166). London: Palgrave Macmillan. doi:10.1111/j.1473-4192.2011.00286.x

Moreno Espinoza, S. (2010b). *Evaluación del vocabulario en redacciones escritas por aprendices del inglés como L2 a través de tecnología*. Unpublished PhD Dissertation. Universidad de la Rioja.

Murillo Rojas, M. (2009). Diversidad del vocabulario en los preescolares: Aportes para valorar su competencia lingüística. *Filología y Lingüística*, *XXXV*(1), 123–138. Retrieved from http://www.kerwa.ucr.ac.cr/bitstream/handle/10669/14415/1271-1903-1-SM.pdf?sequence=1

Myers-Scotton, C. (2006). *Multiple voices: An introduction to bilingualism*. Oxford: Blackwell.

Nakata, T., & Webb, S. (2016). Does studying vocabulary in smaller sets increase learning? The effects of part and whole learning on second language vocabulary acquisition. *Studies in Second Language Acquisition*, *38*, 523–552. doi:10.1017/S0272263115000236

Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, *5*, 12–25. Retrieved from https://www.victoria.ac.nz/lals/about/staff/paul-nation

Nation, P. (1990). *Teaching and learning vocabulary*. New York: Heinle & Heinle.

Nation, P. (1993). Vocabulary size, growth, and use. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (Vol. 6, pp. 115–135). Amsterdam: John Benjamins.

Nation, P. (1996). The four strands of a language course. *TESOL in Context*, *6*(2), 7–12.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, P. (2003). Vocabulary. In D. Nunan (Ed.), *Practical English language teaching* (pp. 129–152). New York: McGraw Hill.

Nation, P. (2006a). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*(1), 59–82. doi:10.3138/cmlr.63.1.59

Nation, P. (2006b). Language education-vocabulary. In K. Brown (Ed.), *Encyclopaedia* of *Language and Linguistics* (2nd ed., Vol. 6, pp. 494–499). Oxford: Elsevier.

Nation, P. (2006c). Vocabulary: Second language. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., Vol. 13, pp. 448–454). Oxford: Elsevier.

Nation, P. (2007a). The four strands. *Innovation in language learning and teaching*, *1*(1), 1–12. doi:10.2167/illt039.0

Nation, P. (2007b). Vocabulary learning through experience tasks. *Indian Journal of Applied Linguistics*, *33*(2), 33–43. doi:10.17507/tpls.0503.13

Nation, P. (2011). Research into practice: Vocabulary. *Language Teaching*, *44*(4), 529–539. doi:10.1017/S0261444811000267

Nation, P. (2013). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, *26*(2), 1–16. Retrieved from http://nflrc.hawaii.edu/rfl/October2014/articles/nation.pdf

Nation, P. and Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.

Nation, P., & Chung, T. (2009). Teaching and testing vocabulary. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 543–559). Oxford: Blackwell.

Nation, P., & Hwang, K. (1995). Where would general service vocabulary stop and special purpose vocabulary begin? *System*, *23*(1), 35–41. doi:10.1016/0346-251X(94)00050-G

Nation, P., & Meara, P. (2010). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics* (2nd ed., pp. 34–52). London: Hodder Education.

Nation, P., & Yamamoto, N. (2012). Applying the four strands to language learning. *International Journal of Innovation in English Language Teaching*, *1*(2), 167–181. Retrieved from https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/yamamoto-four-strands.pdf.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp. 6–19). Cambridge: Cambridge University Press.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle Cengage Learning.

Navarro Pablo, M., & García Jiménez, E. (2018). Are CLIL students more motivated? An analysis of affective factors and their relation to language attainment. *Porta Linguarum*, *29*, 71–90. Retrieved from http://www.ugr.es/~portalin/articulos/PLnumero29/4_MACARENA%20NAVARRO.pdf

Nicolay, A. C., & Poncelet, M. (2013). Cognitive abilities underlying second-language vocabulary acquisition in an early second-language immersion education context: A longitudinal study. *Journal of Experimental Child Psychology*, *115*, 655–671. doi:10.1016/j.jecp.2013.04.002

Nikula, T. (2010). Effects of CLIL on a teacher's classroom language use. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 105–123). Amsterdam: John Benjamins.

Ojeda Alba, J., & Jiménez Catalán, R. M. (2007). The worlds children's words build. *Didáctica Lengua y Literatura*, *19*, 155–172. Retrieved from http://revistas.ucm.es/edu/11300531/articulos/DIDA0707110155A.PDF

Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Read Writ*, *26*, 45–65. doi:10.1007/s11145-012-9392-5

Olmos, C. (2009). An assessment of the vocabulary knowledge of students in the final year of secondary education. Is their vocabulary extensive enough? *International Journal of English Studies*, *9*(3), 73–90.

Özönder, O. (2016). Student EFL teachers' receptive vocabulary size. *Procedia: Social and Behavioral Sciences*, *232*, 444–450. doi:10.1016/j.sbspro.2016.10.061

Paciotto, C., & Delany-Barmann, G. (2012). Minority language education reform from the bottom: two-way immersion education for new immigrant populations in the United States. *Procedia: Social and Behavioral Sciences*, *46*, 17–21. doi:10.1016/j.sbspro.2012.05.059

Palmer, D. (2010). Race, power, and equity in a multiethnic urban elementary school with a dual-language "strand" program. *Anthropology & Education Quarterly*, *41*(1), 94–114. doi:10.1111/j.1548-1492.2010.01069.x

Palmer, D. K., Martínez, M. A., Mateus, S. G., & Henderson, K. (2014). Reframing the debate on language separation: Toward a vision for translanguaging pedagogies in the dual language classroom. *The Modern Language Journal*, *98*(3), 757–772. doi:10.1111/modl.12121

Paribakht, T. S., & Wesche, M. (1979). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174–200). Cambridge: Cambridge University Press.

