

THE INFLUENCE OF CLIL ON RECEPTIVE VOCABULARY: A PRELIMINARY STUDY

KEVIN IGLESIAS DIÉGUEZ

University of the Basque Country (UPV/EHU)

kevinigle94@gmail.com

MARÍA MARTÍNEZ-ADRIÁN

University of the Basque Country (UPV/EHU)

maria.martineza@ehu.eus

ABSTRACT. *Empirical studies have shown that Content and Language Integrated Learning (CLIL) seems to be beneficial to receptive vocabulary, which in turn correlates with a higher level of general competence. However, these studies have mainly compared CLIL and Non-CLIL groups matching in age at testing and without a control of other variables such as amount of exposure. The present study, even though exploratory in nature, sets out to fill this gap by comparing groups with the same onset age as well as controlling for other variables. To test general proficiency, the Quick Placement Test (QPT) was used, and the 1,000 and 2,000 frequency bands of the Vocabulary Levels Test (VLT) were delivered to measure functional vocabulary size. CLIL students were found to outstrip their respective Non-CLIL counterparts at the same educational level and to perform as well as an older Non-CLIL sample. Taking together the level of English language lessons and differences in cognitive maturity and amounts of exposure, it is argued that CLIL instruction has intrinsic benefits for receptive vocabulary.*

Keywords: Content and Language Integrated Learning (CLIL), receptive vocabulary, general proficiency, Third language (L3) English.

LA INFLUENCIA DEL AICLE EN EL VOCABULARIO RECEPTIVO: UN ESTUDIO PRELIMINAR

RESUMEN. *Diversos estudios empíricos sugieren que el Aprendizaje Integrado de Contenidos y Lenguas Extranjeras (AICLE) es beneficioso para el desarrollo del vocabulario receptivo, que a su vez se correlaciona con una mejor competencia general. Sin embargo, estos estudios han comparado mayoritariamente grupos AICLE y No-AICLE pertenecientes al mismo año de instrucción sin un control de otras variables como cantidad de exposición. El presente estudio, de naturaleza exploratoria, pretende superar limitaciones previas comparando grupos con la misma edad de inicio de aprendizaje, además de controlar otras variables. Para evaluar la competencia general, se administró la primera parte del Quick Placement Test (QPT), y las versiones del Vocabulary Levels Test (VLT) correspondientes a las 1.000 y 2.000 palabras más frecuentes para medir el vocabulario receptivo. Los alumnos AICLE obtuvieron mejores resultados que sus homólogos No-AICLE en el mismo curso, resultados similares a una muestra No-AICLE de mayor edad. Teniendo en cuenta el nivel de las clases de inglés y las diferencias en madurez cognitiva y exposición, se argumenta que la instrucción AICLE posee beneficios intrínsecos para el vocabulario receptivo.*

Palabras clave: Aprendizaje Integrado de Contenidos y Lenguas Extranjeras (AICLE), vocabulario receptivo, proficiencia general, inglés como tercera lengua (L3).

Received 23 December 2016

Revised version accepted 22 March 2017

1. INTRODUCTION

Over the last two or three decades, research in Second Language Acquisition (SLA) has experienced a boom in two relatively new areas: vocabulary knowledge as an important part of linguistic competence and Content and Language Integrated Learning (CLIL) as a new type of instruction of a Foreign Language (FL).

Although lexical competence does not guarantee high communicative proficiency, it is a fundamental pillar of language use, which in turn facilitates communication (Nation 1993; Meara 1996; Nation and Waring 1997). The oft-cited image of acquirers carrying dictionaries instead of grammar books (Krashen 1989) is far from being mere anecdotal evidence. Empirical research has found strong positive correlations between vocabulary and the so-called “passive skills” of reading and listening. Laufer (1992) shows strong positive correlations between two vocabulary tests and a reading test. Being part of a large project called

DIALANG, Alderson (2005) reports similar coefficients between different aspects of lexical competence and reading and listening comprehension, which are in accordance with Qian (1999) and Nemati (2010). Although correlations do not point to a cause-effect relationship, it is important to point out that there is always a rather strong correlation between vocabulary and passive comprehension.

Regarding CLIL, a vast amount of research has been devoted to clarifying whether the type of instruction has any effect on the acquisition of a FL. As Ellis (1994: 17) points out, this line of research has been motivated by “a desire to address issues of general theoretical interest to SLA research and also by a desire to improve the efficacy of language pedagogy”. Whereas the main focus has been traditionally given to the study of grammar acquisition, it has been only recently that attention is being paid to the effect of CLIL on vocabulary learning, arguably because of its importance in achieving communication.

However, much of the research has centred only on giving a measurement of the receptive vocabulary size of second language (L2) students. For this reason, studies dealing with the effect of CLIL over traditional teaching methods are still scarce, especially in secondary education. Such studies have mainly compared CLIL and Non-CLIL groups matching in age at testing, which means that other factors could explain the variation found. The present paper will contribute to filling this gap by carrying out a pseudo-longitudinal study in which we compare the size of receptive vocabulary of learners in 1st and 3rd year of Compulsory Secondary Education¹ in two different instructional contexts, namely CLIL and Non-CLIL, while controlling for several variables such as onset age and the number of hours of exposure.

To this end, the paper is organised as follows. Firstly, a brief description of what CLIL is and how it has been put into practice in Europe and in the Basque Country will be provided. Subsequently, previous studies that have been carried out in the field of vocabulary acquisition both in CLIL and Non-CLIL contexts will be reviewed. Finally, the results will be reported together with a discussion of their possible causes and implications.

2. CLIL

Since the mid-90's, there has been a growing concern in the European Union regarding people's ability to communicate in a language that is not their mother tongue. As Ruiz de Zarobe (2008) notes, this interest arises from the need to create a more inclusive and integrative society, mainly as a way to cope with

¹ In Spain, Compulsory Secondary Education is known as *Educación Secundaria Obligatoria* (ESO), and post-secondary education is referred to as *Bachillerato*. Throughout this paper, the terms *ESO* and *Bachillerato* will be used for these educational stages.

a multiethnic reality. However, the European Commission's (June 2012) report shows that most people consider learning a new language beneficial for work or study-related prospects since these answers occupy 4 out of the 5 most chosen options, with "understanding people from other countries" ranking sixth. The pursuit of different objectives, together with little guidance from European institutions (Ruiz de Zarobe 2013), has caused the CLIL type of instruction to be implemented in different ways depending not only on the country, but also on the region and individual ventures, as well as other contextual factors that "influence both their aims and outcomes" (Nikula, Dalton-Puffer and Llinares 2013: 72). As a consequence, CLIL is often used as an umbrella term to describe any 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: 183)².

In the Basque Country, CLIL has been recently implemented on top of the existing 3-model system for bilingual education. Model A offers all subjects in Spanish, Model B teaches half the subjects through Spanish and the other half through Basque, and Model D is entirely carried out in Basque, with the exception of language lessons in all cases. However, schools differ on several factors which make CLIL implementation rather heterogeneous: starting age for CLIL instruction, number of subjects and hours offered in English, and electiveness of such subjects, among others.

Generally, CLIL pedagogy is characterised by a more student-centred approach, as opposed to the traditional teacher-centred one, with the focus on students' participation and interactions using the target language in an attempt to develop their communicative competence. This is achieved by providing them with comprehensible input in addition to a more "natural" context for acquisition and encouraging interaction on the students' part. However, the culture of the CLIL classroom is still that of the first language (L1), and as Ruiz de Zarobe (2013: 237) notes, "the teachers' pragmatic use of the language is sometimes less varied than in the teaching of subjects in the L1", depending on teacher's proficiency in the target language.

