

ASSESSING CONTENT AND LANGUAGE INTEGRATED LEARNING: A STUDY OF THE  
SECOND LANGUAGE ACQUISITION OF ENGLISH IN MONOLINGUAL ANDALUSIA (SPAIN)

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of  
Master of Arts

by

María del Mar Gutiérrez Ortiz

August 2016

© 2016 María del Mar Gutiérrez Ortiz

## ABSTRACT

This study examined the effect of Content and Language Integrated Learning (CLIL) in a group of Andalusian learners' English proficiency and syntax. High school students ( $n = 22$ ) enrolled in CLIL and non-CLIL classes in the same school (I.E.S. Mariana Pineda) took an English proficiency test adapted from the University Entrance Examination. Students were assessed for syntactic development in English using an Elicited Imitation (EI) task in conjunction with an experimental design based on the syntactic properties of English. Pre-intervention academic achievement, socio-economic status (SES), and gender were included as covariates in the analyses. CLIL students scored significantly higher than non-CLIL students on the syntactic task ( $p = 0.002$ ) and the proficiency test ( $p < 0.001$ ). Results revealed a significant positive correlation between the scores obtained in the proficiency test and the EI task for the CLIL group ( $p < 0.01$ ), but not for the non-CLIL group ( $p = 0.39$ ). These findings advance a methodology to assess the English syntax of CLIL students.

## BIOGRAPHICAL SKETCH

This author was born November 30, 1990 in Seville, Spain. She graduated in English from the University of Seville, after being an exchange student at the University of Edinburgh. She received an M. Ed. in Teaching English as a Second Language from the University of Seville. She is currently studying a Fulbright-funded M.A. in Developmental Psychology at Cornell University.

## ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor Prof. Barbara C. Lust for giving me a chance to work under her guidance. I have learned immensely from her rigorous scientific methodology and vast knowledge.

I would also like to thank my committee member Prof. Qi Wang for her insightful comments and suggestions.

I also thank Prof. Suzanne Flynn for the inspiration of her early seminal study.

Sincere thanks are also due to I.E.S. Mariana Pineda for facilitating data collection. In particular I would like to thank Inmaculada Gómez Barea and María del Carmen Gutiérrez Calderón for their helpful collaboration during this process. I also thank the 12<sup>th</sup> grade students who willingly took part in this study.

To all my Ithaca friends, I express thanks for the good times and study breaks. I am especially grateful to Rochelle Cassells for discussing and proof-reading the final text. For his loving understanding, I thank Robert Wijma who always takes such good care of me.

I thank my family for their support and encouragement. Above all, I am profoundly grateful to my father Joaquín Gutiérrez, my “coach” and my inspiration, for his invaluable help in this and all my academic adventures.

## Table of Contents

CHAPTER 1: Introduction .....	1
CHAPTER 2: Background.....	2
2.1. <i>Content and Language Integrated Learning</i> .....	2
2.2. <i>Theoretical framework</i> .....	4
2.2.1. L2 acquisition in bilingual education .....	4
2.2.2. Principal branching direction .....	6
2.3. <i>Literature review</i> .....	7
2.3.1. The CLIL advantage.....	7
2.3.2. The CLIL controversy .....	9
CHAPTER 3: Rationale and design.....	11
3.1. <i>Proficiency Test</i> .....	12
3.2. <i>Elicited Imitation Task</i> .....	12
CHAPTER 4: Method .....	14
4.1. <i>Program description</i> .....	14
4.2. <i>Participants</i> .....	15
4.3. <i>Materials</i> .....	16
4.4. <i>Procedure</i> .....	17
4.5. <i>Coding</i> .....	17
CHAPTER 5: Results .....	19
5.1. <i>Demographic and linguistic information</i> .....	19
5.2. <i>CLIL effects on proficiency</i> .....	21
5.3. <i>CLIL effects on syntactic development</i> .....	22
5.4. <i>Correlation between proficiency and syntactic development</i> .....	24
5.5. <i>Overall difference between CLIL and non-CLIL students</i> .....	24
CHAPTER 6: Discussion.....	25
REFERENCES .....	29
APPENDICES.....	36
APPENDIX I.....	36
APPENDIX II.....	39
APPENDIX III.....	42
APPENDIX IV .....	43

# CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

## CHAPTER 1: Introduction

With 24 official languages and over 60 minority languages, the European Union (EU) is home to a multilingual community. The command of foreign languages is fundamental for communication between the different members of the EU, so the multilingualism policy of the European Commission promotes language learning. To examine multilingual experiences across the different European countries, in 2005 the Directorate-General for Education and Culture of the EU requested a Special Eurobarometer study on multilingualism. This survey revealed that 56% of Spanish citizens did not speak a foreign language well enough to have a conversation, and only 17% spoke at least three languages (Directorate General for Education and Culture, 2006). As a consequence, the Spanish government developed a plan to redefine the language learning system by implementing *Content and Language Integrated Learning* (CLIL) programs across the nation (Ruiz de Zarobe & Lasagabaster, 2010).

Marsh (2002) defined CLIL as a “dual-focused educational context in which a foreign language is used as a medium in the teaching and learning of non-language content.” In CLIL lessons, content and language are integrated; the foreign language is a vehicle to teach content subjects, and content subjects are designed to promote language learning. Although all CLIL programs are based on this premise, methods, materials and curriculum organization vary across and within the different European countries. Therefore, the acquisition of foreign languages in CLIL contexts is determined both by the general features of the approach and the specific characteristics of each particular program.

Previous research has found a consistent advantage of CLIL students in the second language (L2) proficiency, but few studies have focused on the students' command of the L2 grammar. Proficiency, the effectiveness in using the language, is determined both by the students' knowledge of the language grammar and extra-linguistic factors such as the features of the pedagogical intervention (Housen & Kuiken, 2009). For example, oral skills tend to

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

improve in interactive lessons that foster communication between students (Kayi, 2006). Therefore, the proficiency of students attending different CLIL programs may be partially attributed to the specific characteristics of each particular program. In turn, a more direct assessment of the students' command of the grammar may be more revealing of the context-free effects of the CLIL approach on the students' knowledge of the language.

The central objective of this study was to investigate the effect of the participation in a CLIL program on the students' acquisition of English as an L2. The study was conducted in a Spanish high school located in Andalusia, a traditionally monolingual region where CLIL programs have been implemented in the vast majority of public schools to foster the acquisition of English as a second language (L2). CLIL students were compared to their non-CLIL peers on dimensions of English proficiency and knowledge of the language. In addition to standardized school estimates that measure proficiency, an Elicited Imitation (EI) task was designed to evaluate the development of language knowledge in the acquisition of English as a foreign language. This paper will inform us about the development of grammar in CLIL, and by doing so, a new approach is simultaneously advanced for measuring English language acquisition. Results suggest that direct assessment of a students' syntax through the EI task may be an effective approach to evaluating high school CLIL programs in different educational contexts.

### CHAPTER 2: Background

#### **2.1. *Content and Language Integrated Learning***

Bilingual education refers to enriched educational contexts in which an L2 is used as a vehicle to teach content courses, such as history and math (Feinberg, 2002). Thus, students are able to acquire the language through meaningful and authentic communication while gaining competence in the required curricular content (Pinner, 2013). CLIL was developed in Europe in the 1990s based on Canadian and American bilingual education. On the one hand, in Canada, immersion programs emerged with the objective of instructing native speakers of the majority

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

language in a foreign language (Heras & Lasagabaster, 2015). On the other hand, American additive bilingual education programs were implemented for minority students to become proficient and literate in both their mother tongue and the country language (Heras & Lasagabaster, 2015). Despite its historical roots, CLIL emerged as a distinct bilingual program with specific characteristics.

Like any bilingual program, CLIL uses a communicative approach in the classroom to promote the students' proficiency in an L2. In the CLIL approach, language and content are integrated in two ways: "weak CLIL" and "strong CLIL" (Paran, 2013). In weak CLIL, relevant contents (e.g. pop culture or sports) are used to teach the foreign language (language objectives and content focus). In strong CLIL, the L2 is used to teach course content, e.g. history or math (language focus and content objectives): the foreign language is a vehicle to teach the subjects and the subjects are designed to promote the acquisition of the language.

Strong CLIL has been adopted by many European education systems as the most effective way to meet the guidelines of the EU that emphasize multilingualism and language diversity (Ruiz de Zarobe & Lasagabaster, 2010). In line with the European trends, the implementation of CLIL has witnessed a dramatic increase in Spain. The legislative framework guiding the Spanish education system establishes the principles and rights at a national level. All Spanish CLIL programs share a number of key features: a high percentage of the teaching staff is made up of non-native speakers of the target language without training on CLIL programs, many students start in secondary education, and the materials used in class are usually abridged (Lasagabaster & Sierra, 2010). As a result of this, the students enrolled in CLIL aim at becoming competent users of the L2, rather than acquiring native-like proficiency.

