



Universidad Pública de Navarra  
Nafarroako Unibertsitate Publikoa

**FACULTAD DE CIENCIAS HUMANAS, SOCIALES  
Y DE LA EDUCACIÓN**

**Máster Universitario de Profesorado de Educación Secundaria**

## **Trabajo de Fin de Máster**

# **ANALYSING CONTENT ACQUISITION IN CLIL VS. NON-CLIL PROGRAMMES**

**Estudiante: Elia Galera Rodríguez**

Tutora: Izaskun Villarreal Olaizola

**Especialidad: Inglés**

**Junio, 2022**

## **ANALYSING CONTENT ACQUISITION IN CLIL VS. NON-CLIL PROGRAMMES**

**ABSTRACT:** Students following CLIL methodologies are believed to achieve higher levels of linguistic competence than mainstream EFL learners but conceptual content level is less clear, since scarce research has been conducted in this area. Moreover, SES and self-efficacy are variables that may affect the student's academic results. To this end, the Social Sciences content level attained, the socio-economic status (SES) and students self-efficacy perceptions were compared across three groups of students: (1) a non-CLIL group (n=14), (2) a soft-CLIL group (n=14) and (3) a hard-CLIL group (n=18). All students took the same Social Sciences test in Spanish and completed two questionnaires regarding their SES and self-efficacy perceptions. Results revealed that CLIL groups slightly outperformed non-CLIL ones in subject performance, suggesting that CLIL methodology is also beneficial for content learning. Furthermore, no significant correlations were found between non-CLIL and CLIL test scores, their SES and self-efficacy levels. Individually, results revealed that levels of self-efficacy were inversely proportional to soft-CLIL scores, and directly proportional to hard-CLIL scores.

*Key words:* CLIL, content acquisition, CALP, EFL, academic performance

**TABLE OF CONTENTS**

1. Introduction.....	4
2. Literature Review.....	5
CLIL benefits.....	5
CLIL and content acquisition.....	6
Socio-economic background and content acquisition outcomes.....	8
Self-efficacy and content acquisition outcomes.....	8
3. Research questions (RQ).....	10
4. Method.....	11
Context and participants.....	11
Instruments, procedure and data collection.....	12
Social Sciences exam.....	12
Socio-economic survey.....	13
Self-efficacy questionnaire.....	15
5. Results.....	16
Social Sciences exam.....	16
Socio-Economic Status.....	18
Self-efficacy levels.....	18
Relationship between CLIL, SES and Self-Efficacy.....	20
6. Discussion.....	20
7. Conclusion, pedagogical implications and further studies.....	23
8. Limitations.....	24
9. Acknowledgments.....	24
Bibliography.....	25
Appendix.....	28

## 1. Introduction

CLIL stands for “content and language integrated learning”. It is a methodology related to all forms of education in which a second or foreign language (L2) is used for teaching and learning specific subjects. This methodology started to develop in Europe in the 1990’s and its use has increased in recent years. The term CLIL comprehends a wide range of school practices (Coyle, 2012).

This methodology focuses mainly in two areas: content and language. Language acquisition and the communicative skills benefit from CLIL methodology, since students are exposed to a large amount of input in the foreign language. This issue can be seen in a wide range of studies, in which language performance has been tested among CLIL and non-CLIL students. CLIL groups tend to achieve better results than their non -CLIL counterparts on language competence (Martínez Agudo, 2020).

Nevertheless, in the area of content acquisition, whether CLIL is advantageous for subject knowledge development has received little attention. Moreover, the existing studies report mixed findings. A certain trend can be seen in which CLIL pupils in lower grade levels - primary students- tend to show lower content achievement levels than non-CLIL students studying in their L1 (Fernández-Sanjurjo et al., 2017). As courses progress, secondary CLIL students gradually outperform non-CLIL ones in content examinations regardless of the language in which they are tested (Elorza & Muñoa, 2008; Hughes & Madrid, 2019). Moreover, CLIL learners not only show better results in the subject content but also demonstrate a greater capacity for reasoning, abstract and critical thinking (Elorza & Muñoa, 2008). Conversely, studies such as Baranova’s (2019) showed that CLIL students in higher grade levels (undergraduate students) achieved worse results than non-CLIL students. Despite this, the differences between the two groups were not significant. The scarce existing investigations therefore suggest more positive content learning outcomes for CLIL students. Yet, there’s still a very narrow body of research to convincingly state that CLIL affects positively content learning and additional research is needed to determine whether the attested benefits hold for any CLIL programme.