Moreover, teachers of CLIL subjects are not language teachers and concentrate mainly on content rather than form (Navés 2009). For this reason, research suggests that whilst general proficiency is improved, specific aspects of language do not seem to behave in the same way. As for general proficiency, Ruiz de Zarobe

² The term *Content-Based Instruction* (CBI), among others, is also used for the kind of instruction described here. Although each term is associated with its historical genesis, their actual current pedagogies do not differ to such an extent so as to consider them different (Cenoz, Genesee and Gorter 2014; Cenoz 2015).

(2008) compared the performance of CLIL and Non-CLIL students at the end of Secondary and Post-Secondary Education in written and oral production. She reports significant differences in the four aspects of general proficiency tested in favour of CLIL learners. Lasagabaster (2008) found that CLIL students outperformed a sample of Non-CLIL learners matching in number of hours of exposure and another sample matching in grade in terms of speaking, listening, grammar and writing tasks. However, benefits of CLIL do not seem to extend to some specific areas of language such as morphosyntax. García Mayo and Villarreal Olaizola (2010) report no significant differences between students in a CLIL setting and learners in traditional instruction in different tense and agreement morphemes. Similarly, Martínez Adrián and Gutiérrez Mangado (2015) analysed the general proficiency and the production of tense and agreement morphemes in a group of CLIL students, a group of Non-CLIL students matching in year of instruction and hours of exposure, and another group with the same number of hours of exposure but different age at testing. They found that CLIL learners could perform as well as older students and that they outstripped students of the same age but with fewer hours of exposure when tested for general proficiency. Regarding specific aspects of morphosyntax, CLIL students are reported to have obtained similar results to the group with fewer hours of exposure and significantly poorer scores than older learners. In order to solve difficulties found in these aspects, focus-on-form has been proposed by several researchers (García Mayo 2011; Ruiz de Zarobe and Lasagabaster 2010).

It still remains to be seen whether the trend found in specific areas of grammar also occurs in receptive vocabulary, a limitation we tackle in the present paper.

3. STUDIES ON RECEPTIVE VOCABULARY

The study of receptive vocabulary has been mainly done through the Vocabulary Levels Test (VLT) given its acknowledged validity and reliability (Schmitt, Schmitt and Clapham 2001; King and Fulcher 2007).

These studies have generally tested the effect of an ample variety of variables such as the level of motivation; age at which first exposure took place; age at the time of testing; the effects of maturity and memory; and the type of instruction, which revolves around the influence of CLIL on vocabulary over traditional EFL teaching. The last variable has gained special importance in the last decade since CLIL projects are increasingly being implemented in schools throughout Europe. Nonetheless, these studies are more limited when compared to Non-CLIL in the sense that they are aimed at finding vocabulary size estimates, with subjects having received different hours of exposure (both inside and outside the classroom) and having started learning English at different ages.

In this respect, studies in Non-CLIL contexts abound, especially in the last years of Primary Education. Jiménez Catalán and Terrazas (2005-2008) report 4th graders' receptive vocabulary to be around 737 words after 419 hours of instruction in English. Terrazas and Agustín (2009) found a much lower estimate of 361 words for 4th graders after the same amount of exposure. In the same study, the size for students' vocabulary in 5th, 6th Primary and 1st ESO was calculated at 509, 631 and 817 words, respectively. Agustín and Terrazas (2012) conducted a cross-sectional study in all the grades between 4th Primary and 3rd ESO. They report similar results to those in Terrazas and Agustín (2009), while the estimates for 2nd and 3rd ESO are 987 and 1206 words. Canga (2013) analysed lexical knowledge of 4th ESO students and results show a mean of 935 words after 1049 hours of instruction, which is a poorer score than 2nd and 3rd ESO students' in Agustín and Terrazas (2012). In an in-depth investigation about vocabulary tests, López-Mezquita (2005) carried out a not-so-controlled study of students in 4th ESO and 1st and 2nd Bachillerato. As for the first group, results point to a knowledge of 941 words, similar to Canga's (2013) but considerably lower than Agustín and Terrazas (2012). This may be due to the inclusion of students who had failed the subject of English language in previous years but had passed on to the next educational level nonetheless, as well as learners who were in curricular diversification programs with much lower standards than typical 4th ESO groups. It should be noted that no variable was controlled for with the exception of grade. In similar circumstances, estimates for 1st and 2nd Bachillerato amount to 1582 and 1885 words, respectively.

As far as the CLIL variable is concerned, studies on lexical knowledge are scarce mainly due to the difficulty of finding homogeneous groups of subjects. Jiménez Catalán and Ruiz de Zarobe (2009) compared 6th grade students in CLIL and Non-CLIL contexts, with the latter receiving 331 hours of exposure less than the former. Moreover, the CLIL group was composed of only bilinguals whereas the latter were exclusively monolinguals. For the CLIL group, the size of their vocabulary is estimated at 748 words, and for the Non-CLIL, at 602 words. In a similar study, Fernández Fontecha (2014a) set up a group of CLIL students in 5th Primary and another one of Non-CLIL learners in 2nd ESO, both of whom had received approximately 839 hours of instruction. Results show that the Non-CLIL group outperformed the CLIL group: 985 words for the former and 705 for the latter. Canga (2015) compared the scores of three groups: two 6th-grade samples (one with CLIL and the other one with traditional teaching) and one formed by 2nd ESO students in a traditional Non-CLIL context. He reports slightly higher scores for the secondary Non-CLIL group, although statistical analysis shows that this difference is not significant. A summary of the results, together with the variables considered, is shown in Table 1.

Table 1. Summary of word estimates for different grades in vocabulary research.

Study	Year/Grade	CLIL?	Hours of exposure	Vocabulary size
Jiménez Catalán and Terrazas (2005-2008)	4 th Primary	N	419	737
Terrazas and Agustín (2009)	4 th Primary	N	419	361
	5 th Primary	N	524	509
	6 th Primary	N	629	631
	1 st ESO	N	734	817
Agustín and Terrazas (2012)	4 th Primary	N	419	361
	5 th Primary	N	524	527
	6 th Primary	N	629	663
	1 st ESO	N	734	836
	2 nd ESO	N	839	987
	3 rd ESO	N	944	1206
Canga (2013)	4 th ESO	N	1049	935
López-Mezquita (2005)	4 th ESO	N	-	941
	1 st Bachillerato	N	-	1582
	2 nd Bachillerato	N	-	1885
Jiménez Catalán and Ruiz de Zarobe (2009)	6 th Primary	N	629	602
	6 th Primary	Y	960	748
Fernández Fontecha (2014a)	5 th Primary	Y	839	705
	2 nd ESO	N	839	985
Canga (2015)	6 th Primary	N	629	601
	6 th Primary	Y	944	903
	4 th ESO	N	1049	936

As can be seen from López-Mezquita (2005), caution should be taken since insightful comparisons and estimations can only be made when the variables are controlled for; otherwise, results will vary wildly. Apart from the study conducted by López-Mezquita (2005), the rest of the investigations are better-designed in this respect but are nonetheless limited in that they do not take into consideration the rate of acquisition rather than the end result. To our knowledge, no studies have been conducted so far which shed light on whether the CLIL type of instruction is intrinsically beneficial for vocabulary size in addition to allowing a greater number of hours of exposure. The present study purports to fill this gap by comparing CLIL and Non-CLIL subjects who have started learning English at the same age, have received similar amounts of exposure and have not taken any extracurricular activity in English (cf. section 5.1).