In Spain, each of the 17 regional governments adapts the education law in different ways. Andalusia is a monolingual region that in 2014-2015 had developed a CLIL school network of 957 institutions (Portal de Plurilingüismo). The Andalusian government has implemented a "Plan to

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

Foster Multilingualism" [Plan de Fomento del Plurilingüismo] that establishes the guidelines for CLIL schools. All CLIL schools have access to the same resources and materials through the government website. Andalusia is of particular interest given that it is the most populated region in Spain. Additionally, by conducting the experiment in a CLIL school located in a traditionally monolingual community, the effect of CLIL on the acquisition of English as an L2 was examined.

### **2.2. Theoretical framework**

#### 2.2.1. L2 acquisition in bilingual education

In the past decades, a vast number of theories have tried to account for the nature of language development and the role of input in the process of language acquisition (see Lust, 2011 for a review). Research has provided evidence that Universal Grammar (UG), the innate human language faculty, remains intact and constantly accessible during the process of first and second language acquisition (Epstein, Flynn, & Martohardjono, 1996). In order to account for the development of the specific languages despite the invariability of UG, Epstein et al. (1996) highlight the need to draw a distinction between the innate language faculty and the grammar of specific languages. This view is fleshed out in the Grammatical Mapping (GM) paradigm, which reconciles the role of Universal Grammar (UG) and input in the process of language acquisition (Lust, 2012). According to the GM, children use UG to develop the Specific Language Grammar (SLG), after being exposed to the data of a particular language (Lust, 2012). UG is constituted by principles, invariant properties common to all languages, and parameters, binary switches that are set to one of the values depending on the linguistic input (Chomsky, 1980). Principles account for language universals, whereas parameters account for language variation. L2 learners are guided by UG to develop the L2 from exposure, so they need to reset the parameters in the creation of the L2 grammar (Epstein et al., 1996). Since the knowledge of previous languages also plays a role in subsequent language acquisition (Flynn, Foley, &

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

Vinnitskaya, 2004), the SLG of the L2 is determined by UG, influenced by the SLG of the first language, and computed over the L2 data.

Competence is the knowledge of SLG, and performance is the use of such knowledge in specific situations (Chomsky, 1965). In L2 acquisition research, the term performance is frequently associated with proficiency, that is, “skill in performance, the adeptness in using language in comprehension or expression” (Francis, 2011, p. 3). Typically, the L2 is assessed by means of proficiency tests that yield scores according to the complexity, accuracy, and fluency (CAF) of the students’ oral and written productions. Housen and Kuiken (2009) explain that the characteristics of the language task, the personality and socio-psychological features of the learner, and the features of the pedagogical intervention may influence the manifestation of CAF in the use of the L2. Therefore, performance (and proficiency) is not only determined by linguistic knowledge (competence), but also by processing abilities (Francis, 2011) and nonlinguistic factors. Speakers may be unable to produce well-formed utterances despite having good grasp of the language, and L2 learners often handle information in the L2 without full competence (Francis, 2011).

Bilingual education provides conditions for L2 acquisition that resemble the environment of first language acquisition. Students acquire the language implicitly by using it in meaningful interactions. According to the GM paradigm, this additional exposure to the L2 data would result in a faster development of the target language grammar, which would, in turn, boost the students’ performance. However, researchers have argued that, when L2s are acquired in natural settings, age of acquisition may be better predictor of morpho-syntactic level and overall proficiency than “length of residence” or social-psychological factors (DeKeyser & Larson-Hall, 2005). Individuals who are exposed to the language earlier tend to achieve higher levels of proficiency and higher command over some aspects of the L2 morpho-syntax. The “age of acquisition effect” together

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

with the specific characteristics of the CLIL approach only allow for a partial development of the students' L2 (competent use vs. native-like proficiency) in CLIL classrooms.

### 2.2.2. Principal branching direction

In the creation of the SLG of a first or a second language, learners integrate parameter setting with the principles of UG (Lust, 2012). Binary parameters are set to one of the values after exposure to the data and determine certain fundamental properties of the language grammar. The principal branching direction (PBD), also called head initial/head final parameter, is a parameter that classifies languages according to their predominant branching direction (Lust, Bhatia, Gair, Sharma, & Khare, 1995).

Principal Branching Direction refers to the branching direction which holds consistently in unmarked form over major recursive structures of a language, where "major recursive structures" are defined to include embeddings of sentence complements under either NP or S "heads." Specifically, relative clauses in complex NP and adverbial subordinate clauses are critical to the definition of this parameter. (Lust et al., 1995, p. 199)

Thus, right branching languages (head-initial languages) tend to include embedding to the right of their heads, i.e. the unmarked position for subordinate clauses is to the right of the sentence (Lust et al., 1995). Although not all languages have a perfectly consistent branching direction, and all languages appear to allow alternations in the order of components (Flynn, 1983), some languages, like English, allow the characterization of head-initial languages, as exemplified in (1) and (2) (Flynn & Espinal, 1985):

(1) [The child [who is eating rice]] is crying

(2) [The child drank the milk [after he ate the rice.]]

Research has shown that in early child language, the direction of grammatical anaphora accords with the principal branching direction of the language. In English, the acquisition of

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

forward anaphora (3) precedes the acquisition of backward anaphora (4), because it coheres with the head-initial configuration of the language (Lust, 1981):

(3) *John* read the play while *he* smoked a pipe.

(4) While *he* smoked a pipe, *John* read the play.

In the same way, Flynn (1983) has demonstrated that the PBD is also a structural constraint in the acquisition of English as an L2. Thus, Spanish speakers learning English as an L2 are sensitive to the head-direction of the L2, and prefer forward to backward anaphora in complex sentences. Flynn (1986) showed that Spanish L2 learners of English acquired complex sentences with post-posed embedded clauses and forward anaphora, such as in (5), before complex sentences with pre-posed embedded clauses and backward anaphora in (6):

(5) The man answered the boss when he installed the television.

(6) When he entered the office, the professor questioned the man.

Beginners did not show significant differences between the two types of sentences because they had not acquired any of them. Intermediate students mastered sentences with post-posed clauses and forward anaphora (5), significantly more than sentences with pre-posed clauses and backward anaphora (6). Advanced students had command over both types. In the present study, Flynn's findings were used to assess the grammatical knowledge of students in CLIL and non-CLIL groups. The amount correct imitations that students produce for each type of sentence was used to evaluate the students' knowledge of the L2 grammar.

### **2.3. Literature review**

#### 2.3.1. The CLIL advantage

##### *The advantage in proficiency*

A general advantage in English proficiency has been repeatedly reported for learners enrolled in CLIL programs when compared to learners who only receive English through foreign language lessons (Dalton-Puffer, 2008). In the Spanish context, CLIL students have been

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

shown to surpass non-CLIL students on reading comprehension (Lorenzo, Casal, & Moore, 2010; Navés, 2011; Pérez-Vidal & Roquet, 2015; Prieto-Arranz, Rallo Fabra, Calafat-Ripoll, & Catrain-González, 2015), vocabulary (Jiménez-Catalán & Ruiz de Zarobe, 2009; Moreno Espinosa, 2009), fluency, and lexical and syntactic complexity in written productions (Gené-Gil, Juan-Garau, & Salazar-Noguera, 2015; Lasagabaster, 2008; Lorenzo et al., 2010; Navés & Victori, 2010; Pérez-Vidal & Roquet, 2015; Ruiz de Zarobe, 2010), oral production (Lasagabaster, 2008; Lorenzo et al., 2010; Ruiz de Zarobe, 2008) and metacognitive awareness to select appropriate learning strategies (Ruiz de Zarobe & Zenotz, 2012, 2015). Research has obtained contradictory results for listening comprehension: whereas Navés (2011) and Pérez-Vidal and Roquet (2015) did not find a difference between CLIL and non-CLIL learners in the Catalanian context, Lorenzo et al. (2010), Lasagabaster (2008) and Prieto-Arranz et al. (2015) found a positive effect of CLIL on listening comprehension.

### *The advantage in grammar*

The studies that have compared the grammatical development of CLIL and non-CLIL groups in Spain have made use of natural communication. By asking students to retell the picture story "Frog, where are you?", they have obtained disparate results about different aspects of the students' morpho-syntax. In a longitudinal analysis of the oral narratives of Basque/Spanish bilinguals, Ruiz Zarobe (2008) found that CLIL groups outperformed non-CLIL groups in grammatical accuracy. Villarreal and García Mayo (2009) demonstrated that CLIL students omitted affixal –s and –ed less than non-CLIL students, but they did not find significant differences between the omission rate of the suppletive forms. Martínez Adrián and Gutiérrez Mangado (2009) found significant differences between the number of placeholders and embedded clauses used by CLIL and non-CLIL students, but not between the production of null subjects, null objects, and correct negative structures. Lázaro Ibarrola (2012) demonstrated that

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

CLIL students produce higher rates of inflected verbs, higher correction rates in the use of pronouns, and a higher total number of pronouns and subordinate sentences.