Parkes, J. (2008). Who chooses dual language education for their children and why. *International Journal of Bilingual Education and Bilingualism*, *11*(6), 635–660. doi:10.1080/13670050802149267

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye tracking study. *Studies in Second Language Acquisition*, *38*, 97–130. doi:10.1017/S0272263115000224

Pérez-Cañado, M. L. (2016). From the CLIL craze to the CLIL conundrum: Addressing the current CLIL controversy. *Bellaterra Journal of Teaching & Learning Language & Literature*, *9*(1), 9–31. doi:10.5565/rev/jtl3.667

Pérez-Vidal, C. (2015). Languages for all in education: CLIL and ICLHE at the crossroads of multilingualism, mobility and internationalization. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual environments* (pp. 31–50). New York: Springer International. doi:10.1007/978-3-319-11496-5_3

Pfenninger, S. (2016). All good things come in threes: Early English learning, CLIL and motivation in Switzerland. *Cahiers de I'ILSL*, *48*, 119–147.

Pichette, F., De Serres, L., & Lafontaine, M. (2012). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics*, *33*(1), 66–82. doi:10.1093/applin/amr037

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*(1), 1–28.

Polat, N., & Cepik, S. (2016). An exploratory factor analysis of the sheltered instruction observation protocol as an evaluation tool to measure teaching effectiveness. *TESOL Quarterly*, *50*(4), 817–843. doi:10.1002/tesq.248

Prieto-Arranz, J. I., Rallo-Fabra, L., Calafat-Ripoll, C., & Catrain-González, M. (2015). Testing progress of receptive skills in CLIL and non-CLIL contexts. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual environments* (pp. 123–137). New York: Springer International. doi:10.1007/978-3-319-11496-5_8

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*(3), 513–536.

Quesada Pacheco, A. (2013). La enseñanza y aprendizaje del idioma inglés: la investigación y su impacto en la realidad costarricense. *Revista de Lenguas Modernas*, *19*, 393–408.

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 12–25. doi:10.1177/003368828801900202

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41–60). New Jersey: Lawrence Erlbaum.

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaarts & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209–227). Amsterdam: John Benjamins.

Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies*, *7*(2), 105–125. Retrieved from http://revistas.um.es/ijes/issue/view/4811

Reindorp, R. C. (1943). Teaching English by Radio in Costa Rica. *Hispania*, *26*(4), 425–428. https://doi.org/10.2307/333598

Reindorp, R. C. (1944). Teaching English as a foreign language in Costa Rica. *The Modern Language Journal*, *28*(3), 227–232. https://doi.org/10.1111/j.1540-4781.1944.tb03895.x

Reynolds, K. M. (2015). *Approaches to inclusive English classrooms: A teacher's handbook for content-based instruction*. Bristol, UK: Multilingual Matters.

Roy, S. (2008). French immersion studies: From second-language acquisition (SLA) to social issues. *The Alberta Journal of Educational Research*, *54*(4), 396–406.

Roy, S., & Galiev, A. (2011). Discourses on bilingualism in Canadian French immersion programs. *The Canadian Modern Language Review*, *67*(3), 351–376. doi:10.3138/cmlr.67.3.351

Roy, S., & Schafer, P. C. (2015). Reading as a skill or as a social practice in French immersion? *Language and Education*, *29*(6), 527–544. doi:10.1080/09500782.2015.1059435.

Ruiz de Zarobe, Y. (2010). Written production and CLIL: An empirical study. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (Vol. 7, pp. 191–209). Amsterdam: John Benjamins.

Ruiz, N. (2012). It's different with second language learners: Learning from 40 years of experience. In C. Dudley-Marling & S. Michaels (Eds.), *High expectation curricula: Helping all students succeed with powerful learning* (pp. 145–161). New York: Teachers College Press.

Sailor, K. M. (2013). Is vocabulary growth influenced by the relations among words in a language learner's vocabulary? *Journal of Experimental Psychology*, *39*(5), 1657–1662. doi:10.1037/a0032993

Sánchez Avendaño, C. (2005). Los problemas de redacción de los estudiantes costarricenses: Una propuesta de redacción desde la lingüística de texto. *Filología y Lingüística*, *XXXI*(1), 267–295.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524780

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Nagami, … I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721–737). Singapore: National University of Singapore, Centre for Language Studies.

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, *48*(2), 281–317.

Schmitt, N. (2007). Current perspectives on vocabulary teaching and learning. In J. Cummins & C. Davison (Eds.), *International handbook of English language teaching* (pp. 745–759). New York: Springer. doi:10.1007/978-0-387-46301-8_55

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329-363. doi:10.1177/136216880 8089921

Schmitt, N. (2010a). Key issues in teaching and learning vocabulary. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 1–20). Bristol, UK: Multilingual Matters.