4. RESEARCH QUESTIONS

The present paper aims at overcoming the limitations of previous studies and pointing towards a clearer answer to those questions that have been raised in the literature. Firstly, we aim to check the existence of a correlation between receptive vocabulary and general proficiency, as has been previously reported (Qian 1999; Nemati 2010). In addition to confirming the relationship between these two aspects, this study also serves the purpose of assessing the representativeness of the sample. Secondly, the results of CLIL and Non-CLIL groups with the same age at testing will be compared to ascertain the academic success and effectiveness of CLIL in receptive vocabulary. Since previous studies have mainly focused on the end result (Jiménez Catalán and Ruiz de Zarobe 2009; Canga 2015), this paper includes a pseudo-longitudinal analysis to reach more insightful conclusions. Finally, we compare a CLIL and a Non-CLIL group with the same onset age and number of hours of exposure but a difference in testing age to assess the impact of CLIL on receptive lexical knowledge, overcoming the limitations found in previous research done in CLIL settings, and similar in design to Lasagabaster's (2008) assessment of the relationship between CLIL and general proficiency. In short, this study addresses the following questions:

1. Is there a relationship between receptive vocabulary and general proficiency?
2. Do students in a CLIL context outperform their counterparts in a traditional EFL classroom at the same educational level?
3. Do CLIL students outperform older Non-CLIL students when they have been exposed to the same number of hours and all other factors are held constant?

5. METHOD

5.1. PARTICIPANTS

The sample consisted of 55 Basque-Spanish bilingual learners of third language (L3) English from three different schools, of which two are state-funded (school A and B) and the other private (school C). These schools are located in the Basque Country where Basque is the language of instruction for all subjects except for Spanish and English language courses. The context in which the subjects are immersed has been defined as additive trilingualism (Cenoz and Valencia 1994), where Basque, the language of instruction, is a minority language of Spain. Spanish is the majority language, and English is taught as a foreign language. It is the case that some learners have Basque and Spanish as their L1s, others have Basque as L1 and Spanish as L2, while a third set of learners has Spanish as their L1 and Basque as their L2. In all cases, the additive context in which these learners live leads to balanced bilingualism. The participants are middle-class students with a

very similar socioeconomic status. The socioeconomic status of the participants was established by means of a questionnaire where the participants described their parents' educational level and profession. None of the subjects attended an academy or any extracurricular activities related to English, nor had they had stays in English-speaking countries. They had not either gone through a selection process during the course of their studies.

Participants were divided into four groups considering their type of instruction and their current year of instruction, which determines the number of hours of exposure³: (a) a Non-CLIL 1 group (n=10) of 12 year-olds in 1st ESO; (b) a CLIL 1 group (n=15) with the same age as the previous group but more hours of exposure; (c) a Non-CLIL 2 group (n=15) of 14 year-olds in 3rd ESO with a similar number of hours of exposure to the CLIL 1 group; and (d) a CLIL 2 group (n=15) with students of the same age as the Non-CLIL 2 group but more hours of exposure. Only participants who started learning English at the age of 3 have been included in the sample. In doing so, this study purports to overcome the limitations that have arisen in previous studies dealing with the effect of CLIL instruction on receptive vocabulary due to the lack of a match between the number of hours of exposure and the onset age. Participants' characteristics are displayed in Table 2.

Table 2. Participants' characteristics.

Group	Onset Age	Mean age at testing (range)	Length of exposure (in years)	Hours of Exposure	Gender	School of origin
Non-CLIL 1 (n=10)	3	12.4 (12-13)	9	972	6 male/4 female	A
CLIL 1 (n=15)	3	12.1 (12-13)	9	1,116	8 male/7 female	C
Non-CLIL 2 (n=15)	3	14.2 (14-15)	11	1,173	10 male/5 female	B
CLIL 2 (n=15)	3	14.1 (14-15)	11	1,451	9 male/6 female	C

³ The variable gender has not been taken into consideration when dividing the participants into groups since differences between males' and females' learning behaviours in lexical learning seem to be test-dependent (Sunderland 2010), with only small differences arising at some stages due to psychological changes characteristic of puberty and motivational factors related to adolescence (Agustín and Terrazas 2012).

Non-CLIL groups had 2 hours of English a week during their first three years of formal education and 3 hours a week of EFL lessons in Primary and Secondary Education. In addition to these hours of formal instruction of English, the CLIL groups had 1 hour a week of CLIL in Social Sciences and Creative Arts during the last 3 years of Primary Education and throughout Secondary Education. As a result, the CLIL 1 group had a total of 4 years of CLIL instruction, whereas the CLIL 2 group were exposed to CLIL for 6 years. In all groups, the materials and approach used in the English lessons were the same, with the occasional use of Spanish or Basque when needed.

5.2. INSTRUMENTS

Data were gathered by means of three instruments. To measure the participants' general proficiency, the first part of the QPT (version 1) was used. Part 2 was not handed out since it corresponds to proficiency levels of mastery⁴, which are beyond the scope of Secondary Education. This test has been extensively used in SLA research to assess general proficiency (López-Mezquita 2005; Martínez Adrián and Gutiérrez Mangado 2015). In addition, two different VLTs were used, namely the 1,000 and 2,000 frequency bands, to measure the size of students' receptive vocabulary (Appendices 1 and 2). The 1,000 VLT consisted in translation of words to avoid difficulties arising from not understanding the definitions rather than the target vocabulary items⁵. The 2,000 VLT is a slightly modified version of the test developed by Schmitt, Schmitt and Clapham (2001) that has been previously used in Fernández Fontecha (2014a). Tests for both frequency bands were used with the purpose of getting an insightful understanding of students' functional vocabulary regarding the 1,000 and 2,000 most frequent words, since scoring at least 15 (max=30) in the 2,000 VLT is claimed to show that students master the whole 1,000 most frequent words, whilst this may not always be the case. Although students are reported to learn the most frequent words first (Read 1988), they may also possess some significant word knowledge pertaining to the 2,000 frequency band while not mastering the previous 1,000⁶. These tests have been empirically proved to be reliable and valid as a measure of the intended functional vocabulary (Schmitt, Schmitt and Clapham 2001; Xing and Fulcher 2007) and are widely used in vocabulary research (Qian 2002; Jiménez and Terrazas 2005-2008; Terrazas and Agustín 2009).

⁴ C1 and C2 levels according to the CEFR scale (https://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf).

⁵ It has been translated by the GLAUR research group based at the University of La Rioja under the supervision and approval of Paul Nation. This version has also been used by other researchers at the University of the Basque Country (see Martínez Adrián and Gallardo del Puerto 2010).

⁶ Lower frequency bands (5,000 and 10,000) have not been used since less frequent words hardly ever appear in Secondary Education textbooks or class materials, including CLIL, where difficulty of vocabulary is kept to a minimum to facilitate content learning, according to teachers.

5.3. PROCEDURE

All tests were done in one session during class time, except for a background questionnaire, which students were asked to complete at home with their parents. They were told that the results of these tests would not affect in any way their marks in English or any other subject, and were also told to miss out any item to which they did not know the answer. For each test, they were given clear instructions, together with an example, both in written form and orally in Spanish to clarify what they were being asked to do.