By eliciting the narration of the picture story, all these studies compared the morpho-syntax that spontaneously arises in the students' discourse. Since the students' productions are not only determined by their knowledge of the language, but also by processing abilities and external factors, these studies provided a measure of language in use, i.e. a measure of performance and not a measure of linguistic competence. In fact, a study comparing successive and simultaneous bilinguals found that the children's performance in the narration task was not always correlated with their actual knowledge of the grammar (Kim, Park, & Lust, 2016). Given that elicited narration does not necessarily reflect language knowledge, a different task has to be used to measure the students' linguistic competence more directly.

### 2.3.2. The CLIL controversy

After an initial praise of CLIL, a number of researchers began to question the validity of these previous studies (see Pérez Cañado, 2016 for a review). Paran (2013) argues that in order for a CLIL program to work, a number of conditions must be fulfilled. On the one hand, CLIL programs are effective in countries with more exposure to English, where students have higher academic achievement, and proficient teachers in the L2 are educated on the CLIL approach. However, Spain ranked among the bottom in foreign language knowledge across European countries (TNS Opinion & Social, 2012), and below the average in the Programme for International Student Assessment (PISA) survey that assesses the competencies of 15-year-olds in reading, mathematics and science in 65 countries and economies (OECD, 2014). In addition, most of the teacher training programs in Spain do not include specific courses on CLIL education (Lasagabaster & Sierra, 2010). Teachers are now required to hold an advanced certificate (C1) to teach CLIL groups, but the acquisition of a foreign language was not mandatory in education degrees until 2010 (European Commission/EACEA/Eurydice, 2015).

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

Although regional authorities attempt to compensate for this lack of formal training by offering language learning courses, seminars, conferences, and methodology courses (Lasagabaster & Sierra, 2010), this first set of criteria established by Paran (2013) is mostly unfulfilled in the Spanish context.

On the other hand, although CLIL was devised for learners of a variety of backgrounds (Coyle, Hood, & Marsh, 2010), CLIL works better where it is implemented selectively, and tends to be more effective for high achievers (Paran, 2013). In Andalusia, CLIL is currently implemented in all compulsory public education stages and it is accessible to all students (Pérez Cañado, 2016). However, does this accessibility guarantee the egalitarianism of the approach? Because students have the choice of enrolling in the CLIL or the non-CLIL program at the beginning of secondary education, CLIL groups are “self-selected” (Pérez Cañado, 2016). Thus, CLIL groups tend to have more motivated students with higher SES and better academic outcomes (Bruton, 2011a, 2011b, 2013, 2015). If CLIL groups are self-selected, then the reasons that led certain students to enroll in CLIL programs may also have an impact on their scores. The advantage of CLIL students may not only be due to their participation in the CLIL program, but to pre-existing differences between groups. In fact, a study conducted in the Catalanian context found that CLIL students of grades 7 and 9 achieve equal or higher levels than non-CLIL learners who were in higher grades on writing and overall proficiency (Navés & Victori, 2010). Over time, the differences were ameliorated, which shows that the initial results may have reflected a pre-existing imbalance. Most of the studies showing the advantages of the program have not controlled for moderating variables (Pérez Cañado, 2016). Paran (2013) and Pérez Cañado (2016) highlight need for new research that ensures the homogeneity of the experimental and control groups by means of statistical procedures.

The dissimilarity between CLIL programs constitutes an additional difficulty for research. Because CLIL education is broadly defined as a dual language approach that integrates

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

language and content (Marsh, 2002), many different CLIL programs have been implemented across educational contexts (see Pérez Cañado, 2016 for a discussion on the vagueness of the term). Programs not only differ on the courses taught via the foreign language and the methodologies used in the classroom (Pérez Cañado, 2016), but also on the language background of the students. If different versions of the approach focus on certain linguistic skills, the students' performance on proficiency tests may vary remarkably across educational contexts. Additionally, all the research on the morpho-syntactic development by CLIL students has been conducted in bilingual regions, where students are already bilingual in Spanish and Basque or Catalan and learn English as a third language (L3). Therefore, to the best of my knowledge, no study to date has explored the effect of CLIL participation in the acquisition of the L2 grammar in Spain.

Even though a large body of research has argued for the advantage of the CLIL approach in the Spanish context, all studies have examined the students' performance in proficiency tests and none of them has looked into the acquisition of grammar specifically. Additionally, the morpho-syntax of CLIL students has only been explored in bilingual regions where students acquire English as an L3. Variability among the different versions of the approach and lack of control for pre-intervention differences in previous studies call for new research that uses comprehensive methodology to measure the students' competence and performance across educational contexts, while including pre-intervention SES, English proficiency, and academic achievement as covariates.

### CHAPTER 3: Rationale and design.

The present study exemplifies the use of a more comprehensive methodology to measure the effect of the CLIL approach in the students' level of English as an L2. CLIL and non-CLIL students are compared for the scores obtained in a proficiency test and on an EI task that taps into their grammatical knowledge through an experimental design. The research

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

design was cross-sectional and quasi-experimental as the assignment to each condition (intervention vs. no intervention) was by self-selection. Since quasi-experimental designs lack random assignment, a comparison group was identified that was as similar as possible to the experimental group in terms of pre-intervention characteristics. Thus, the CLIL group was compared with 12<sup>th</sup>-grade students that attend the non-CLIL program in the same high school. To ensure the comparability between groups, SES, and academic achievement before the intervention were included as covariates.

### **3.1. Proficiency Test**

The students' English performance was evaluated using the mandatory University Entrance examination. This test is designed to evaluate the minimum level that all 12<sup>th</sup>-grade students need to achieve at the end of secondary education. Because passing the test contributes to the students' access to higher education, the results of the present study are informative about the potential impact of CLIL on the students' future careers. In addition to learning about the effect of CLIL on the students' English proficiency, we intend to gain further insight into the effect of CLIL on the youth's potential to enter a globalized job market.

The independent variables are the program (CLIL or non-CLIL) X gender with covariates, SES, and Grade Point Average (GPA). The dependent variable is the score obtained in the proficiency test. To check the effect of the participation in CLIL on the students' performance in each part of the test, we used the score obtained in the different parts of the tests as dependent variables.

### **3.2. Elicited Imitation Task**

The EI task is an experimental task used in conjunction with experimental designs to assess grammatical knowledge (Lust, Flynn, & Foley, 1996). In the EI task, the researcher reads aloud a sentence and asks the participant to repeat it according to standardized methods for administration. This task is based on the assumption that in order to repeat the sentence, the

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

participant must analyze the structure of the sentence they have heard (including vocabulary, syntax, and meaning), and then reconstruct this structure in their production of the model. The stimulus sentences must be too long to be stored in the short-term memory without being analyzed (Flynn & Espinal, 1985). So in order to "imitate" a structure, the structure must be part of the speaker's grammatical competence (Lust et al., 1996). The researcher can analyze the participants' knowledge of specific aspects of the grammar by modeling sentences that vary only in critical factors with all other held constant.

The critical factors that vary in our experimental design are based on Flynn's findings (1986): our two groups had to repeat a series of complex sentences with adverbial subordinate clauses and pronoun anaphora. The direction of the pronoun anaphora was coherent with the position of the subordinate clause in relation to the main clause. Thus, in post-posed clauses the antecedent preceded the pronoun (forward anaphora), and in the pre-posed clauses the pronoun preceded the antecedent (backward anaphora):

(7) Post-posed clause with forward anaphora: The man answered the boss when he installed the television.

(8) Pre-posed clause with backward anaphora: When he installed the television, the man answered the boss.

Because English is a right-branching language, post-posed clauses with forward anaphora (7) are unmarked, and so they are acquired before pre-posed clauses with backward anaphora (8). This developmental pattern was used to determine the students' level of syntax by computing not only the total number of sentences imitated, but also the type of sentences that each group produced correctly. In line with previous findings (Flynn, 1986), significant differences between (7) and (8) are only expected for the non-CLIL group (equivalent to Flynn's (1986) intermediate group). If the CLIL group has already acquired the two types of sentences, they will produce a similar number of correct imitations for all of them.

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

Even though Flynn (1983) did not find an independent effect of anaphora or branching direction on the number of correct imitations, a third type of sentence with pre-posed subordinate clause and forward anaphora (9) have been included to control whether the direction of the anaphora (backward vs. forward) alone had an effect on the number of correct imitations within left branching clauses and whether the position of the clause (pre-posed vs. postposed) had an effect in the production of sentences with forward anaphora:

(9) Pre-posed clause with forward anaphora: When the doctor received the results, he called the gentleman

The independent variables are the program (CLIL or non-CLIL) X gender with covariates, SES and GPA. The dependent variable is the total number of correct imitations. To measure the difference in the number of correct imitations for each type of sentence, we used a generalized mixed model with subject as a random effect and type of sentence, gender, SES, and GPA as fixed effects.

### **3.3. Hypotheses**

The first hypothesis of this study is that students in the CLIL group would outperform students in the non-CLIL group in the proficiency test and the EI task. The second hypothesis is that students in the CLIL group would score similarly for the three types of complex sentences in the EI imitation task, whereas non-CLIL students would be weaker on complex sentences with pre-posed clauses and backward anaphora. The third hypothesis is that if both the standardized proficiency tests and the EI test measure language knowledge, then that results of the proficiency test and the EI task would be strongly correlated.