Schmitt, N. (2010b). *Researching vocabulary: A vocabulary research manual*. New York: Palgrave Macmillan. doi:10.1057/9780230293977

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913–951. doi:10.1111/lang.12077

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(i), 26–43. doi:10.1111/j.1540-4781.2011.01146.x

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *SSLA*, *20*, 17–36.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484–503. doi:10.1017/S0261444812000018

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88. doi:10.1177/026553220101800103

Schuetze, U. (2015). Spacing techniques in second language vocabulary acquisition: Short-term gains vs. long-term memory. *Language Teaching Research*, *19*(1), 28–42. doi:10.1177/1362168814541726

Skutnabb-Kangas, T., & McCarty, T. L. (2008). Key concepts in bilingual education: Ideological, historical, epistemological, and empirical foundations. In J. Cummins & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 5, pp. 3–17). New York: Springer.

Song, K. H. (2016). Systematic professional development training and its impact on teachers' attitudes toward ELLs: SIOP and guided coaching. *TESOL Journal*, *7*(4), 767–799. doi:10.1002/tesj.240

Song, X., Cheng, L., & Klinger, D. (2015). DIF investigations across groups of gender and academic background in a large-scale high-stakes language test. *Papers in Language Testing and Assessment*, *4*(1), 97–124. Retrieved from http://www.altaanz.org/uploads/5/9/0/8/5908292/song_et_al.pdf

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, *36*(2), 139–152. doi:10.1080/095717308023 89975

Sunderland, J. (2000). Issues of language and gender in second and foreign language education. *Language Teaching*, *33*, 203–233. doi:10.1017/S02614448000 15688

Sunderland, J. (2010). Theorizing gender perspectives in foreign and second language learning. In R. M. Jiménez Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 1–19). London: Palgrave Macmillan. doi:10.1111/j.1473-4192.2011.00286.x

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125–144). Oxford: Oxford University Press.

Swain, M. (2000a). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford: Oxford University Press.

Swain, M. (2000b). French immersion research in Canada: Recent contributions to SLA and applied linguistics. *Annual Review of Applied Linguistics*, *20*, 199–212. doi:10.1017/s0267190500200123

Swain M. (2006). Languaging, agency and collaboration in advanced second language proficiency. In Byrnes H. (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky*, 95-108. London: Continuum.

Swain, M. (2013). The inseparability of cognition and emotion in second language learning. *Language Teaching*, *46*(2), 195–207. doi:10.1017/S02614448 11000486

Swain, M., & Deters, P. (2007). "New" Mainstream SLA Theory: Expanded and Enriched. *The Modern Language Journal*, *91*(Focus Issue), 820–836. doi:10.1111/j.1540-4781.2007.00671.x

Swain, M., & Johnson, R. K. (1997). Immersion education: A category within bilingual education. In R. K. Johnson & M. Swain (Eds.). *Immersion education: International perspectives* (pp. 1–16). Cambridge, MA: Cambridge University Press. doi:10.1017/cbo9781139524667.003

Swain, M., & Lapkin, S. (1989). Canadian immersion and adult second language teaching: What's the connection? *The Modern Language Journal*, *73*(ii), 151–159. doi:10.1111/j.1540-4781.1989.tb02537.x

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards Second Language Learning. *Applied Linguistics*, *16*(3), 371–391. doi:10.1093/applin/16.3.371

Swain, M., & Lapkin, S. (2005). The evolving sociopolitical context of immersion education in Canada: Some implications for program development. *International Journal of Applied Linguistics*, *15*(2), 169–186. doi:10.1111/j.1473-4192.2005.00086.x

Swain, M., & Suzuki, W. (2008). Interaction, output, and communicative language learning. In B. Spolsky & M. H. Francis (Eds.), *The handbook of educational linguistics* (pp. 557–570). Oxford: Blackwell. doi: https://doi.org/10.1002/9780470694138.ch39

Swain, M., & Watanabe, Y. (2013). Languaging: Collaborative dialogue as a source of second language learning. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Oxford: Blackwell. doi:10.1002/9781405198431.wbeal0664

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, *84*, 265–278. doi:10.1016/j.neuroimage.2013.08.023

Tedick, D. J. (2014). Language immersion education. A research agenda for 2015 and beyond. *Journal of Immersion and Content-Based Language Education*, *2*(2), 155–164. doi:10.1075/jicb.2.2.00int

Terrazas Gallegos, M., & Agustín Llach, M. P. (2009). Exploring the increase of receptive vocabulary knowledge in the foreign language: A longitudinal study. *International Journal of English Studies*, *9*(1), 113–133.

Torres-Guzmán, M. E., Kleyn, T., Morales-Rodríguez, S., & Han, A. (2005). Self-designated dual-language programs: Is there a gap between labeling and implementation? *Bilingual Research Journal*, *29*(2), 453–474. doi:10.1080/15235882.2005.10162844

Ullman, M. T., Miranda, R. A., & Travers, M. L. (2008). Sex differences in the neurocognition of language. In J. B. Becker, K. J. Berkley, N. Geary, E. Hampson, J. P. Herman, & E. Young (Eds.), *Sex on the brain: From genes to behaviour* (pp. 291–309). New York: Oxford University Press. doi:10.1093/acprof:oso/9780195311587.003.0015

Valdés, G. (2013). Dual-language immersion programs: A cautionary note concerning the education of language minority students. In C. A. Klee, A. Lynch, & E. Tarone (Eds.). *Research and practice in immersion education: Looking back and looking ahead* (pp. 15–38). Minneapolis, MN: Board of Regents of the University of Minnesota. doi:10.17763/haer.67.3.n5q175qp86120948