They were first given the QPT, to be completed in 30 minutes, followed by the VLTs in order of frequency band, for which the time allotted was 10 minutes each.

Total scores and vocabulary size estimates were obtained. To this end, Nation's (1990: 78) formula was applied. Individual data were entered into SPSS for descriptive and inferential statistical analysis. The results of the Kolmogorov-Smirnov tests revealed that all groups were normally distributed in all tests. Since they complied with the normality assumption, Pearson's correlation coefficient was calculated between the QPT and the VLTs. Independent samples t-test was also implemented to check for any significant differences between the groups' means.

6. RESULTS AND DISCUSSION

6.1. RESEARCH QUESTION I

The first research question aimed to ascertain whether a relationship existed between receptive vocabulary and general proficiency. The results for the correlations are shown in Table 3 below.

In all groups, the 2,000 VLT strongly correlated with the QPT at a significant level, which suggests that vocabulary plays an essential role in proficiency. These figures (.731-.856) agree with previous studies that have purported to answer this question (Nemati 2010; Qian 1999).

In the case of the present study, a correlation was expected to a certain extent, since some of the questions in the QPT tap specifically on vocabulary knowledge. However, such strong coefficients point to a great relevance of vocabulary for other parts of general proficiency than simply the vocabulary compartment, highlighting the importance of lexical knowledge in successful communication as proficiency in English increases.

Moreover, by including the 1,000 VLT, interesting results arise. Its correlation with the QPT is lower than that of the 2,000 VLT (coefficients ranging from .520-.659), and the relationship between these two tests is rather weak and

non-significant in the Non-CLIL 1 group. Lower coefficients may suggest that knowledge of the 1,000 most frequent words does not affect general proficiency to the extent that the next frequency band does. This seems to signal that knowledge of vocabulary from different frequency bands has a differing impact on general proficiency. For this reason, caution should be taken when reporting and interpreting results in these correlations between receptive vocabulary and general proficiency when only one test is used for lexical knowledge. By and large, knowledge of the 1,000 most frequent words seems to have a lesser effect on general proficiency than knowledge of the next 1,000.

Table 3. Pearson Correlation Coefficients for the QPT and VLTs by group (two-tailed).

		QPT	1,000 VLT	2,000 VLT
Non-CLIL 1	QPT	1		
	1,000 VLT	0.198	1	
	2,000 VLT	.811**	.279	1
CLIL 1	QPT	1		
	1,000 VLT	.659**	1	
	2,000 VLT	.856**	.666**	1
Non-CLIL 2	QPT	1		
	1,000 VLT	.520*	1	
	2,000 VLT	.731**	.758**	1
CLIL 2	QPT	1		
	1,000 VLT	.562*	1	
	2,000 VLT	.762**	.49	1

* $p < .05$ ** $p < .01$

As for the impact of CLIL, it seems that this type of instruction slightly increases the correlation between general proficiency and receptive vocabulary knowledge, suggesting that a bigger size of receptive vocabulary relates to greater proficiency. However, these data should be submitted to further statistical analysis to check for any significant difference in the Pearson coefficients between groups that only differ in this variable.

All in all, functional receptive vocabulary seems to be part of general proficiency. In this respect, CLIL in this paper is argued to increase this type of lexical knowledge when compared to traditional EFL. Since the main purpose of the educational system is to allow students to communicate in a foreign language, this type of instruction

should be implemented in all schools if it really improves receptive vocabulary knowledge as reported. This is analysed in the next research question.

6.2. RESEARCH QUESTION II

The second research question was concerned with the differences in terms of amount of exposure between the CLIL groups and their Non-CLIL counterparts at the same year of instruction (cf. Table 2). These differences amount to 144 hours in the case of the CLIL 1 and Non-CLIL 1 groups, whereas the CLIL 2 and Non-CLIL 2 groups differ in 278 hours.

The results obtained by the Non-CLIL 1 and CLIL 1 groups in the QPT are shown in Table 4:

Table 4. Descriptive and inferential statistics for the QPT in 1st ESO.

	Non-CLIL 1	CLIL 1	t
Mean	13.30	18.47	
SD	2.58	3.48	
Min	10	14	-4.004
Max	18	26	(p=.001)
Range	8	12	

The Non-CLIL 1 group got a mean score 13.30 in the QPT, which according to the score guide, corresponds to a beginner's level (A1). This is rather disturbing since higher mastery is expected after 972 hours of exposure and learning. The CLIL 1 group obtained a higher mean of 18.47, which signals that students have achieved an elementary level (A2) after 1,116 hours of exposure. Inferential statistics showed that the difference in general proficiency was significant ($p=.001$) in favour of CLIL learners, which agrees with previous studies (Lasagabaster 2008; Ruiz de Zarobe 2008; Martínez Adrián and Gutiérrez Mangado 2015).

Regarding the results obtained in the VLTs (see Tables 5 and 6), the means were also higher for CLIL students. In the 1,000 VLT, the mean for the Non-CLIL students was 19.40, which yields a vocabulary estimate of 647 words. In the 2,000, students averaged 9.73 points, which using Nation's (1990) formula gives a receptive vocabulary size of 648 words out of 2,000. This estimate is lower than those found by Agustín and Terrazas (2012) and Jiménez Catalán and Ruiz

de Zarobe (2009) for students with approximately the same number of hours of exposure⁷. The proximity of both estimates suggests that participants had hardly acquired any words from the 2,000 most frequent words that do not pertain to first 1,000. As far the CLIL group is concerned, they obtained a mean score of 24.13 points in the 1,000 VLT, corresponding to a receptive vocabulary of 804 words. According to the results of the 2,000 VLT, this group has a functional vocabulary of 1,035 words since the average score is 15.53. This means that students have already mastered the majority of the 1,000 most frequent words and have acquired slightly over 200 of the next frequency band. Independent-samples t-tests showed that the CLIL group outperformed the Non-CLIL group at a significant level in 1st ESO ($p=.000$).

Table 5. Descriptive and inferential statistics for the 1,000 VLT in 1st ESO.

	Non-CLIL 1	CLIL 1	t
Mean	19.40	24.13	
SD	3.86	1.81	
Min	12	21	-4.143
Max	24	27	($p=.000$)
Range	12	6	

Table 6. Descriptive and inferential statistics for the 2,000 VLT in 1st ESO.

	Non-CLIL 1	CLIL 1	t
Mean	9.73	15.53	
SD	2.68	4.03	
Min	2	10	-5.520
Max	12	25	($p=.000$)
Range	10	15	

As for the Non-CLIL2 and CLIL2 groups (aged 14), the results obtained in the QPT are shown in Table 7.

⁷ As no observation data were gathered in these two studies which could show differences in the type of input received, any explanations for the different performance found would remain tentative and speculative.

Table 7. Descriptive and inferential statistics for the QPT in 3rd ESO.

	Non-CLIL 2	CLIL 2	t
Mean	19.93	23.33	
SD	3.13	4.19	
Min	15	16	-2.520
Max	25	34	(p=.018)
Range	10	18	

As can be observed, at the moment of testing, students in the Non-CLIL group had achieved an elementary level (A2) with a mean of 19.93 points, whereas the CLIL group was in the threshold between elementary and lower intermediate (B1) with 23.33 points on average. As in the previous case with 12-year-olds (groups Non-CLIL1 and CLIL1), the CLIL group performed significantly better than the Non-CLIL group ($p=.018$).