## CHAPTER 4: Method

### **4.1. Program description**

We investigated the achievements of the students in I.E.S. Mariana Pineda, a public high school located in Montequinto, a town near Seville (the main city in Andalusia). The school

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

offers a CLIL program from 1° E.S.O. (7th grade) through 2° de Bachillerato (12th grade). The content courses taught in the foreign language vary every year:

- 1° E.S.O (7th grade): Music, Natural Sciences, and Social Sciences.
- 2° E.S.O (8th grade): Technology, Music, Natural Sciences and Social Sciences.
- 3° E.S.O (9th grade): Technology, Natural Sciences, and Social Sciences.
- 4° E.S.O (10th grade): Social Sciences, Ethics and Integrated Project.
- 1° Bachillerato (11th grade): Science of Contemporary World, Philosophy and Integrated Project.
- 2° Bachillerato (12th grade): Integrated Project.

All students attended English as foreign language classes (EFL) during six years of primary education. In secondary education, the experimental group attended CLIL content classes (CLIL) and EFL, whereas the control group only learned English during the lessons of EFL (table 1).

	1° E.S.O.		2° E.S.O.		3° E.S.O.		4° E.S.O.		1° Bac.		2° Bac		TOTAL
	EFL/CLIL		EFL/CLIL		EFL/CLIL		EFL/CLIL		EFL/CLIL		EFL/CLIL		
<b>Non-CLIL</b>	4		4		4		4		3		3		22
<b>CLIL</b>	4	8	4	11	4	8	4	6	3	6	3	1	62

Table 1. Number of CLIL and EFL hours that each group receives weekly during secondary school.

### 4.2. Participants

Participants were 22 Spanish students between the ages of 17 and 18 years. The CLIL group ( $n = 11$ ) had been attending a CLIL program for 6 years, although one of the students has only been in the program for 2 years. The non-CLIL group ( $n = 11$ ) received all content lessons in Spanish and EFL three hours a week. All the parents of underage students and 18-year-old students completed consent forms before testing began. No monetary compensation was given for students' participation in this study.

The differences between groups were controlled with a linguistic and demographic questionnaire. In Spain, the fact that both groups attend the same public high school determines largely their SES. Even so, students provided their parents' occupation and highest academic attainment. SES was obtained from the fathers' occupation and education using the

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

International Socio-Economic Index of Occupational Status (Ganzeboom, De Graaf, & Treiman, 1992). Since the students' GPA prior to entering the program could not be retrieved, the grades obtained in the first semester of high school were used as covariates. High school grades usually reflect academic performance better than school grades, which take into account attitudinal factors that are irrelevant to the present study. In addition, at that point, students had only attended three CLIL classes for three months (mid-September to mid-December). Because we did not have pre-intervention scores for English proficiency, the students completed a survey about their linguistic background that did not reveal significant differences between the groups.

### **4.3. Materials**

*Language questionnaire.* A questionnaire was used for gathering linguistic and demographic background information of students (see Appendix I). Students provided the age and context of the second and L3 acquisition, as well as information about their use of the language outside the classroom. Information was gathered about the parents' education and occupation prior to enrollment in the program. Students reported their GPA from the first year of compulsory education.

*Proficiency test.* The proficiency test was adapted from the reserve English test in the University Entrance Examination of September 2014. The exam was shortened to one hour, so students had to read a text, answer four reading comprehension questions (instead of five), complete 11 use of English exercises (vocabulary and grammar), and write a text of 100 words (instead of 150 words). The complete test is given in Appendix II.

*Elicited Imitation.* This experimental task was adapted from Flynn (1986) and consisted of a battery of 9 sentences (3 iterations for each type) to specifically examine the learners' knowledge of embedded clauses in English. Participants were asked to repeat the sentences one by one after the experimenter. All the sentences were matched for length (15 syllables). The stimuli were counterbalanced across trials using an online randomizer. See Appendix III for a full list of sentences.

### **4.4. Procedure**

Two teachers in Mariana Pineda high school informed students that they would have the opportunity of participating in a study that would evaluate the effect of the CLIL program on the acquisition of English as an L2. Students that expressed willingness to participate received consent forms. Once the consent forms had been signed, the researcher attended the school to answer specific questions about the procedure. During the same session, the researcher examined students about the vocabulary of the EI task, and asked them to memorize the words they did not know. After a short interval of time, a specific computerized link for the proficiency test was assigned to the participants. In the first part of the experiment, participants completed the proficiency test during the Computer Science class. Students were not allowed to consult any extra materials and only questions about the format of the exam were solved. The exam was proctored by the collaboration between teachers and the researcher.

After an average of three days, students were called individually to do the second part of the experiment. First, students were administered an online questionnaire about demographic information and previous experience with the language. Then they listened to the list of complex sentences in random order and repeated them one by one. Their answers were recorded with *Audacity*. The recordings and their transcriptions were coded by researchers trained in language analyses according to standardized procedures.

### **4.5. Coding**

*Language Questionnaire.* The highest academic qualification received by both parents was divided into five levels: 1- None/Primary, 2- Secondary/Intermediate professional training, 3- Pre-university/Advanced professional training, 4- Tertiary and 5-Postgraduate. In addition, the occupation of the father was coded using the ISEI scores for Occupation categories (Ganzeboom et al., 1992), which transforms the father's education and occupation into SES. The ISEI score was standardized for the descriptive tables and centered for the statistical

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

analyses. The students reported their GPA from the first year of secondary education on a scale of 0 to 10, and this score was centered for the statistical analyses.

*Proficiency test.* The guidelines established for the correction of the University Entrance examination have been adapted for this proficiency test. In the reading comprehension section, students could obtain up to 3 points, 1.5 for two multiple choice questions and 1.5 for two true or false questions. The 11 use of English questions were worth 4 points. A specific set of guidelines were taken into account for the writing task. The score (up to 3 points) was based on grammatical correction (repeated mistakes were only taken into account once), lexical richness and accuracy, and textual and communicative aspects. The maximum total score in the proficiency test was 10. To ensure scoring reliability, the composition was scored by the researcher and a research assistant. Disagreements were resolved by discussion, and in the event that consensus could not be reached, the two grades were averaged.

*Elicited Imitation task.* In order for an EI response to be scored correct (1 point), the participant had to repeat the stimulus sentence given without any major syntactic or semantic change. Lexical substitutions that did not imply a substantial change in the meaning of the sentence were not considered a semantic error. For example, the substitution of *gentleman* for *man* did not constitute an error, but the substitution of *actor* for *lawyer* did. Incidental changes made by the participant which did not alter the syntactic structure or meaning in ways that are relevant to the focus of this research, such as singular to plural or present tense to past tense, and mispronunciations were not be scored as incorrect. Changes that altered the syntactic structure or meaning of the original stimulus sentences (e.g., repetition of only one clause, alteration of the original anaphora/antecedent relation, change in clause order, or repetition of a lexical item not considered to be a synonym) were scored as incorrect (0 points). See Appendix IV for coding criteria. Because there were three iterations of each type of sentence, the maximum score students could obtain was 9. The researcher transcribed all the sentences. Reliability was

determined by having a native Spanish speaker who was bilingual in English and a native English speaker code all the data. Agreement between coders was 100%.

CHAPTER 5: Results

This chapter presents the results of the statistical analyses conducted for the study. The first section elucidates the students’ linguistic background and demographic information. Then, the results of the proficiency test, the EI task, and the correlation between them are reported.

Before all the analyses, assumption checks were performed. Normality of residuals was checked by graphical methods (normal q-q plot). A plot of the residuals against the predicted value revealed fairly linear relationships between continuous data. There was not a discernable pattern to the Scale Location plots (square root of the standardized residuals vs. fitted values), so the homoscedastic error assumption was not violated. Neither the Cook’s D plot, nor the Residuals vs. Leverage plot pointed to the presence of influential outliers.

**5.1. Demographic and linguistic information**

The questionnaire revealed the students’ homogenous background. All of them are native speakers of Spanish who started attending English as a foreign language (3 hours a week) in the first year of primary education (age 6-7), and 16 students (CLIL = 9 and non-CLIL = 7) took French in secondary education. None of the students had spent more than one week in an English speaking country.

<b>GROUP</b>	<b>GENDER</b>	<b>Extracurricular Mean (SD)</b>	<b>SES (z-score) Mean (SD)</b>	<b>GPA Mean (SD)</b>
<b>Non-CLIL</b>	Male ( <i>n</i> = 8)	0.37 (0.99)	-0.25 (1.12)	6.85 (0.65)
	Female ( <i>n</i> = 3)	0.00	0.64 (1.52)	6.54 (0.42)
	Both genders ( <i>n</i> = 11)	0.27 (0.86)	0.008 (1.22)	6.77 (0.61)
<b>CLIL</b>	Male ( <i>n</i> = 6)	3.5 (3.54)	-0.28 (0.66)	7.17 (0.84)
	Female ( <i>n</i> = 5)	1.6 (3.2)	0.36 (8.38)	7.69 (0.82)
	Both genders ( <i>n</i> = 11)	2.63 (3.52)	-0.008 (0.78)	7.41 (0.87)

Table 2. Summary of the background information of the students by group and gender.