Vagh, S. B., Pan, B. A., & Mancilla-Martinez, J. (2009). Measuring growth in bilingual and monolingual children's English productive vocabulary development: The utility of combining parent and teacher report. *Child Development*, *80*(5), 1545–1563. doi:10.1111/j.1467-8624.2009.01350.x

van der Slik, F. W. P., van Hout, R. W. N. M., & Schepens, J. J. (2015). The gender gap in Second Language Acquisition: gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PLoS ONE*, *10*(11), 1–22. doi:10.1371/journal.pone.0142056

Velasco, P., & García, O. (2014). Translanguaging and the writing of bilingual learners. *Bilingual Research Journal*, *37*, 6–23. doi:10.1080/15235882.2014.893270

Vitevitch, M. S., & Goldstein, R. (2014). Keywords in the mental lexicon. *Journal of Memory and Language*, *73*, 131–147. doi:10.1016/j.jml.2014.03.005

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Wei, L. (2005). Dimensions of bilingualism. In L. Wei (Ed.), *The bilingualism reader*. London: Routledge.

Wei, L. (2008). Research perspectives on bilingualism and multilingualism. In L. Wei & M. G. Moyer (Eds.), *The Blackwell guide to research methods in bilingualism and multilingualism*. Oxford: Blackwell.

Wojtinnek, P. R., & Pulman, S. (2011). Semantic relatedness from automatically generated semantic networks. In *Proceedings of the Ninth International Conference on Computational Semantics* (pp. 390–394). Association for Computational Linguistics. Retrieved from https://dl.acm.org/citation.cfm?id= 2002669.2002718

APPENDICES

APPENDIX 1

VNiVERSiDAD D SALAMANCA

Investigadora responsable: Damaris Castro García
Universidad de Salamanca
Tel: (506) 8837 5360/(+34) 637 672 276
E-mail: dcastrog@usal.es

Estudiante: _____ **Colegio**: _____

Grupo: _____ **Tiempo**: 20 minutos

1. Prueba de vocabulario receptivo: En la siguiente prueba usted debe escoger la palabra correcta (de la columna izquierda) que vaya con el significado correspondiente en las opciones de la derecha. Escriba el número de la palabra junto a su significado.

Vea el siguiente ejemplo:

1. business	
2. clock	_____ part of the house
3. horse	_____ animal with four legs
4. pencil	_____ something used for writing
5. shoe	
6. wall	

Usted debe responder de la siguiente forma:

1. business	
2. clock	6 part of the house
3. horse	3 animal with four legs
4. pencil	4 something used for writing
5. shoe	
6. wall	

Algunas palabras están en la prueba solo para incrementar su dificultad. Usted no debe encontrar el significado para estas palabras. En el ejemplo arriba, estas palabras son *business, clock* y *shoe*.

Si usted no tiene una idea sobre el significado de una palabra, no trate de adivinar. Pero, si usted piensa que podría saber el significado, entonces puede intentar encontrar la respuesta.

The Vocabulary Levels Test: Version 2, Level 2000 (Schmitt et al., 2001)

1. copy _____ end or highest 2. event point 3. motor _____this moves a car 4. pity _____ things made to 5. profit be like another 6. tip	1. accident _____ loud deep 2. debt sound 3. fortune _____ something you 4. pride must pay 5. roar _____ having a high 6. thread opinion of yourself
1. coffee 2. disease _____ money for work 3. justice _____ a piece of 4. skirt clothing 5. stage _____ using the law in 6. wage the right way	1. arrange 2. develop _____ grow 3. lean _____ put in order 4. owe _____ like more 5. prefer than something else 6. seize
1. clerk 2. frame _____ a drink 3. noise _____ office worker 4. respect _____ unwanted 5. theater sound 6. wine	1. blame 2. elect _____ make 3. jump _____ choose by 4. threaten voting 5. melt _____ become like 6. manufacture water
1. dozen 2. empire _____ chance 3. gift _____ twelve 4. tax _____ money paid 5. relief to the government 6. opportunity	1. ancient 2. curious _____ not easy 3. difficult _____ very old 4. entire _____ related to God 5. holy 6. social
1. admire _____ make wider or 2. complain longer 3. fix _____ bring in for 4. hire the first time 5. introduce _____ have a high 6. stretch opinion of someone	1. slight 2. bitter _____ beautiful 3. lovely _____ small 4. merry _____ liked by many 5. popular people 6. independent

APPENDIX 2

VNiVERSiDAD D SALAMANCA

Investigadora responsable: Damaris Castro García
Universidad de Salamanca
Tel: (506) 8837 5360/(+34) 637 672 276
E-mail: dcastrog@usal.es

Reading Test: Receptive Vocabulary in Context (Cambridge, *ESOL SKILLS for LIFE*, 2011)

Estudiante: _____ Grupo: _____

Colegio: _____Time: 1:15 minutes

Each task tells you how much to write, how much time to spend, how many points you can get.

Part 1 (About 20 minutes)
Zohra sends an email to her friend James. Look at Text A.

TEXT A.

james@mymail.com

Hi!!!