In the VLTs (Tables 8 and 9), students in the CLIL group also obtained higher scores. The Non-CLIL group averaged knowledge of 796 words out of the 1,000 most frequent and 1,146 words from the 2,000 most frequent ones, which means that students have not fully acquired understanding of all of the 1,000 most frequent ones but have nonetheless demonstrated knowledge of 350 words belonging to the lower frequency band. The CLIL group, on the other hand, achieved the slightly higher score of 25.93 points in the 1,000 VLT and 20.87 in the 2,000 VLT, which stand for 864 and 1,391 words, respectively. Inferential analyses showed that the difference in the knowledge of the higher frequency band was not significant (although a statistical tendency was found: $p=.060$), and that students in the CLIL group had significantly larger vocabularies ($p=.009$).

Table 8. Descriptive and inferential statistics for the 1,000 VLT in 3rd ESO.

	Non-CLIL 2	CLIL 2	t
Mean	23.87	25.93	
SD	3.82	1.44	
Min	18	23	-1.963
Max	29	29	(p=.060)
Range	11	6	

Table 9. Descriptive and inferential statistics for the 2,000 VLT in 3rd ESO.

	Non-CLIL 2	CLIL 2	t
Mean	17.20	20.87	
SD	4.16	2.92	
Min	9	15	-2.792
Max	23	27	(p=.009)
Range	14	12	

In almost all cases, learners in a CLIL context outperformed their Non-CLIL counterparts to a significant extent, with the only exception of the 1,000 VLT in the 14-year-old groups. Although the difference for this frequency band was only tendential ($p=.060$), the CLIL 2 group were found to know on average over 60 words more than their Non-CLIL 2 counterparts. In the light of these findings, the CLIL type of instruction seems to help consolidate knowledge of the 1,000 frequency band.

All in all, CLIL instruction does indeed help grow a bigger receptive vocabulary, as well as improve general proficiency. In 1st ESO there is a sharp difference of almost 400 words between their functional vocabularies. Such disparity may stem from the decontextualised use of English in traditional EFL lessons. In this type of instruction, textbooks tend to present vocabulary items that are essential to the requirements of the curriculum leaving out other less frequent words; hence, it is very likely that students may not have been very much exposed to words that are beyond the 1,000 most frequent ones (Alcaraz Mármol 2009) since learning words in isolation does not require the need for words that are not specifically intended to be learnt, specially taking into consideration that teachers rarely divert from the limited vocabulary presented in textbooks. On the other hand, CLIL instruction necessarily makes use of words of higher frequency bands since words in the 1,000 frequency band are not sufficient to convey the required content in CLIL subjects. Moreover, these may in turn reinforce the vocabulary to be learnt in the English class, since repetition of occurrence seems to play a key role in vocabulary learning (Saragi, Nation and Meister 1978; Webb 2007). In other words, the most frequent and thus repeated words appear both in the traditional English classroom (mostly decontextualised) and in the other subjects taught through English (in context).

However, a pseudo-longitudinal analysis of the results reveals that CLIL students do not make the most of the greater number of hours of exposure. In other words, they seem to have a lower rate of acquisition. In the Non-CLIL groups, students in 3rd of ESO know 149 more words among the 1,000 most frequent ones and 498 more belonging to the next frequency band in 201 hours, which respectively

averages 74 and 247 words per 100 hours of exposure to the target language. In the CLIL groups, these differences amount to a total of 60 words for the lower frequency band and 356 for the next one in 335 hours, with an average of 18 and 106 words per 100 hours of exposure.

This seems to point to CLIL not only being ineffective but rather slowing down vocabulary acquisition. However, there are two reasons that may explain these results. Firstly, the study is not entirely longitudinal, meaning that the subjects tested in the CLIL 1 group are not the same as those in the Non-CLIL 2 group. Since three schools are involved in this study, they may have contributed unequally to the conformation of the groups, which may show up and magnify slight differences in teaching that possibly make this small sample not wholly representative. For more solid grounds, a longitudinal study is warranted.

Secondly, CLIL students are not as exposed to other words other than those appearing in the English class as could be expected. A qualitative analysis of end-of-degree projects (TFGs) and Master's Dissertations dealing with CLIL didactic units has revealed that the vocabulary used almost entirely coincides with the vocabulary presented in the English language textbooks for the same year of education⁸ (e.g. Lázaro Gómez 2013; Calvario Pérez 2014). This suggests that the difference in vocabulary between CLIL and Non-CLIL students is caused by incidental learning of vocabulary used in extra material provided by the teacher and class dynamics (such as group discussions) rather than explicit learning. Considering that English classes are the same for both CLIL and Non-CLIL students, the effect of CLIL can be deemed remarkable given the little attention paid to increasing receptive vocabulary. Generally, it seems that CLIL subjects reinforce the vocabulary used in the English class by allowing the repetition of these words in addition to present students with a few more words that they learn incidentally.

Nonetheless, this apparent effectiveness of CLIL instruction as regards receptive vocabulary may also be attributable to the greater number of hours of exposure to the target language. It still remains to be seen whether students in Non-CLIL settings would achieve the same results if they had an additional hour of English instruction per week (and no other subject taught through English). Since the Government decides on the hours needed for each subject, this kind of study is rendered impossible. Therefore, the only approach that may succeed in unveiling the effect of type of instruction on receptive vocabulary is to find subjects that have started learning English at the same age and have been exposed to the same number of hours, notwithstanding the difference in testing age. This is the aim of the third research question.

⁸ This is especially noticeable in those exercises that focus on vocabulary, where the difference between words being asked for in a traditional-English-class exercise and a CLIL subject is virtually nonexistent.

6.3. RESEARCH QUESTION III

This question aimed at clarifying whether CLIL instruction was beneficial for receptive vocabulary knowledge by overcoming the limitations found in previous studies. As already stated, subjects had been exposed to the English language for the same amount of time and have started learning English at the same age but differ in age at testing. In turn, any difference between the groups' performance, or lack thereof, could arguably be attributed to the type of instruction. Thus, the CLIL 1 group was compared with the Non-CLIL 2. All students had first been exposed to the English language at the age of 3 and, in spite of having a different age at the time of testing, had received similar hours of English instruction, either explicitly (language lessons) or through content subjects (Social Sciences and Creative Arts). More specifically, both groups shared 972 hours of formal instruction in English, with the remaining difference amounting to 144 hours of CLIL subjects in the case of the CLIL 1 group, and additional 201 hours of formal instruction in English in the Non-CLIL 2 group. Descriptive statistics for these groups can be seen in Tables 4-9, whereas word estimates and relevant inferential statistics are shown in Tables 10, 11 and 12 below⁹.

Table 10. Descriptive and inferential statistics for the QPT.

	CLIL 1	Non-CLIL 2	t
Mean	18.47	19.93	
SD	3.48	3.13	-1.214
Min	14	15	(p=.235)
Max	26	25	
Range	12	10	

Table 11. Word estimates and inferential statistics for the 1,000 VLT.

	CLIL 1	Non-CLIL 2	t
Mean	804	796	
SD	30	127	-.245
Min	700	600	(p=.808)
Max	900	966	
Range	200	366	

⁹ Since the score of the QPT does not have a specific meaning beyond assessing proficiency level in a 6-category scale, descriptive statistics are repeated for easy reference, together with the result of the t-test.