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

Students in the CLIL group attended extracurricular English classes for more years than students in the non-CLIL group before entering secondary education, but results of a t-test did not indicate significant differences between groups ( $t(20) = 1.58, p = 0.13$ ). In addition, parents in the CLIL group had a slightly higher education (fig.1 and 2). However, according to the ISEI class scheme, students in the CLIL class did not have significantly higher SES than students in the non-CLIL class,  $t(20) = 0.37, p = 1.00$ . Although students in the CLIL group had slightly higher GPA than students in the non-CLIL group at the onset of the intervention, the difference was not significant ( $t(20) = 1.9, p = 0.70$ ).

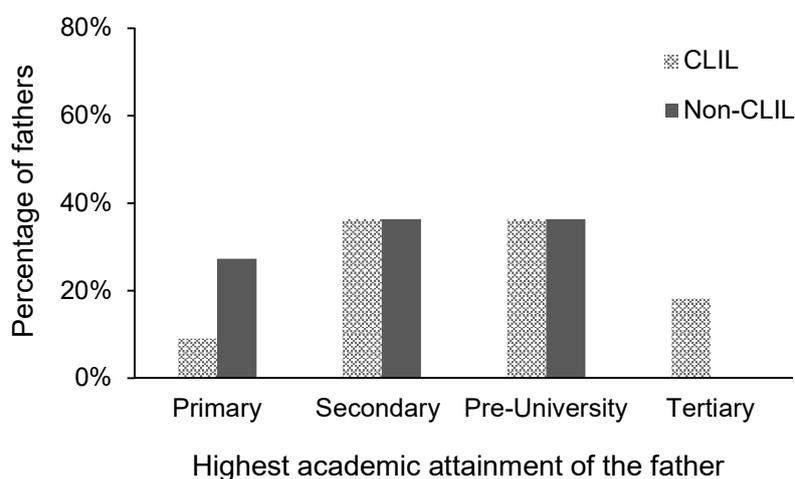


Figure 1. Highest academic achievement of the fathers in the CLIL and non-CLIL groups.

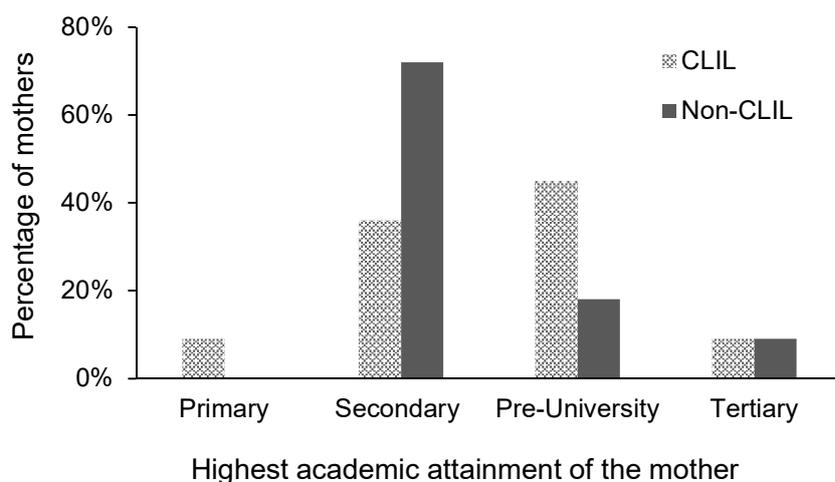


Figure 2. Highest academic achievement of the mothers in the CLIL and non-CLIL groups.

**5.2. CLIL effects on proficiency**

Table 3 shows that students in the CLIL group outscored their peers in the non-CLIL group in all the parts of the proficiency test.

<b>GROUP</b>	<b>Reading (max 3)</b> Mean (SD)	<b>Use (max 4)</b> Mean (SD)	<b>Writing (max 3)</b> Mean (SD)	<b>Overall (max 10)</b> Mean (SD)
<b>Non-CLIL</b>	1.57 (0.93)	1.27 (0.58)	0.84 (0.55)	3.70 (1.56)
<b>CLIL</b>	2.52 (0.58)	3.05 (0.58)	2.30 (0.53)	8.00 (1.26)
<b>Total</b>	2.05 (0.91)	2.16 (1.06)	1.57 (0.91)	5.85 (2.57)

Table 3. Mean score obtained by students of both groups in each part of the proficiency test and in the complete test (overall).

A generalized Analysis of Covariance (ANCOVA) was conducted to determine a statistically significant difference between CLIL and non-CLIL students on the overall score of the proficiency test and scores of every part of the proficiency test including gender, GPA, and SES as covariates. There was a statistically significant main effect of CLIL on the overall score of the proficiency test,  $F(1, 17) = 44.28, p < .001$ . Neither gender, nor GPA or SES had a significant effect on the proficiency scores (all  $ps. < 0.05$ ).

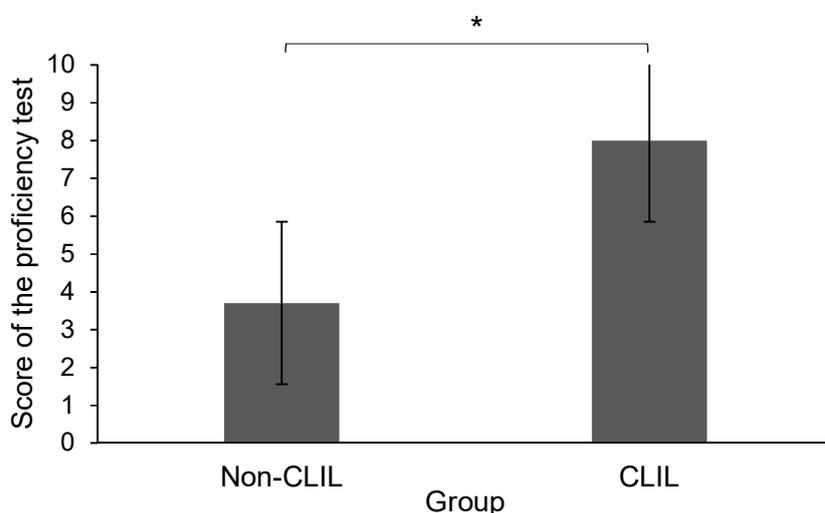


Figure 3. Mean score obtained by students in the CLIL and non-CLIL groups in the proficiency test.

A significant effect of intervention was found for reading comprehension ( $F(1, 17) = 7.55, p = 0.01$ ), use of English ( $F(1, 17) = 42.91, p < 0.001$ ) and written production ( $F(1, 17) = 44.06,$

$p < 0.001$ . In sum, the CLIL group scored significantly above the non-CLIL group in the three parts of the test (reading, writing and use of English).

### 5.3. CLIL effects on syntactic development

In keeping with the design, grammatical language knowledge was tested through an EI task with the experimental design explained above. Students in the CLIL group surpassed students in the non-CLIL group for each type of sentence (table 4). As predicted, students repeated more sentences with post-posed clauses and forward anaphora than sentences with pre-posed clauses and backward anaphora. Three was the maximum score students could obtain for each type of sentence, and nine the maximum overall score (table 4).

<b>GROUP</b>	<b>Pre-posed backward</b> Mean (SD)	<b>Pre-posed forward</b> Mean (SD)	<b>Postposed forward</b> Mean (SD)	<b>Total</b> Mean (SD)
<b>Non-CLIL</b>	0.45 (0.66)	0.91 (0.79)	1.18 (0.94)	2.54 (1.78)
<b>CLIL</b>	2 (1.04)	2.09 (1)	2.18 (0.94)	6.27 (2.6)
<b>Total</b>	1.23 (1.17)	1.5 (1.07)	1.68 (1.06)	4.41 (2.9)

Table 4. Mean score obtained by students of both groups in each type of the EI task and total score.

A generalized ANCOVA indicated a significant effect of the intervention on the total score of the EI task ( $F(1, 17) = 12.15, p = 0.003$ ), fig 4. However, there was no effect of gender ( $F(1, 17) = 0.29, p = 0.6$ ), GPA ( $F(1, 17) = 0, p = 0.98$ ) or SES ( $F(1, 17) = 0.04, p = 0.85$ ).

The number of correct imitations for each type of sentence was analyzed using a generalized mixed model where gender, GPA, SES and type of sentence were included as fixed effects, and student ID as a random effect. In the overall, the scores of both groups were higher for the sentences that are acquired earlier, i.e. sentences with post-posed clauses and forward anaphora in right branching languages, fig. 5. A marginal effect of type ( $F(2, 42) = 2.63, p = 0.08$ ) and a significant difference between pre-posed backward and post-posed forward sentences ( $t(42) = 2.28, p = 0.03$ ) were found.