On the other hand, apart from the CLIL variable, we cannot disregard other variables that also affect students' academic results. In terms of socio-economic status, studies such as Fernández-Sanjurjo et al.'s (2018) show that students with lower socio-economic status obtain lower scores than those coming from more privileged backgrounds. Furthermore, we find authors such as Bruton (2019) who state that "it is precisely when CLIL is optional that there is implicit selection". Thus, it could be said that due to the difficulty for students to study a subject in a language other than L1, it is students with better academic results who are enrolled

in the first instance in language programmes with this methodology. Due to the aforementioned selection of students, other variables should be taken into account when comparing non-CLIL and CLIL learners, such as “(parental) socio-economic status, content scores, FL scores, overall academic scores/IQ, motivational levels and external exposure to the FL” (Bruton, 2019).

Additionally, another variable that can affect students' academic results is self-efficacy (Ohlberger & Wegner, 2019). Accordingly, the results of Jaekel's (2020) study indicate that it is CLIL learners who report higher levels of self-efficacy in comparison to non-CLIL learners. Yet having other variables in play such as IQ (higher for CLIL students than their non-CLIL counterparts) and socio-economic status (also higher for CLIL students), the self-efficacy variable had the strongest positive effect on students' language proficiency.

Therefore, as will be explained later, it can be said that lower socio-economic backgrounds and low self-efficacy may have a detrimental impact on the academic performance of students (Hughes & Madrid, 2019; Jaekel, 2020).

## **2. Literature Review**

### **CLIL benefits**

According to Lasagabaster (2008), CLIL methodology enhances the competence of Cultural Awareness and Expression, boosts motivation of students in order to study foreign languages and triggers communication between students and teachers, helping to develop all the language skills in a natural way, especially the oral ones.

It has been reported that learning a subject with a foreign language (FL) as the vehicular one, can help students to improve language performance. This is the case of Spain, in which the FL - in our particular case English - has little or no presence outside school, however many studies report that English language competence in students following these programmes improves rapidly, eventually outperforming the students who do not take part in the aforementioned CLIL programmes (Lasagabaster, 2008; Martínez Agudo, 2020). Although different variables such as the socio-economic background, which will be discussed later, play a very important role in language performance, research studies that involve learners with similar socio-economic characteristics tend to favour CLIL as a more efficient way for language acquisition. Conversely, other authors state that CLIL does not provide language learning instructional support, therefore it cannot be guaranteed that this methodology will favour language acquisition (Roussel et al., 2017 as cited in Bruton, 2019).

Related to language acquisition, Cummins' (2008) classic distinction between Basic Interpersonal Communicative Skills (BICS) and Cognitive Academic Language Proficiency

(CALP) has to be taken into account. This distinction refers to the two types of language used in a lesson conducted in a second language, and it was emphasised to draw educators' attention to the evolution of second language learners as they progress through these language programmes (Cummins, 2008).

In the first few years of CLIL programmes, the student is exposed to a large amount of input in a second language, therefore being able to cope with everyday language -BICS-. This language includes the conversational language of everyday classroom life, and some general academic language. It is in older years when students enrolled in CLIL programmes do acquire a better level of a foreign language, but it emerges with the use of concrete language. In higher course levels specialised, abstract and subject-specific vocabulary arises -CALP- (Ball et al., 2015). The fact that CALP acquisition develops years later may suggest that subject-specific language knowledge and thus learning strategies related to academic competence in general arise in higher levels of CLIL classrooms. This specific language, related to more cognitively demanding topics, challenges the learner as it involves a higher cognitive burden (Jaekel, 2020). Not only a better level is achieved as he or she is able to express themselves in a more accurate way, but also the use of this language involves the development of the student's critical and abstract thinking, which is beneficial in general for the academic development. In this way, we find that students who benefit from a CLIL methodology could have advantages derived from language acquisition that could be applied to other fields, including content acquisition.