Table 12. Word estimates and inferential statistics for the 2,000 VLT.

	CLIL 1	Non-CLIL 2	t
Mean	1035	1,146	
SD	259	277	-1.114
Min	666	600	(p=.275)
Max	1666	1533	
Range	1000	933	

For the three tests administered, no significant differences were found between the two groups, which indicates that CLIL does not pose any threat to lexical knowledge nor general proficiency. However, a more in-depth analysis and discussion of the results will suggest that the CLIL type of instruction has more benefits than simply allotting more hours to English in fewer years of academic study¹⁰.

As regards general proficiency, both groups have obtained similar results, although the Non-CLIL 2 group has scored slightly higher. However, only the first part of the QPT was administered, which means that students have not been assessed on listening, writing or speaking skills. Had these tasks been included, it is likely that the CLIL group could have outperformed the Non-CLIL one since the cornerstone of CLIL instruction is participation and interaction, rather than more controlled activities. The nature of the instrument used has thus conditioned the results to some extent, as previous research has found compelling evidence that shows CLIL learners can perform as well as or even better than traditional EFL students with the same numbers of hours of exposure (Lasagabaster 2008; Ruiz de Zarobe 2010; Martínez Adrián and Gutiérrez Mangado 2015).

As for receptive vocabulary within the 1,000 most frequent words, CLIL and Non-CLIL students seem to be on equal grounds since the difference is not significant. As argued below, this could have resulted from the application of the same standards in vocabulary to the same educational level irrespective of whether students are taking more subjects in English.

In the 2,000 frequency band, the Non-CLIL 2 group scored on average slightly higher, although CLIL students had higher minimum and maximum. These results become remarkable if we take into account that the CLIL group has been exposed to

¹⁰ It is still not clear whether the better performance of the CLIL group is due to CLIL instruction or the intensity of the L3 exposure implicit in CLIL. However, the nature of the study, as well as contextual limitations, makes it impossible to tease apart intensity and CLIL instruction (e.g. Martínez Adrián and Gutiérrez Mangado 2015).

English for 57 hours less, approximately half a year of formal English instruction. The following discussion points to several advantages stemming from CLIL instruction rather than a higher number of hours of exposure in traditional EFL teaching.

The CLIL educational approach is described as allowing a more naturalistic and contextual learning of the language by focusing on content, much like required if learning happened in an English-speaking country. However, research initially suggested that context does not play a role in vocabulary learning. Seibert's (1930) longitudinal study, with tests delivered after one hour, two, ten and forty days found that students learning word pairs consistently outperformed students working with words in context. In the same vein, Gershman (1970) reports a non-significant difference between word pair and contextual learning. These empirical studies used a very narrow account of context, providing words in a sentence or with a drawing. More recent research uses context in much broader terms, encompassing and "simulating" CLIL situations. Coady (1997: 286) carried out a survey of previous research and reached the conclusion that "if the language is authentic, rich in content, enjoyable, and, above all, comprehensible, then learning is more successful", which coincides with the CLIL environment in subject matter lessons. More research seems to support this claim (Nagy 1995; Webb 2008).

Lexical knowledge also seems to be bolstered by incidental learning, since attention is drawn to content rather than to vocabulary items (Vidal 2011). The results of the present study show that CLIL learners have acquired almost the same number of words as the Non-CLIL students have, even though the former have received 57 hours less of exposure than the Non-CLIL group. Considering that traditional teaching draws their attention to vocabulary items of higher frequency bands (since less frequent words are presented in subsequent years of academic study), we entertain that CLIL learners could have performed better if vocabulary standards in English lessons were raised and lexical complexity in CLIL subjects were not so limited (or at least increased steadily). Furthermore, it would possibly be beneficial to implement incidental learning in English lessons for both CLIL and Non-CLIL students.

In addition to providing a real context for English learning and use, CLIL has been shown to increase students' motivation levels, which in turn facilitate vocabulary acquisition (Fernández Fontecha 2014b). However, the actual extent to which instruction affects motivation highly depends on the subject matter: whereas studies in Physical Education have found no significant differences between CLIL and Non-CLIL students (Heras and Lasagabaster 2015), research in other subjects report a significantly higher level of motivation in CLIL students (Lasagabaster and Sierra 2009; Lasagabaster 2011). The roots of this motivational growth are related to the meaningful and teleological use of language, since students in a traditional EFL

setting find some exercises boring, unrealistic and non-significant (Lasagabaster and Sierra 2009: 13). Hence, CLIL learners are more prone to acquiring and increasing their vocabulary.

All in all, CLIL seems to be a promising approach for the growth of functional receptive vocabulary. Results reported here are remarkable since CLIL learners have performed as well as Non-CLIL students, who have received slightly more hours of exposure and have developed greater cognitive maturity.

7. CONCLUSION AND PEDAGOGICAL IMPLICATIONS

With CLIL experiencing a boom in the last decade and communication in a FL becoming essential, the present study set out to clarify the benefits of this type of instruction over traditional EFL teaching as regards receptive vocabulary, for this plays a key role in understanding meaning as well as being significantly related to general proficiency. Comparisons between groups of different characteristics (cf. Table 2) suggest that CLIL students not only outstrip their Non-CLIL counterparts in the same year of instruction but also perform equally well when their results are compared with older learners who have been exposed to the English language for approximately the same amount of time. Although results suggest CLIL students' rate of acquisition is slower, we have proposed that this downside stems from the pseudo-longitudinal nature of the present study (rather than purely longitudinal) and the limitations of the English syllabus. Taking into account the greater number of hours that the Non-CLIL 2 students have received, their intrinsic cognitive maturity and higher complexity presented in their English lessons, we have arguably attributed to CLIL instruction methodological characteristics that favour vocabulary learning: contextual presentation of lexical items; focus on content, which allows incidental learning; and increasing motivation levels by giving language use a communicative purpose.

In this paper, we have also suggested that the benefits of CLIL as far as receptive vocabulary is concerned can be further exploited by including English lessons especially designed to meet the needs of CLIL students. These would take into consideration the further hours of exposure to the target language and the higher number of repetitions available that are needed to learn a word. In consequence, the English curriculum for CLIL students should have higher standards as far as vocabulary is concerned. In addition, CLIL materials should not be so constrained by English lessons standards in this respect and vocabulary of higher complexity should be included progressively.

This study is exploratory in nature and, as such, there are certain shortcomings that need to be considered, among others, the size of the sample and an unequal

distribution of males and females. In fact, this variable has started to receive attention on the part of CLIL researchers (Fernández Fontecha and Canga Alonso 2014; Heras and Lasgabaster 2015; Roquet, Llopis, and Pérez Vidal 2016). Note also that results might not be fully generalisable to other contexts as CLIL implementations are heterogeneous. A follow-up study should include a bigger sample, with the same number of males and females in each group. Moreover, receptive vocabulary has only been assessed on the basis of individual words, hence neglecting formulaic language. A more comprehensive estimate could be obtained by delivering more vocabulary tests, such as the Word Associates Test, Size Test, Eurocentres Vocabulary Test or a test based on the PHRASE and PHaVE lists¹¹. Additionally, it would also be interesting to test the impact of CLIL on students' academic vocabulary as well as productive vocabulary. In addition, further research should also focus on the effect of CLIL on the learning of content.