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

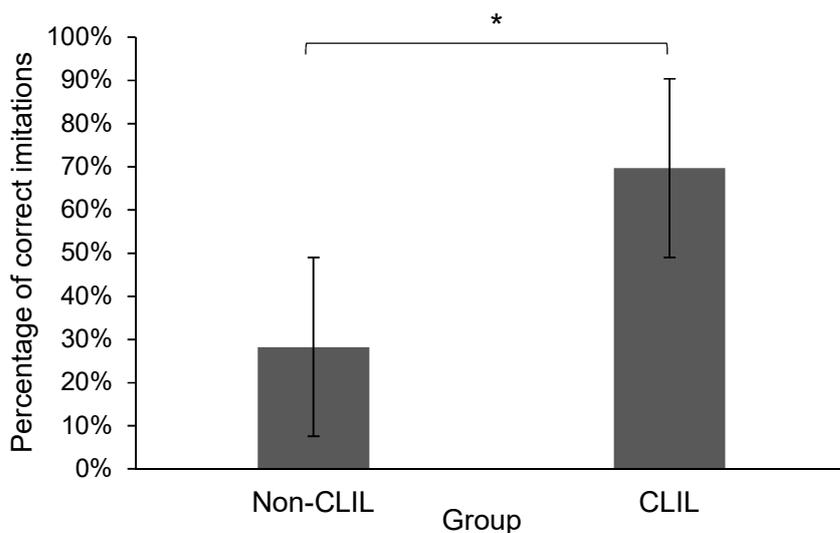


Figure 4. Percentage of total sentences correctly imitated by students in the CLIL and non-CLIL groups.

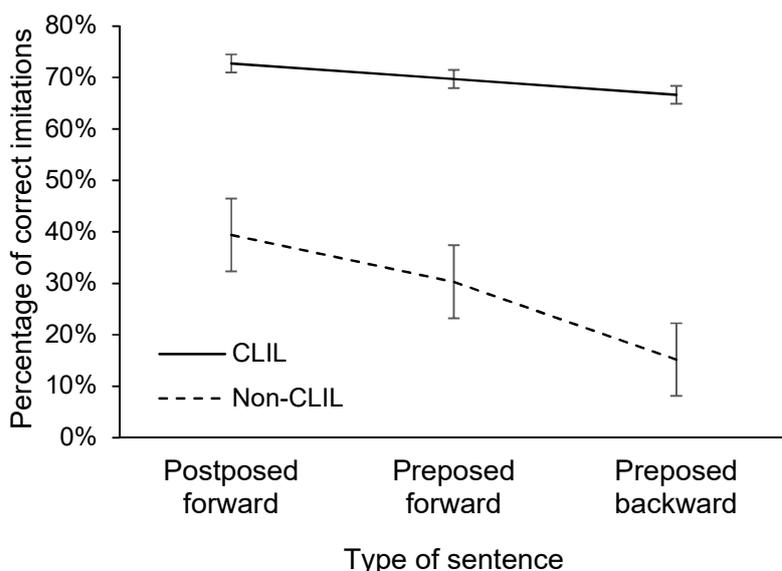


Figure 5. Percentage of each type of sentence correctly imitated by the CLIL and the non-CLIL groups.

In addition, even though the intervention by type interaction was not significant ( $F(2, 40) = 0.97, p = 0.39$ ), a pairwise comparison was performed to test the hypothesis that the difference between complex sentences with pre-posed clauses (backward anaphora) and complex sentences with post-posed clauses (forward anaphora) would only be significant for the non-CLIL group because they are at an earlier stage in the development of the language. As predicted, only students in the non-CLIL group differed significantly in the number of correct

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

imitations for the two types of sentence ( $t(40) = -2.58, p = 0.04$ ). The fact that no significant differences were found between the two types of sentences with pre-posed subordinate clauses ( $t(42) = 1.37, p = 0.18$ ) or the two types of sentences with forward anaphora ( $t(42) = 0.91, p = 0.37$ ) confirms that anaphora direction or clause position alone do not have an effect on the number of correct imitations.

In sum, students in the CLIL group produced significantly more correct imitations in the EI task than their counterparts in the non-CLIL class. In addition, students in the non-CLIL group produced significantly fewer sentences with pre-posed clauses and backward anaphora than sentences with post-posed clauses and forward anaphora. This difference was not observed for students in the CLIL group, which confirms our prediction that CLIL students were able to produce all types of sentences because their English grammar is more advanced.

### **5.4. Correlation between proficiency and syntactic development**

In order to assess the relationship between the scores obtained by students in both tasks, a Pearson product-moment correlation coefficient was obtained to measure the strength of the linear association between the proficiency scores and the EI scores. There was a strong positive correlation between the results of the proficiency test and the results of the overall EI task for the CLIL group,  $r(9) = 0.71, p < 0.01$ . However, the correlation was not significant for the non-CLIL students ( $r(9) = 0.28, p = 0.39$ ).

### **5.5. Overall difference between CLIL and non-CLIL students**

Students' responses to the linguistic and demographic questionnaire did not reveal significant differences between groups; the participants were native speakers of Spanish who started learning English at the same age in school. Students did not have significantly different grades in the first semester of high school. There were no significant differences between groups in SES or years in extracurricular English classes prior to enrollment in the program. There was a statistically significant effect of the intervention (CLIL) on the overall score of the

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

proficiency test and the total number of correct imitations (EI task). Students in the CLIL group outscored their counterparts in the non-CLIL group on the total score of the proficiency test and each of its parts (reading comprehension, use of English and written production). In addition, whereas students in the CLIL group were able to repeat a similar rate of sentences with pre-posed clauses (backward anaphora) and sentences with post-posed clauses (forward anaphora), students in the non-CLIL group were weaker at sentences with pre-posed clauses (backward anaphora). Collectively, these results imply that the CLIL participation has an effect on the students' proficiency and grammatical development. Finally, a strong positive correlation was found between the scores of both tests for the CLIL group, suggesting that both are measuring the students' knowledge of the English language.

### CHAPTER 6: Discussion

The goal of the present research was to test the hypothesis that the CLIL approach enhances the students' level of English as an L2. As predicted, we have provided evidence that students enrolled in the CLIL group outscore their counterparts in the non-CLIL group both in a proficiency test which measures performance in reading comprehension, use of English, and written production, and in an EI task which more directly taps into grammatical knowledge. Simultaneously, we have advanced a methodology to evaluate the language grammar of students enrolled in a CLIL group by designing an EI task based on the acquisition of adverbial subordinate clauses in English.

This is the first study to successfully use the EI task based on a syntactic developmental pattern as an assessment method. Järvinen (2005) used an EI in which CLIL and non-CLIL students had to repeat a set of complex sentences with relative clauses of varying levels of difficulty. However, the mean score obtained by the two groups of students in each type of sentence did not cohere with the hypothesized difficulty of the sentences. In the present experiment, the constraints of PBD on the acquisition of the L2 syntax have provided a useful

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

pattern to measure the acquisition of syntax by Spanish students. Students in the CLIL group produced a higher total of correct imitations and showed a more advanced knowledge of the English syntax than their peers in the non-CLIL group. It is possible that an interaction by type interaction was not found due to the sample size. However, a pairwise t-test revealed a significant difference between sentences with pre-posed clauses (backward anaphora) and sentences with post-posed clauses (forward anaphora) only for students in the non-CLIL group.

As we explained above, proficiency tends to be influenced by extra-linguistic factors, such as pedagogical intervention, and so it may change across different CLIL programs. In turn, an assessment of the grammar with the EI task produces generalizable results about the context-free effect of the CLIL approach on the L2 acquisition. The strong correlation between tests for the bilingual group suggests that the EI task in conjunction with an experimental design may be sufficient to evaluate the effectiveness of CLIL programs. The low level of students in the non-CLIL group could explain the non-significant correlation between the two tasks. Given that on average non-CLIL students did not pass the proficiency test and only repeated 40% of the sentences in the EI task, these tests are likely too advanced to adequately evaluate the English level of these students. Either complemented with a proficiency test or on its own, the EI task can be used to test language knowledge by consulting developmental patterns uncovered by studies on L2 acquisition. Future research could evaluate different aspects of the students' grammar by developing batteries of sentences that adequately measured L2 acquisition. Differences between the first, the second, or even the L3 grammar would need to be taken into account in the creation of the stimuli. In sum, assessing the acquisition of a foreign language would become an interdisciplinary endeavor that integrates L2 teaching, linguistics, and experimental psychology to broaden teaching and learning effectiveness.

Results of this experiment have limitations and raise the necessity for further research. Having conducted the study only in a high school necessarily limits the generalizability of the

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

results. Although there is a possibility that the differences are due to specific characteristics of this particular high school, the fact that the regional government established the same guidelines and provided the same materials for all schools in Andalusia makes our results relevant for the most populated region in the country. Thus, we have shown the context-specific gains of the approach for English proficiency and competence in Andalusia, but future studies should extend this experiment to other CLIL schools located in different areas of the Spanish territory (other cities and towns) with students of different SES.