Paragraph 1
Good news!! – I got the job in the chemist! I'm delighted (I can't stop smiling) because it's only the second interview I've ever had. Only a few more terrible shifts in the supermarket! I might have a party to celebrate –I'll call you when I've organized it.
Paragraph 2
Thanks for helping me find the new class. My teacher thinks that I really need to work hard on reading next term as it's my weakest area. I love books (I've got loads – including a great new dictionary) but I'm not sure about what I can do to improve. I'm lucky because my kids support me with this and sometimes translate things for me.
Paragraph 3
I know you always have your head in a book. Have you got any reading tips for me?
Paragraph 4
Let me know what you think on Saturday. Let's meet in that new place – the burgers are great. It's opposite the park, about two minute's walk from the station.

Take care,
Zohra

Question 1: Why is Zohra writing to James? Check (✓) one box. (1 point)

She wants to find a new class.	
She wants James to plan a party.	
She wants advice about studying.	

Question 2: Where does Zohra suggest meeting James? Check (✓) one box. (1 point)

at the station	
in a café	
in the park	

Question 3: Look at these sentences. Are they true or false? Check (✓) TRUE or FALSE for each sentence. (4 points)

	TRUE	FALSE
A. Zohra has been to many job interviews before.		
B. Zohra doesn't like reading.		
C. Zohra's family help her study.		
D. Zohra says she needs to buy some new books.		

Question 4: Where does Zohra work at the moment?

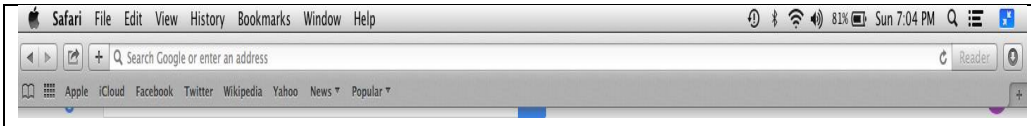
_____ (1 point)

Question 5: Look at Paragraph 1 of Text A. Which word means ***very happy***?

Write ONE word: _____ (1 point)

James finds a website and sends the link to Zohra. Look at text B.

TEXT B



Reading is really important for English learners as it helps other language areas.

Paragraph 2

Try to read lots of different things. Pick up a newspaper and read about something that interests you. It's a good idea to watch the news on TV and then read about the same thing in a newspaper –this can help you understand the general meaning. Borrow one of the latest novels from the library – something that really grabs your interest.

Paragraph 3

New/difficult words –some strategies:

- Avoid looking up all the new words in the text you are reading. This will save you time searching through the dictionary.
- Does the word look like other words you know? It might have a similar meaning.

Paragraph 4

Try to increase your reading speed. See how long it takes you to read 100 words. Note the time and then read the same number more quickly the following week. You can improve the time it takes you to read by following this technique –it really works.

Do	Don't
-read daily	- translate every word
-read for pleasure	- read too slowly

Question 6: What is the best title for Text B? Check (✓) ONE box. (1 point)

The best books to read	☐
Developing reading skills	☐
Where to find reading courses	☐

Question 7: How many different types of books does Text B mention? Check (✓) ONE box. (1 point)

2	☐
3	☐
4	☐

Question 8: Look at Paragraph 3 of Text B. What does This mean? Check (✓) ONE box. (1 point)

reading new words	☐
developing reading skills	☐
where to find reading courses	☐

Question 9: How often does Text B say you should read?
(1 point)

_____

Question 10: There is a spelling mistake in Paragraph 2 of Text B. Which word is not correct? Write the word here.
(1 point)

_____

Question 11: Look at Paragraph 4 of Text B. Which word means ***way of doing something***? Write ONE word.
(1 point)

_____

Part 2 **(About 25 minutes)**

James finds an article and sends it to Zohra. Look at Text C.

TEXT C

Have you visited your local shopping center recently and noticed a new retailer amongst the shops there? Dave Pringle reports on the new book superstores opening around the U.S –over 35 in the last 12 months. He asks why this is happening. What is attracting people to these kinds of shops?

Ozgur Durson, a regular shopper in his local book superstore, said "It's not really the size of these stores that is important to me, or the latest books –it's the wonderful café where I can relax and have a look at my favorite magazines."

However, shopping in these new book superstores isn't a pleasure for everyone. It can actually be the opposite. Jasna Khan's views are typical of many. "I hate the huge bookshops. *Books for You* has just opened in my high street and in my opinion it's unfriendly and far too noisy. In contrast, I much prefer smaller shops because there is always someone to assist you. The big ones sell all sorts of CDs and DVDs and they always play loud music throughout the store, which in all honesty gives me a headache!"

Question 12: What is the best title for Text C? Check (✓) ONE box. (1 point)

People are buying more books	☐
Giant bookshops opening everywhere	☐
Reading books is good for your mind	☐

Question 13: What does Ozgur Durson like most about the book superstores? Check (✓) ONE box. (1 point)

The sell the latest books.	☐
They have a café.	☐
They are big.	☐

Question 14: Why does Jasna Khan say she prefers smaller bookshops? Check (✓) ONE box. (1 point)

You can buy DVDs.	☐
They are cheaper.	☐
You can get help.	☐

Question 15: Look at Line 10 of Text C? What does IT mean? Check (✓) ONE box. (1 point)

shopping	
a pleasure	
the opposite	

Question 16: What information does Text C tell you? Check (✓) YES or NO for each. (4 points)

	YES	NO
A. what time of day people go to book superstores		
B. what the name of a book superstore is		
C. what the most popular books are		
D. how many book superstores have opened in the last year		

Here is a newspaper article about libraries. Look at Text D.