### **CLIL and content acquisition**

CLIL programmes were created to promote the acquisition of a foreign language and its students were said to attain comparable levels of content knowledge to their non-CLIL counterparts. To date, however, scarce research comparing the subject knowledge of CLIL and non-CLIL students, has been undertaken to be able to convincingly affirm it. Therefore, the improvement in subject knowledge is less clear than in language performance (Fernández-Sanjurjo et al., 2017). Recent studies report mixed results (Baranova, 2019; Fernández-Sanjurjo et al., 2017; Hughes & Madrid 2019; Martínez Agudo, 2020), and students' age or level of instruction seems to be a core variable to determine the success of the CLIL program in content attainment.

In the case of Baranova's (2019) research in which the experiment was conducted with 3<sup>rd</sup> year undergraduate students at university, the control group -non-CLIL students- coped better than the experimental group -CLIL students- regarding their levels of "professional discipline testing" (content knowledge). Despite this, "it is important to emphasise that the difference in the results of the control and experimental groups turned out to be insignificant"

(Baranova, 2019). These results may seem contradictory if we look at the age of the participants, but it is worth mentioning that the participants had not participated in a CLIL programme before, therefore students were at the early stages of CLIL methodology. In this case, a longitudinal experiment would show whether in later years these CLIL students outperform non-CLIL ones, or whether this methodology is detrimental to their knowledge acquisition with significant differences.

Regarding secondary students, an experiment was conducted by Elorza and Muñoa (2008) within a *Social Science and Language Integrated Curriculum* programme in Basque secondary schools. According to the results of both a language competence and subject knowledge assessments, the control group -non-CLIL learners- performed worse than CLIL learners, despite the fact that the content exam was conducted in Basque (L1), the language non-CLIL students used to learn the content. The aspects in which they performed better were the ones in which written production and cognitive demanding skills were needed. These results may suggest that the methodology used in CLIL programmes can be beneficial for students in a wide range of competences (Elorza & Muñoa, 2008).

Conversely, contradictory results were obtained in the study of Fernández-Sanjurjo et al. (2017), in which non-CLIL primary students performed slightly better than CLIL ones in Science content knowledge. Besides, a strong relationship between socio-economic status (SES) and better content performance was reported in this study. Similar findings were also reported by Hughes and Madrid (2019) who showed that although there were no significant differences between non-CLIL and CLIL primary students' content learning outcomes, mean scores were slightly higher for the non-CLIL group. In contrast, in the same study conducted in secondary schools, non-CLIL students were significantly outperformed by non-CLIL students, particularly in public schools (Hughes & Madrid, 2019). These secondary school students, therefore, had been studying through CLIL programmes for a longer period of time. It seems possible that these results are due to the aforementioned later acquisition of CALP. As students grow older, they are able to express themselves better in a second language. Considering the multiple benefits that CLIL programmes can provide for learners (cognitively, socially, academically and professionally), the short-term decline in performance could be understood as a natural part of the development of future plurilingual citizens (Hughes & Madrid, 2019).

Since results between primary, secondary and undergraduate students tend to differ, and very little was found in the literature regarding the question of content acquisition of secondary CLIL students, it can be suggested that this paper contributes to data collection on this issue, providing information of secondary students that benefit from long-term CLIL

programmes. Furthermore, the expected sample of data collection will provide different student's profiles, taking into account three different learning programmes: (1) a non-CLIL programme, (2) a Plurilingual-CLIL programme and (3) a full-immersion-CLIL programme in which the hours of exposure to English are different. In addition, a range of socio-economic status among these students will be found. Moreover, we will consider the variable of self-efficacy as well, which may play an important role in the student's academic results.