Additionally, this study shares the shortcomings of any quasi-experimental design. By assuming the absence of additional confounders, we take the risk of ignoring a common cause that is responsible both for the students' decision to enter a CLIL program and the results obtained. L2 proficiency does not only depend on the students' contact with the language (as we measured in the questionnaire), but on other factors such as motivation (Clément, 1980). If motivation had influenced the students' decision to enroll in the CLIL program and their outcomes in the test, it would become a potential confounder in our experiment. An ideal longitudinal study would have allowed us to obtain pre-intervention scores on English proficiency and motivation for L2 learning.

Notwithstanding these caveats, the results of this study have implications for the educational system in Spain. Given the position of Spain in the European language rankings, we need to find ways of improving the system of foreign language teaching. The study shows the benefits for the CLIL group, but there is cause for concern about the non-CLIL group. At the time of testing, almost none of the students in the non-CLIL group would have passed the English test that is part of the Spanish University entrance examination. The EI task and the experimental design have shown that the difference between groups may be partially attributed the students' command of the English grammar. This research may serve as a call of attention

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

for policymakers, teachers and researchers to look into the impact of CLIL not only on the students who enroll in the program, but also on the students who remain in non-CLIL classes.

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

### REFERENCES

- Bruton, A. (2011a). Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore (2010). *Applied Linguistics*, 32(2), 236–241.
- Bruton, A. (2011b). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, 39(4), 523–532.
- Bruton, A. (2013). CLIL: Some of the reasons why ... and why not. *System*, 41(3), 587–597.
- 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.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass.: M. I. T. Press.
- Chomsky, N. (1980). *Rules and Representations*. New York, N.Y.: Columbia University Press.
- Coyle, D., Hood, P., & Marsh, D. (2010). A window on CLIL. In *CLIL: Content and Language Integrated Learning*. Cambridge, England; New York, N.Y.: Cambridge University Press.
- Dalton-Puffer, C. (2008). Outcomes and Processes in Content and Language Integrated Learning (CLIL): Current Research from Europe. In L. V. W. Delanoy (Ed.), *Future Perspectives for English Language Teaching* (pp. 139–157). Heidelberg, Germany: Carl Winter.
- DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. Kroll & A. DeGroot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 88–108). Oxford, England: Oxford University Press.
- Directorate General for Education and Culture. (2006). *Special Eurobarometer. Europeans and their Languages* (pp. 1–65).

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

- Clément, R., C, R., & Smythe, P. C. (1980). Social and individual factors in second language acquisition. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, 12(4), 293–302.
- Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Universal Grammar and second language acquisition: The null hypothesis. *Behavioral and Brain Sciences*, 19(4), 746–758.
- European Commission/EACEA/Eurydice. (2015). *The European Higher Education Area in 2015: Bologna Process. Implementation Report*. Luxembourg: Publications Office of the European Union.
- Feinberg, R. C. (2002). *Bilingual Education: A Reference Handbook*. Santa Barbara, Calif.; Denver, Colo.; Oxford, England: ABC-CLIO.
- Flynn, S. (1983). *A Study of the Effects of Principal Branching Direction in Second Language Acquisition: the Generalization of a Parameter of Universal Grammar from First to Second Language Acquisition*. Cornell University, Ithaca.
- Flynn, S. (1986). Production vs. Comprehension: Differences in Underlying Competences. *Studies in Second Language Acquisition*, 8(2), 135–164.
- Flynn, S., & Espinal, I. (1985). Head-initial/head-final parameter in adult Chinese L2 acquisition of English. *Second Language Research*, 1(2), 93–117.
- Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The Cumulative-Enhancement Model for Language Acquisition: Comparing Adults' and Children's Patterns of Development in First, Second and Third Language Acquisition of Relative Clauses. *International Journal of Multilingualism*, 1(1), 3–16.
- Francis, N. (2011). *Bilingual Competence and Bilingual Proficiency in Child Development*. Cumberland, R.I.: M.I.T. Press.

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

- Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1–56.
- 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 educational environments* (pp. 139–161). Berlin, Germany: Springer.
- Heras, A., & Lasagabaster, D. (2015). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*, 19(1), 70–88.
- Housen, A., & Kuiken, F. (2009). Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30(4), 461–473.
- Järvinen, H. M. (2005). Language Learning in Content Based Instruction. In A. Housen & M. Pierrard (Eds.), *Investigations in Instructed Second Language Acquisition*. Berlin, Germany; Boston, Mass: De Gruyter Mouton.
- Jiménez-Catalán, R. M., & Ruiz de Zarobe, Y. (2009). The receptive vocabulary of EFL learners in two instructional contexts: CLIL vs. Non-CLIL. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe*. (pp. 81–92). Bristol, England: Multilingual Matters.
- Kayi, H. (2006). Teaching speaking: activities to promote speaking in a second language. *Internet TESL Journal*, 12(11).
- Kim, A., Park, A., & Lust, B. (2016). Simultaneous vs. successive bilingualism among preschool-aged children: a study of four-year-old Korean–English bilinguals in the USA. *International Journal of Bilingual Education and Bilingualism*, 1–15.

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

- Lasagabaster, D. (2008). Foreign Language Competence in Content and Language Integrated. *The Open Applied Linguistics Journal*, 1(1), 30–41.
- Lasagabaster, D., & Sierra, J. M. (2010). Immersion and CLIL in English: more differences than similarities. *ELT Journal*, 64(4), 367–375.
- Lázaro Ibarrola, A. (2012). Faster and further morphosyntactic development of CLIL vs. EFL Basque-Spanish bilinguals learning English in high-school. *International Journal of English Studies*, 12(1), 79+.
- Lorenzo, F., Casal, S., & Moore, P. (2010). 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.
- Lust, B. (1981). Constraints on anaphora in child language: A prediction for a universal. In S. Tavakolian (Ed.), *Language acquisition and linguistic theory* (pp. 74–96). Cambridge, Mass.: M.I.T. Press.
- Lust, B. (2011). Acquisition of Language. In P. Hogan (Ed.), *Cambridge Encyclopedia of the Language Sciences*. (pp. 56–64). Cambridge, England: Cambridge University Press.
- Lust, B. (2012). Tracking universals requires grammatical mapping. In *Linguists of Tomorrow: Selected Papers from the 1st Cyprus Postgraduate Student Conference in Theoretical and Applied Linguistics* (p. 105). Newcastle upon Tyne, England: Cambridge Scholars Publishing.
- Lust, B., Bhatia, T., Gair, J., Sharma, V., & Khare, J. (1995). Children's acquisition of Hindi anaphora: A parameter-setting paradox. In V. Ghambir (Ed.), *Teaching and acquisition of South Asian languages* (pp. 172–189). USA: University of Pennsylvania Press.

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

- Lust, B., Flynn, S., & Foley, C. (1996). What Children Know about What They Say: Elicited Imitation as a Research Method for Assessing Children's Syntax. In D. McDaniel, C. McKee, & H. Smith Cairns (Eds.), *Methods for Assessing Children's Syntax* (pp. 55–76). Cambridge, Mass.: M.I.T. Press.
- Marsh, D. (2002). *Information Bulletin 9: The relevance and potential of content and language integrated learning (CLIL) for achieving MT+2 in Europe*. Jyväskylä Iliopisto, FI.
- Martínez Adrián, M., & Gutiérrez Mangado, J. (2009). The acquisition of English syntax by CLIL learners in the Basque Country. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 176–196). Bristol, England: Multilingual Matters.
- Moreno Espinosa, S. (2009). Young learners' L2 word association responses in two different learning contexts. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 93–111).
- Navés, T. (2011). How promising are the results of integrating content and language for EFL writing and overall EFL proficiency? In Y. Ruiz de Zarobe, J. M. Sierra, & J. M. Gallardo del Puerto (Eds.), *Content and foreign language integrated learning: Contributions to multilingualism in European contexts* (pp. 155–186). Bern, Switzerland: Peter Lang.
- Navés, T., & Victori, M. (2010). CLIL in Catalonia: an overview of research studies. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds.), *CLIL in Spain: Implementation, results and teacher training* (pp. 30–54). Newcastle upon Tyne, England: Cambridge Scholars Publishing.

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

- OECD. (2014). *PISA 2012 Results: What Students Know and Can Do Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014)*. PISA: OECD Publishing.
- Paran, A. (2013). Content and Language Integrated Learning: Panacea or Policy Borrowing Myth? *Applied Linguistics Review*, 4(2), 317–342.
- 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.
- Pérez-Vidal, C., & Roquet, H. (2015). CLIL in context: Profiling language abilities. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 237–255). Berlin, Germany: Springer.
- Pinner, R.S. Authenticity of Purpose: CLIL as a way to Bring Meaning and Motivation into EFL Contexts. *Asian EFL Journal* 15 (4), 138-160.
- Prieto-Arranz, J. I., Rallo Fabra, L., Calafat-Ripoll, C., & Catrain-González, M. (2015). Testing progress on receptive skills in CLIL and non-CLIL contexts. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 123–137). Berlin, Germany: Springer.
- Ruiz de Zarobe, Y. (2008). CLIL and Foreign Language Learning: A Longitudinal Study in the Basque Country. *International CLIL Research Journal*, 1(1), 60–73.
- 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* (pp. 191–209). Amsterdam, Netherlands; Philadelphia, Pa.: John Benjamins.