TEXT D

Soon we will stop using the word 'library'. Not because we don't love books anymore but because we will call new libraries 'idea stores' and they will have cafés and child care centers for children.

A new library in Bow, East London, has brought color and light to one part of the capital. When you walk through the modern glass doors, you don't see shelves of books or busy librarians. Instead there is a café and a helpful security guard answering questions.

Technology has changed things too. Now there are computers for the internet and an area for DVDs and CDs.

"The majority of people have been very positive since we opened and only a few have said they prefer the old-style library," says manager Zoinul Abidin.

Paragraph 5

Rabia Puddicombe, a local journalist, says, "The librarians seem much friendlier than in the past – they're always happy to chat. And wherever I'm looking after my granddaughter, I bring her here to play with the other children. The child care worker supervises her while I sit with a cup of tea."

Question 17: What is the best title for Text D? Check (✓) ONE box. (1 point)

Learning in libraries	☐
Working for a library	☐
Libraries of the future	☐

Question 18: Why will libraries be called something different? Check (✓) ONE box. (1 point)

because people are not interested in books	☐
because they will offer many new services	☐
because more children will go to libraries	☐

Question 19: What does Zoinul Abidin say about library visitors? Check (✓) ONE box. (1 point)

Many people preferred the old library.	☐
The new library is popular with most people.	☐
She hopes more people will come.	☐

Question 20: How many different jobs can you see in Paragraph 5 of Text D? Check (✓) ONE box. (1 point)

1	☐
2	☐
3	☐

Question 21: Here is a list of bookshops: You want to call the bookshops. You look in the phone book for the telephone numbers. The phone book is in alphabetical order (A-Z) (4 points)

List of bookshops:

University Bookshop
Steve's Bookshop
Mary's Bookshop
Uptown Bookshop
Smith's Bookshop

A: Which bookshop is FIRST in the phone book?
_____

B: You find "Smith's Bookshop." Which bookshop is NEXT in the phonebook?
_____

C: Which bookshop is LAST in the phonebook?
_____

Part 3 **(About 30 minutes)**

Look at Text E. Three students write articles for their college magazine.

TEXT E

Ali Some of my friends never pick up a newspaper but I think they're great. I always turn to the back pages because I like to check the scores and see who has won. My colleague and I buy the paper before getting the bus to work and then look at it together, chatting about the interesting articles. Paragraph 2 I think reading newspapers has greatly improved my English so it's very important for me. My teachers are always telling students that they should pick up the free ones and read an article a day. I always ignore words I don't know instead of looking up all the words in the dictionary – I think that saves times. Unlike most of my friends. I much prefer reading to listening to the radio or watching TV.	
Fabiano Recently I've become interested in reading about famous people and celebrities in different countries. Some people think that there are too many stories about them and not enough serious new stories. However, I'm crazy about music so I always like to know what my favorites singers are doing and what the latest gossip is. After I finish reading, my sister always borrows my papers and magazines –she is interested in the sports pages. It's quite frustrating because she never buys her own! Reading the newspaper has helped me develop my language skills. I always push myself to read as quickly as possible and I look up all the new words in the dictionary. Then I like to talk to my wife about the stories which interest me most.	
Ding Paragraph 1 I mainly look at newspapers for the politics because it is essential for us to know what's going on. My library has a vast selection of papers (lots of local, national and international ones) so I don't have to pay for them. I usually go there after work. They also lend books and CDs and have a good range of DVDs. The only problem is you can't take the papers home. Paragraph 2 Reading newspapers has definitely helped me develop my English. I write down words I don't know and then ask my teacher about the meaning when I go to class. She's always pleased that I'm reading independently in my own time. Sometimes she recommends interesting news stories for us to read. Paragraph 3 I think it's really interesting how UK news stories are so unlike articles in papers in china. It's not just the content, but the writing styles aren't similar either. In fact, the layout is the only thing that's the same.	

Question 22: Why are all the students writing? Check (✓) one box. (1 point)

to recommend their favorite newspapers	☐
to compare books and newspapers	☐
to explain why they read newspapers	☐

Question 23: Who says he lends the newspapers to someone? Check (✓) one box. (1 point)

Ali	☐
Fabiano	☐
Ding	☐

Question 24: Who says he does **NOT** buy newspapers? Check (✓) one box. (1 point)

Ali	☐
Fabiano	☐
Ding	☐

Question 25: Who says he reads the newspapers with someone else? Check (✓) one box. (1 point)

Ali	☐
Fabiano	☐
Ding	☐

Question 26: Look at Ali's text. What does <u>ones</u> mean? Check (✓) one box. (1 point)

newspapers	☐
teachers	☐
students	☐

Question 27: Look at Paragraph 3 of Ding's text. What point is he making? Check (✓) one box. (1 point)

Chinese and the UK news articles are almost the same.	
Chinese and the UK news articles have many differences.	
Chinese news articles are better than UK ones.	