### **Socio-economic background and content acquisition outcomes**

As stated above, different variables affect the academic performance of students. One of the clearest and most visible variables is the (parental) socio-economic level (SES) (Bruton, 2019) of the student. SES is a measure of social status that includes factors such as the level of family income, the education of the student's parents or guardians and the resources available to them at home, which have a significant influence on students' ability to achieve higher grades. Studies such as Fernández-Sanjurjo et al.'s (2017) show that students with less favoured socio-economic backgrounds obtain lower academic grades than those with better socio-economic backgrounds.

While it is true that attending CLIL programmes removes first of all the barriers of the mother tongue therefore hosting "a broader range of learners" (Lorenzo et al., 2010 as cited in Bruton, 2013 p.593), there is a tendency for families with higher SES to enrol their children in this type of programme. It is true that studying a CLIL subject through language programmes is optional for all students, but also the selection of these students is common, since a certain level of English is required to be able to keep up with the pace of the class. Therefore, students with a higher SES, with greater opportunities for study such as private tutoring, tend to access these programmes. Accordingly, SES of the students should be taken into account in order to compare the academic results between groups, since the student body may have "radically different profiles at the outset" (Bruton, 2019) depending on the programme in which they are enrolled.

### **Self-efficacy and content acquisition outcomes**

Self-efficacy and self-concept are variables that can affect academic performance, (Bandura, 1997 as cited in Ohbelger & Wegner, 2019). According to Bandura, self-efficacy is the individual's belief in their own competence to achieve a certain objective. This concept has to do with the perception of oneself. In this sense, self-efficacy influences the student's thoughts, motivation and behaviour towards the academic context. In general, students with higher levels of self-efficacy are those who follow self-regulation and study strategies, such as language learning strategies (LLS), which ultimately translate into better academic results



(Donat et al, 2008 as cited in Olhbelger & Wegner, 2019), both in language and content performance.

Levels of self-efficacy evolve over the years, and are reaffirmed by academic success. In relation to language programmes, traditionally students attending CLIL programmes, with the prestige attributed to them (Bruton, 2013), tend to show higher levels of self-efficacy with respect to their non-CLIL counterparts (Jaekel, 2020), as they are assumed to have a higher academic level, better grades and greater chances for higher education. Apart from student's different profiles at the outset (Bruton, 2019), this methodology can help students to develop their own learning strategies (Jaekel, 2020). CLIL is a methodology that encourages communication between students and autonomous work, as well as following a series of beneficial practices for the student's own academic development. An example of this can be the creation of concept maps, a common practice in CLIL programmes, in which the learner works not only on the specific vocabulary but also on the structures he or she will need to express him or herself correctly. These practices help the learner and facilitate the acquisition of content, as well as providing the learner with a powerful LLS that can be used in other academic areas (Jaekel, 2020).

On the other hand, students' stress or anxiety can affect their levels of self-efficacy. Authors such as Jaekel (2020) state that students who are put under more pressure to achieve high academic results may have lower levels of self-efficacy. In this regard, CLIL students generally have to put more effort not only into acquiring the subject content, but also into coping with a FL. Especially in the first years of a CLIL programme, students need to learn a large amount of specific vocabulary, and “need to move rapidly from basic interpersonal communication skills (BICS) to cognitive academic language proficiency (CALP)” (Jaekel, 2020). With regard to lower levels of self-efficacy, students may misjudge tasks as too difficult and may even try to avoid those (Ohlberger & Wegner, 2019).

Bearing in mind the SES and Self-Efficacy variables, the current study examines student's content acquisition of Social Sciences depending on the programme in which students are enrolled (non-CLIL, soft-CLIL or hard-CLIL programmes, which will be explained later).