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

- Ruiz de Zarobe, Y., & Lasagabaster, D. (2010). The emergence of CLIL in Spain: An educational challenge. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training* (pp. ix–xvi). Newcastle upon Tyne, England: Cambridge Scholars Publishing.
- Ruiz de Zarobe, Y., & Zenotz, V. (2012). *Estrategias de lectura en CLIL*. Presented at the International Symposium on Practical Approaches in CLIL, Pamplona, Spain.
- Ruiz de Zarobe, Y., & Zenotz, V. (2015). Reading strategies and CLIL: the effect of training in formal instruction. *The Language Learning Journal*, 43(3), 319–333.
- TNS Opinion & Social. (2012). *Europeans and their Languages*. (No. Special Eurobarometer 386.).
- Villareal, I., & García Mayo, M. P. (2009). Tense and agreement morphology in the interlanguage of Basque/Spanish Bilinguals. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 157–175). Bristol, England: Multilingual Matters.

# CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

## APPENDICES

### APPENDIX I

Language questionnaire.

1. What languages do you speak and/or understand (including your native language)?

When and where did you learn them? (For language(s) that you heard from birth, please write "0 (zero)" under age.)

Languages	Age when you began learning this language	Place of exposure
Language 1		
Language 2		
Language 3		
Language 4		

2. Indicate if you had content classes in languages other than Spanish in any of the following educational levels.

Grades	Language	Participantt
1° de Primaria		
2° de Primaria		
3° de Primaria		
4° de Primaria		
5° de Primaria		
6° de Primaria		
1° de E.S.O.		
2° de E.S.O.		
3° de E.S.O.		
4° de E.S.O.		
1° de Bachillerato.		

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

3. Have you attended extra-curricular lessons of English as a foreign language? If you have, for how many months/years? How many hours a week?
4. Have lived in a foreign country for more than three months?

Country	Length of stay	Language

5. Do you use English outside school? If you do, report, when, where and how many hours every week (do not count extra-curricular lessons of English as a Foreign Language).
6. What was your GPA last semester? Please, tell us how many As, Bs, Cs, Ds and Fs you obtained.
7. What was your GPA in the first semester of the first year of high school?
8. What were your parents' highest academic achievement when you started high school?

FATHER OR LEGAL GUARDIAN	MOTHER OR LEGAL GUARDIAN
Primary <input type="checkbox"/>	Primary <input type="checkbox"/>
Secondary <input type="checkbox"/>	Secondary <input type="checkbox"/>
Professional training (intermediate) <input type="checkbox"/>	Professional training (intermediate) <input type="checkbox"/>
Professional training (advanced) <input type="checkbox"/>	Professional training (advanced) <input type="checkbox"/>
Associate degree <input type="checkbox"/>	Associate degree <input type="checkbox"/>
Bachelor's degree <input type="checkbox"/>	Bachelor's degree <input type="checkbox"/>
Master's degree <input type="checkbox"/>	Master's degree <input type="checkbox"/>
PhD <input type="checkbox"/>	PhD <input type="checkbox"/>

9. What was your father's job when you started high school?

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

10. What is your father current job?

11. What was your mother's job when you started high school?

12. What is your mother current job?

APPENDIX II

Proficiency Test

**WOMEN WARRIORS**

It is estimated that some six hundred women served during the American Civil War. They had signed up disguised as men. Hollywood has missed a significant chapter of cultural history here – or is this history ideologically too difficult to deal with? Historians have often found it hard to deal with women who do not respect gender distinctions, and nowhere is gender distinction more strongly affected than in the question of armed female fighters.

But from antiquity to modern times, there are many stories of female warriors. The best known find their way into the history books as warrior queens, rulers and leaders. They have been forced to act as any Churchill or Roosevelt: Semiramis, queen of Assyria, and Boadicea, to mention just two. Semiramis is said to have conquered Ethiopia, Egypt and much of Asia. Among her many great achievements, some also mention the building of Babylon and of its legendary Hanging Gardens, although other authors claim that the gardens had not been built in her time but long after her reign.

Boadicea was a Celtic queen who revolted against the Roman invaders of her territory after they had robbed, beaten and sold into slavery many of her people. Her story, narrated by Roman historiographer Tacitus, was nearly forgotten for many centuries and became popular during the reign of another English queen who headed an army against foreign invasion, Queen Elizabeth I. Boadicea is honoured with a statue on the Thames at Westminster Bridge, right opposite Big Ben.

Although history is quite reticent about women having significant roles in battle either as warriors or leaders, hardly a war has been fought without women soldiers in the ranks.

**I. Comprehension**

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

CHOOSE AND WRITE THE CORRECT OPTION (A, B, C or D). (0.75 points each)

1. Historians find the subject of women warriors to be a difficult topic because ...
  - a. gender distinctions do not exist today.
  - b. women warriors show no respect.
  - c. women who fight go beyond the traditional female role.
  - d. women warriors are stronger than male warriors.
  
2. Boadicea...
  - a. invaded Roman land and turned Romans into slaves.
  - b. reacted against Roman abuse on her people.
  - c. led an army against the invasion of Queen Elizabeth I.
  - d. had a statue made by Tacitus.

ARE THESE STATEMENTS TRUE OR FALSE? JUSTIFY YOUR ANSWERS WITH THE PRECISE WORDS OR PHRASES FROM THE TEXT (0.75 points each)

3. Hollywood movies have been often inspired by historical women warriors.
4. Many wars have been fought with no women soldiers in their armies.

### II. Use of English.

5. (0.25 points) FIND IN THE TEXT ONE WORD THAT HAS THE FOLLOWING DEFINITION: "to be in charge of, to lead others"
6. (0.25 points) FILL IN THE GAP WITH A CORRECT FORM OF THE VERB IN BRACKETS: "She let them ..... (win) the war."
7. (0.25 points) GIVE A NOUN WITH THE SAME ROOT AS "lose" (verb).
8. (0.25 points) WHICH WORD IS NOT AN ADJECTIVE? happy/ trendy/ hardly / lucky
9. (0.25 points) FIND IN THE TEXT ONE SYNONYM FOR "relevant" (adjective).
10. (0.25 points) FILL IN THE GAP WITH THE CORRECT OPTION: "I never go to the gym .... the week" at / during / in / on

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

11. (0.5 points) USE THE WORDS IN THE BOXES TO MAKE A MEANINGFUL SENTENCE. USE ALL AND ONLY THE WORDS IN THE BOXES WITHOUT CHANGING THEIR FORM:

better	Men	considered	than	always	women	have	been
--------	-----	------------	------	--------	-------	------	------

12. (0.5 points) TURN THE FOLLOWING SENTENCE INTO THE PASSIVE VOICE: “Two builders are repairing the broken wall.”
13. GIVE A QUESTION FOR THE UNDERLINED WORDS: Six hundred women served during the American Civil War.
14. JOIN THE FOLLOWING SENTENCES USING A RELATIVE. MAKE CHANGES IF NECESSARY. “That’s the travel agent’s. I booked my holiday tickets there.”
15. COMPLETE THE FOLLOWING CONDITIONAL SENTENCE: “If women had ruled the world, ...”

### III. Production.

16. WRITE A COMPOSITION OF APPROXIMATELY 100 WORDS ABOUT THE TOPIC PROPOSED (3 points). YOU MUST FOCUS STRICTLY ON IT: Do you think women and men have the same opportunities in life?

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

### APPENDIX III

#### Elicited Imitation task

##### Pre-posed/backward anaphora:

1. When he entered the office, the professor questioned the man.
2. When he delivered the message, the man questioned the lawyer.
3. When he prepared the breakfast, the doctor called the professor.

##### Pre-posed/forward anaphora:

1. When the doctor received the results, he called the gentleman.
2. When the lawyer delivered the plans, he answered the worker.
3. When the professor opened the package, he answered the man.

##### Post-posed/forward anaphora:

1. The man answered the boss when he installed the television.
2. The mayor questioned the president when he entered the room.
3. The man introduced the actor when he delivered the plans.

## CONTENT AND LANGUAGE INTEGRATED LEARNING IN ANDALUSIA

### APPENDIX IV

#### ELICITED IMITATION TASK: CODING CRITERIA

- In order for an elicited imitation response to be scored correct, the participant has to repeat the stimulus sentence given without any major syntactic or semantic change. Lexical substitutions that do not imply a substantial change in the meaning of the sentence will not be considered a semantic error. For example, the substitution of gentleman for man will not constitute an error, but the substitution of actor for lawyer will.
- Incidental changes made by the participant which do not alter the syntactic structure or meaning in ways that are relevant to the focus of this research, such as singular to plural or present tense to past tense, will not be scored as incorrect, if this is all that is wrong with the imitation.
- Mispronunciations will not be scored as incorrect.
- Changes altering the syntactic structure or meaning of the original stimulus sentences (e.g., repetition of only one clause, alteration of the original anaphor/antecedent relation, change in clause order, or repetition of a lexical item not considered to be a synonym) will be scored as incorrect.