Question 28: What information does Text E tell you? Check (✓) Yes or No for each. (3 points)

	YES	NO
A. which sport Ali likes reading about		
B. why Fabiano likes reading about celebrities		
C. what kind of news stories Ding likes reading		

Question 29: What do the students say they do when they find difficult words in a newspaper? Write **ONE** thing for each person. (3 points)

Ali: _____

Fabiano: _____

Ding: _____

Question 30: Look at Paragraph 1 of Ding's text. Which word means *very big?* Write ONE word. (1 point)

_____

Question 31: There is a spelling mistake in Paragraph 2 of Ali's text. Which word is not correct? Write the word here. (1 point)

_____

Question 32: There is a mistake in Paragraph 3 of Ding's text. Which word needs a capital letter? (1 point)

_____

APPENDIX 3

VNiVERSiDAD D SALAMANCA

Investigadora responsable: Damaris Castro García
Universidad de Salamanca
Tel: (506) 8837 5360/(+34) 637 672 276
E-mail: dcastrog@usal.es

Estudiante: _____

Colegio: _____ Grupo: _____

Prueba de vocabulario productivo (Laufer& Nation, 1999)
Complete las palabras que están subrayadas. Cuentan con un ejemplo completado para que vean lo que deben hacer. Ejemplo:

Oración dada:	**Su respuesta:**
He was riding a bic_____.	He was riding a bicycle.

Productive Vocabulary Test Level 2000

1. I'm glad we had this opp_____ to talk.
2. There are a doz_____ eggs in the basket.
3. Every working person must pay income t_____.
4. The pirate buried the trea_____in a desert island.
5. Her beauty and cha_____ had a powerful effect on men.
6. La_____ of rain led to a shortage of water in the city.
7. He takes cr_____ and sugar in his coffee.
8. The rich man died and left all his we_____ to his son.
9. Pup_____ must hand in their papers by the end of the week.
10. This sweater is too tight. It needs to be stret_____.
11. Ann intro_____ her boyfriend to her mother.
12. Teenagers often adm_____ and worship pop singers.
13. If you blow up that balloon any more it will bur_____.
14. In order to be accepted into the university, he had to impr_____ his grades.
15. The telegram was deli_____ two hours after it had been sent.
16. The differences were so sl_____ that they went unnoticed.
17. The dress you're wearing is lov_____.
18. He wasn't very popu_____ when he was a teenager, but he has many friends now.

APPENDIX 4

VNiVERSiDAD D SALAMANCA

Investigadora responsable: Damaris Castro García
Universidad de Salamanca
Tel: (506) 8837 5360/(+34) 637 672 276
E-mail: dcastrog@usal.es

Prueba para medir Vocabulario productivo: Composición

Estudiante: _____Colegio: _____
Sección: _____ Tiempo: 1 hour

Escoja uno de los siguientes temas y escriba una composición en inglés donde discuta el tema y presente su opinión al respecto.

Temas:
1. "What could a foreigner find and enjoy if he visited Costa Rica?" Discuss the things and places that Costa Rica has to offer to visitors.
2. "Should a government be allowed to limit the number of children a family can have?" Discuss this idea considering basic human rights and the danger of population explosion. (Laufer & Nation, 1995)
3. "A person cannot be poor and happy, because money is always needed to gain something that is important to that person." Argue for and against this idea. (Laufer & Nation, 1995)

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

APPENDIX 5

VNiVERSiDAD D SALAMANCA

Investigadora responsable: Damaris Castro García
Universidad de Salamanca
Tel: (506) 8837 5360/(+34) 637 672 276
E-mail: dcastrog@usal.es

La participación en estas pruebas es **VOLUNTARIA y CONFIDENCIAL.**

Estas pruebas forman parte de una investigación doctoral que se lleva a cabo en la Universidad de Salamanca, España, como requisito para obtener el Doctorado en Estudios Ingleses Avanzados, Lenguas y Culturas en Contacto. Todos los datos recogidos serán utilizados únicamente con fines de investigación y se asegura la confidencialidad de los participantes.

El propósito de las diferentes pruebas es determinar las características del vocabulario, en inglés, de los estudiantes de quinto año de este centro educativo. Además, se recoge información general sobre los antecedentes de los estudiantes participantes, así como información sobre la motivación en su clase de inglés. Al completar este documento, usted acepta formar parte de esta investigación.

En cada caso las instrucciones aparecen antes de iniciar los cuestionarios o pruebas; lea con atención antes de responder y aclare cualquier duda antes de iniciar.

GRACIAS POR SU COLABORACIÓN.

Más información sobre la investigación:

Investigadora responsable: Damaris Castro García
Universidad de Salamanca; Facultad de Psicología
Departamento de Filología Inglesa
Facultad de Filología
Placentinos, 18,
37008 Salamanca
Tel: (506) 8837 5360 / (34) 637 672 276
E-mail: dcastrog@usal.es

Complete el siguiente cuestionario con su información personal

Nombre completo: _____ Edad: _____

Centro Educativo: _____ Grupo: _____

Lugar de residencia:_____

Sexo: Masculino_____ Femenino_____

¿Cuál es la ocupación de sus padres?

Madre_____ Padre_____

¿Cuál es el nivel más alto de educación de sus padres? Marque una opción.

Madre:	Padre:
___ Ninguna	___ Ninguna
___ Primaria	___ Primaria
___ Secundaria	___ Secundaria
___ Universitaria	___ Universitaria

¿A qué edad empezó usted a estudiar inglés? _____

Marque todos los que se aplican, deje en blanco los que no se aplican. ¿Cuándo usa usted inglés?

☐ a. En clase	☐ c. Con los amigos
☐ b. En los recreos	☐ d. En su casa

¿Ha recibido usted clases privadas de inglés? _____Sí _____No

Si las ha recibido, ¿por cuánto tiempo? _____

Marque todos los que se aplican: Usted tiene acceso a internet:

☐ a. En su casa	☐ c. En su celular
☐ b. En el colegio	☐ d. No tienen acceso

¿Cuáles de las siguientes actividades realiza usualmente en inglés?