All in all, CLIL seems to be beneficial both in terms of general proficiency and receptive vocabulary knowledge, and since all learners could benefit from this type of instruction, its implementation should be further encouraged from educational institutions to improve language proficiency.

ACKNOWLEDGMENTS

We would like to thank the three schools that have decided to take part in this study and the teachers that let us intrude in some of their classes. The second author wants to acknowledge the grants awarded by the Spanish Ministry of Economy and Competitiveness (FFI2012-32212; FF12016-74950-P), the University of the Basque Country (UFI 11/06) and the Basque Government (IT904-16).

REFERENCES

- Agustín, M. P. and M. Terrazas. 2012. "Vocabulary knowledge development and gender differences in a second language". *Estudios de Lingüística Inglesa Aplicada* 12: 45-75.
- Alcaraz Mármol, G. 2009. "Vocabulary input in EFL textbooks: Frequency levels." <<http://www.um.es/lacell/aelinco/contenido/pdf/52.pdf>> (Accessed 21 February 2017)
- Alderson, J. C. 2005. *Diagnosing Foreign Language Proficiency*. London: Continuum.
- Calvario Pérez, J. J. 2014. *Elaboration of materials for a Conocimiento del Medio class*. Zaragoza: University of Zaragoza.

¹¹ Accessible at <http://www.norbertschmitt.co.uk/resources.html>.

- Canga, A. 2013. "Receptive vocabulary size of secondary Spanish EFL learners". *Revista de Lingüística y Lenguas Aplicadas* 8: 66-75.
- Canga, A. 2015. "Receptive vocabulary of CLIL and Non-CLIL Primary and Secondary school learners". *Complutense Journal of English Studies* 23: 59-77.
- Cenoz, J. 2015. "Content-Based Instruction and Content and Language Integrated Learning: The same or different?" *Language, Culture and Curriculum* 28 (1): 8-24.
- Cenoz, J., Genesee, F. and D. Gorter. 2014. "Critical analysis of CLIL: Taking stock and looking forward". *Applied Linguistics* 35 (3): 243-262.
- Cenoz, J. and J. Valencia. 1994. "Additive trilingualism: Evidence from the Basque Country". *Applied Psycholinguistics* 15: 195-207.
- Coady, J. 1997. "L2 vocabulary acquisition: A synthesis of the research". *Second Language Vocabulary Acquisition*. Eds. J. Coady and T. Huckin. Cambridge: Cambridge University Press. 273-290.
- Dalton-Puffer, C. 2011. "Content-and-Language Integrated Learning: From practice to principles?" *Annual Review of Applied Linguistics* 31: 182-204.
- Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- European Commission. June, 2012. *Special Eurobarometer 386: Europeans and their languages*. Retrieved from: http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf.
- Fernández Fontecha, A. 2014a. "Receptive vocabulary knowledge and motivation in CLIL and EFL". *Revista de Lingüística y Lenguas Aplicadas* 9: 23-32.
- Fernández Fontecha, A. 2014b. "Motivation and gender effect in receptive vocabulary learning: An exploratory analysis in CLIL primary education". *Latin American Journal of Content and Language Integrated Learning* 7 (2): 27-49.
- Fernández-Fontecha, A. and A. Canga Alonso. 2014. "A preliminary study on motivation and gender in CLIL and non-CLIL types of instruction". *International Journal of English Studies* 14 (1): 21-36.
- García Mayo, M. P. 2011. "The relevance of attention to form in communicative classroom contexts". *Estudios de Lingüística Inglesa Aplicada* 11: 11-45.
- García Mayo, M. P. and I. Villarreal Olaizola. 2010. "The development of suppletive and affixal tense and agreement morphemes in the L3 English of Basque-Spanish bilinguals". *Second Language Research* 27 (1): 129-149.
- Gershman, S. J. 1970. *Foreign language vocabulary learning under seven conditions*. Columbia University: Doctoral thesis.
- Heras, A. and D. Lasagabaster. 2015. "The impact of CLIL on affective factors and vocabulary learning". *Language Teaching Research* 19 (1): 70-88.

- Jiménez Catalán, R. M. and M. Terrazas. 2005-2008. "The receptive vocabulary of English Foreign Language young learners". *Journal of English Studies* 5-6: 173-191.
- Jiménez Catalán, R. M. and Y. Ruiz de Zarobe. 2009. "The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus Non-CLIL instruction". *Content and Language Integrated Learning: Evidence from Research in Europe*. Eds. Y. Ruiz de Zarobe and R. M. Jiménez. Bristol: Multilingual Matters. 81-92.
- Krashen, S. D. 1989. "We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis". *The Modern Language Journal* 73 (4): 440-464.
- Lasagabaster, D. 2008. "Foreign language competence in content and language integrated courses". *The Open Applied Linguistics Journal* 1: 30-41.
- Lasagabaster, D. 2011. "English achievement and student motivation in CLIL and EFL settings". *Innovation in Language Learning and Teaching* 5 (1): 3-18.
- Lasagabaster, D. and J. M. Sierra. 2009. "Language attitudes in CLIL and traditional EFL classes". *International CLIL Research Journal* 1 (2): 4-17.
- Laufer, B. 1992. "How much lexis is necessary for reading comprehension?" *Vocabulary and Applied Linguistics*. Eds. P. J. L. Arnaud and H. Béjoint. London: Macmillan. 126-132.
- Lázaro Gómez, M. C. 2013. *La enseñanza de contenidos en lengua inglesa: Estrategias, técnicas y recursos*. Valladolid: University of Valladolid.
- López-Mezquita, M. T. 2005. *La evaluación de la competencia léxica: Tests de vocabulario. Su fiabilidad y validez*. Universidad de Granada: Doctoral dissertation.
- Martínez Adrián, M. and F. Gallardo del Puerto. 2010. "L2 English receptive and productive vocabulary in senior learners". *ITL, International Journal of Applied Linguistics* 160: 112-131.
- Martínez Adrián, M. and M. J. Gutiérrez Mangado. 2015. "Is CLIL instruction beneficial in terms of general proficiency and specific areas of grammar?" *Journal of Immersion and Content-Based Language Education* 3 (1): 51-76.
- Meara, P. 1996. "The dimensions of lexical competence". *Performance and Competence in Second Language Acquisition*. Eds. G. Brown, K. Malmkjaer and J. Williams. Cambridge: Cambridge University Press. 35-53.
- Nagy, W. E. November. 1995. "On the role of context in first- and second-language vocabulary learning". *Center for the Study of Reading Technical Report* 627. Illinois: University of Illinois at Urbana-Champaign.
- Nation, I. S. P. 1990. *Teaching and Learning Vocabulary*. New York: Newbury.

- Nation, I. S. P. 1993. "Vocabulary size, growth and use". *The Bilingual Lexicon*. Eds. R. Schreuder and B. Weltens. Amsterdam/Philadelphia: John Benjamins. 115-134.
- Nation, I. S. P. and R. Waring. 1997. "Vocabulary size, text coverage, and word lists". *Vocabulary: Description, Acquisition and Pedagogy*. Eds. N. Schmitt and M. McCarthy. Cambridge: Cambridge University Press. 6-19.
- Navés, T. 2009. "Effective Content and Language Integrated Learning (CLIL) Programmes". *Content and Language Integrated Learning: Evidence from Research in Europe*. Eds. Y. Ruiz de Zarobe and R. M. Jiménez. Bristol: Multilingual Matters. 22-40.
- Nemati, A. 2010. "Proficiency and size of receptive vocabulary: Comparing EFL and ESL environments". *International Journal of Educational Research and Technology* 1 (1): 46-53.
- Nikula, T., Dalton-Puffer, C. and A. Llinares. 2013. "CLIL classroom discourse: Research from Europe". *Journal of Immersion and Content-Based Language Education* 1 (1): 70-100.
- Qian, D. D. 1999. "Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension". *Canadian Modern Language Review* 56: 282-308.
- Qian, D. D. 2002. "Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective". *Language Learning* 52 (3): 513-536.
- Read, J. 1988. "Measuring the vocabulary knowledge of second language learners". *RELC Journal* 19 (2): 12-25.
- Roquet, H., Llopis, J. and C. Pérez Vidal. 2016. "Does gender have an impact on the potential benefits learners may achieve in two contexts compared: formal instruction and formal instruction + content and language integrated learning?" *International Journal of Bilingual Education and Bilingualism* 19 (4): 370-386.
- 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". *Language use in CLIL*. Eds. C. Dalton-Puffer, T. Nikula and U. Smit. Amsterdam: John Benjamins. 191-212.
- Ruiz de Zarobe, Y. 2013. "CLIL implementation: From policy-makers to individual initiatives". *International Journal of Bilingual Education and Bilingualism* 16 (3): 231-243.

- Ruiz de Zarobe, Y. and D. Lasagabaster. 2010. "CLIL in a bilingual community: The Basque autonomous region". *CLIL in Spain: Implementation, results and teacher training*. Eds. D. Lasagabaster and Y. Ruiz de Zarobe. Newcastle upon Tyne: Cambridge Scholars Publishing. 12-29.
- Saragi, T., Nation, I. S. P. and G. F. Meister. 1978. "Vocabulary learning and reading". *System* 6: 72-78.
- Schmitt, N., Schmitt, D. and C. Clapham. 2001. "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test". *Language Testing* 18 (1): 55-88.
- Seibert, L. C. 1930. "An experiment on the relative efficiency of studying French vocabulary in associated pairs versus studying French vocabulary in context". *Journal of Education Psychology* 21: 297-314.
- Sunderland, J. 2010. "Theorizing gender perspectives in foreign and second language learning". *Gender Perspectives on Vocabulary in Foreign and Second Languages*. Ed. R. M. Jiménez Catalán. New York: Palgrave Macmillan. 1-19.
- Terrazas, M. and M. P. Agustín. 2009. "Exploring the increase of receptive vocabulary knowledge in the Foreign Language: A longitudinal study". *International Journal of English Studies* 9 (1): 113-133.
- Vidal, K. 2011. "A comparison of the effects of reading and listening on incidental vocabulary acquisition". *Language Learning* 61 (1): 219-258.
- Webb, S. 2007. "The effects of repetition on vocabulary knowledge". *Applied Linguistics* 28 (1): 46-65.
- Webb, S. 2008. "The effects of context on incidental vocabulary learning". *Reading in a Foreign Language* 20 (2): 232-245.
- Xing, P. and G. Fulcher. 2007. "Reliability assessment for two versions of Vocabulary Levels Tests". *System* 35: 182-191.

APPENDIX 1. 1,000 VLT

1,000 WORD LEVEL TEST
 2015/2016

INSTITUTO: _____
CURSO: _____ **FECHA:** _____
APELLIDOS: _____ **NOMBRE:** _____

Este es un test de vocabulario. En la parte izquierda te presentamos grupos de seis palabras inglesas y a su derecha, la traducción en castellano de sólo tres de ellas. **Escribe** junto a cada traducción, el **número** de la palabra inglesa que tenga el mismo significado. Observa el siguiente ejemplo:

EJEMPLO	RESPUESTA CORRECTA
1 dog _____ negro 2 house _____ nariz 3 girl _____ casa 4 fork _____ 5 black _____ 6 nose _____	→
1 could _____ podía/pude 2 during _____ durante 3 this _____ 4 piece _____ para 5 of _____ 6 in order to _____	1 kill _____ avanzar 2 reply _____ responder 3 advance _____ matar 4 appoint _____ 5 divide _____ 6 receive _____
1 indeed _____ mi 2 what _____ en efecto 3 along _____ algo 4 my _____ 5 some _____ 6 away _____	1 moment _____ separado 2 separate _____ momento 3 worse _____ amarillo 4 free _____ 5 heavy _____ 6 yellow _____
1 church _____ coche 2 scene _____ problema 3 hour _____ hecho 4 trouble _____ 5 fact _____ 6 car _____	1 spring _____ hermana 2 danger _____ peligro 3 stone _____ piedra 4 product _____ 5 sister _____ 6 subject _____
1 meet _____ poner 2 leave _____ dar 3 put _____ utilizar 4 give _____ 5 use _____ 6 begin _____	1 example _____ anchura 2 breadth _____ miedo 3 fear _____ ayuntamiento 4 desert _____ 5 bit _____ 6 town hall _____
1 wind _____ hombre 2 room _____ línea 3 line _____ noche 4 enemy _____ 5 night _____ 6 man _____	1 surround _____ quedar bien 2 shoot _____ advertir 3 paint _____ disparar 4 fit _____ 5 command _____ 6 warn _____

APPENDIX 2. 2,000 VLT

2,000 WORD LEVEL TEST
2015/2016

INSTITUTO: _____

CURSO _____ FECHA: _____

APELLIDOS: _____ NOMBRE: _____

Este es un test de vocabulario. En la parte izquierda te presentamos grupos de seis palabras inglesas y a su derecha, los significados de sólo tres de ellas. **Escribe** junto a éstos, el **número** de la palabra inglesa correspondiente a dichos significados. Observa el siguiente ejemplo:

EJEMPLO	→	RESPUESTA CORRECTA
1 business 2 clock _____ part of a house 3 horse _____ animal with 4 legs 4 pencil _____ something used for writing 5 shoe 6 wall	→	1 business 2 clock _____ 6 part of a house 3 horse _____ 3 animal with 4 legs 4 pencil _____ 4 something used for writing 5 shoe 6 wall

1 coffee 2 disease _____ money for work 3 justice _____ a piece of clothing 4 skirt _____ using the law in the right way 5 stage 6 wage	1 adopt 2 climb _____ go up 3 examine _____ look at closely 4 pour _____ be on every side 5 satisfy 6 surround
1 choice 2 crop _____ heat 3 flesh _____ meat 4 salary _____ money paid regularly for doing a job 5 secret 6 temperature	1 bake 2 connect _____ join together 3 inquire _____ walk without purpose 4 limit _____ keep within a certain size 5 recognize 6 wander
1 cap 2 education _____ teaching and learning 3 journey _____ numbers to measure with 4 parent _____ going to a far place 5 scale 6 trick	1 burst 2 concern _____ break open 3 deliver _____ make better 4 fold _____ take something to someone 5 improve 6 urge
1 attack 2 charm _____ gold and silver 3 lack _____ pleasing quality 4 pen _____ not having something 5 shadow 6 treasure	1 original 2 private _____ first 3 royal _____ not public 4 slow _____ all added together 5 sorry 6 total
1 cream 2 factory _____ part of milk 3 nail _____ a lot of money 4 pupil _____ person who is studying 5 sacrifice 6 wealth	1 ancient 2 curious _____ not easy 3 difficult _____ very old 4 entire _____ related to God 5 holy 6 social