☐ a. Ver televisión	☐ c. En su celular
☐ b. Oír música	☐ d. No tienen acceso

Evalúe su nivel actual de inglés. Escoja ***solo una*** de las siguientes opciones dadas.

1 = no entiendo lo que escucho y lo que leo y no puedo hablar o escribir en inglés.
2 = entiendo un poco lo que escucho y lo que leo pero no puedo hablar o escribir en inglés.
3 = entiendo lo que escucho y lo que leo pero no puedo hablar ni escribir.
4 = entiendo lo que escucho y lo que leo y puedo hablar y escribir pero con mucha dificultad.
5 = entiendo lo que escucho y lo que leo y puedo hablar y escribir con un poco de dificultad.
6 = entiendo lo que escucho y lo que leo y puedo hablar y escribir sin dificultad.
7 = entiendo lo que escucho y lo que leo y puedo hablar y escribir en inglés igual que en español.

¿Piensa que es importante aprender inglés? Sí _____ No _____

¿Para cuáles de las siguientes actividades piensa usted que el inglés sería importante?

☐ a. Para visitar otros países
☐ b. Para su futuro trabajo
☐ c. Para su superación personal
☐ d. Para conocer extranjeros

APPENDIX 6

VNiVERSiDAD D SALAMANCA

Investigadora responsable: Damaris Castro García
Universidad de Salamanca
Tel: (506) 8837 5360/(+34) 637 672 276
E-mail: dcastrog@usal.es

Encuesta sobre motivación para el aprendizaje de inglés como segunda lengua (Adaptada de Gardner, 2004)

Estudiante: _____
Colegio: _____ Grupo: _____

Complete el siguiente instrumento pensando en su comportamiento en sus clases de inglés o en las que son enseñadas en inglés. En cada caso por favor asigne una valoración de 1 a 5 marcando el número que corresponde.
Asigne un 5 si la información es muy cierta o un 1 si no es para nada cierta para usted.
Puede usar los valores intermedios si lo considera necesario para su caso personal.

5: Completamente de acuerdo
4. De acuerdo
3. A veces de acuerdo
2. Poco de acuerdo
1. Completamente en desacuerdo

Ejemplo: Estudio mucho aun cuando una clase no me gusta. 5 4 3 2 1

1	No pongo mucha atención a las observaciones/ correcciones que recibo en mis clases de inglés.	5 4 3 2 1
2	Aprender inglés es realmente bueno.	5 4 3 2 1
3	Estudiar inglés es importante porque me permitirá sentirme más cómodo(a) con la gente que habla inglés.	5 4 3 2 1
4	Tengo un gran deseo de saber todos los aspectos del idioma inglés.	5 4 3 2 1
5	Hago un esfuerzo para tratar de entender todo lo que veo o escucho en inglés.	5 4 3 2 1
6	Estudiar inglés es importante porque lo voy a necesitar para mi carrera.	5 4 3 2 1
7	Saber inglés no representa un propósito importante en mi vida.	5 4 3 2 1

8	No me preocupo por revisar mis tareas cuando las recibo corregidas por mi profesor de inglés.	5 4 3 2 1
9	Realmente disfruto al aprender inglés.	5 4 3 2 1
10	Estudiar inglés es importante porque me permitirá conocer y hablar con más y mayor variedad de gente.	5 4 3 2 1
11	Me mantengo al día con el inglés porque trabajo con este idioma casi todos los días.	5 4 3 2 1
12	Estudiar inglés es importante porque me permitirá tener una mejor formación.	5 4 3 2 1
13	Retraso trabajar en mis tareas de inglés tanto como puedo.	5 4 3 2 1
14	Estudiar inglés es importante porque me permitirá entender y apreciar mejor el modo de vida de sus hablantes.	5 4 3 2 1
15	Quiero aprender inglés tan bien que me sea natural hablarlo.	5 4 3 2 1
16	Estudiar inglés es importante porque será útil para obtener un trabajo.	5 4 3 2 1
17	Estoy perdiendo cualquier interés que alguna vez tuve en aprender inglés.	5 4 3 2 1
18	Cuando tengo un problema para entender algo en mi clase de inglés siempre le pregunto a mi profesor(a).	5 4 3 2 1
19	Quisiera entender tanto inglés como sea posible.	5 4 3 2 1
20	Usualmente me rindo y no pongo más atención cuando no entiendo alguna explicación de mi profesor(a) de inglés.	5 4 3 2 1
21	Aprender inglés es una pérdida de tiempo.	5 4 3 2 1
22	Para ser sincero(a), realmente no tengo ningún interés de aprender inglés.	5 4 3 2 1
23	Realmente trabajo mucho para aprender inglés.	5 4 3 2 1
24	Solo deseo aprender los aspectos básicos del inglés.	5 4 3 2 1
25	Estudiar inglés es importante porque me permitirá comunicarme más fácilmente con hablantes de inglés.	5 4 3 2 1
26	Deseo poder hablar en inglés fluidamente.	5 4 3 2 1
27	Estudiar inglés es importante porque otra gente me respetará más si hablo inglés.	5 4 3 2 1
28	Cuando salga del colegio, dejaré de estudiar inglés porque no estoy interesado(a) en el idioma.	5 4 3 2 1