### 3. Research questions (RQ)

The aim of this study is to assess the level of non-CLIL and CLIL students in the area of Social Sciences comparing three different language programmes (non-CLIL, soft-CLIL and hard-CLIL) in which the hours of exposure to English (FL) and to CLIL methodology differ. SES and Self-Efficacy variables among the programmes will be taken into account to determine whether they are related to academic scores.

**RQ1:** Do students attending non-CLIL, soft-CLIL and hard-CLIL programmes attain parallel levels of content as demonstrated by their results in a content knowledge exam carried out in Spanish (L1)? If not, are there significant differences between non-CLIL and CLIL students? And between soft and hard-CLIL programmes?

**RQ2:** What is the socio-economic level of the students in the three programmes? Are there differences across the programmes?

**RQ3:** Is the socio-economic level of students related to their academic results?

**RQ4:** What is the self-efficacy level of the students in the three programmes? Are there differences across the programmes?

**RQ5:** Is students' self-efficacy related to their academic results?

Regarding RQ1, given the results of Elorza and Muñoa (2008) and Hughes and Madrid (2019), we expected CLIL students to outperform non-CLIL ones, especially in cognitively-demanding questions. Despite this, there was no evidence to anticipate the relative success of the soft and hard-CLIL programmes.

Secondly, in tune with RQ2 and RQ3, we expected the results on socio-economic status to be similar to those of Fernández-Sanjurjo et al. (2018). Comparing the different groups, we expected non-CLIL students to attain lower levels of SES. Despite this, due to the lack of evidence comparing CLIL programmes we could not predict the differences between soft and hard-CLIL groups, yet higher levels of SES were expected in comparison to the non-CLIL group. In addition, there was some evidence to suggest that socio-economic status would affect students' subject knowledge performance (Fernández-Sanjurjo et al., 2018). Learners from more educated backgrounds tend to have advantages when it comes to learning. According to Wells and Chang-Wells (1992) as cited in Ball et al. (2015), "the presence of books, the high status of reading and the use of 'literate' form of family talk between parents and children" does influence students' performance.

As for RQ4 and RQ5, we expected CLIL students to achieve higher levels of self-efficacy, in tune with the results of Jaekel's study (2020). In this study, CLIL students showed higher self-efficacy levels, which had an impact on their language proficiency. In our study, we

expected these results to influence CLIL students' subject knowledge as well. Besides, there was no evidence to predict self-efficacy differences between soft and hard-CLIL groups. Moreover, several studies (e.g. Zimmerman, 1995 p. 213 as cited in Ohlberger & Wegner, 2019) support the idea that self-efficacy plays a role in the academic results of the students. Therefore, we expected students with higher rates of self-efficacy to outperform their low self-efficacy counterparts.

## **4. Method**

### **Context and participants**

This experiment was conducted in IES Basoko, a state-funded secondary school located in a central neighbourhood of Pamplona, Navarre. Its main distinguishing feature is the variety of linguistic programmes it offers: (1) the G/A Model, in which the vehicular language is Spanish; with the option of taking Basque as a second language apart from three hours of English language arts class and German or French as a foreign language. (2) The Plurilingual programme, in which almost half of the subjects are conducted in English, apart from three hours of English language arts class per week and French or German as a foreign language and (3) the British programme, in which all the lessons are conducted in English, apart from five hours of English language arts class per week plus German or French as a foreign language. The experiment was implemented in three classes of 2<sup>nd</sup> of ESO (average age of students 13.7 years old). Hereafter, we will refer to the students of these three programmes as non-CLIL students, soft-CLIL students and hard-CLIL students respectively. The distinction between soft and hard was coined by Ball et al. (2015) and refers to the amount of time devoted to CLIL methodology in the curriculum. In total, students from group (1) receive 10%, group (2) 45% and group (3) 75% of the teaching hours in English. The difference in exposure to English leads to A2+, B1 and B1+ levels of English depending on the linguistic programme, according to the Common European Framework of Reference (CEFR).

This course has been chosen due to several factors. In the first place, both soft-CLIL and hard-CLIL students conduct the subject of Social Sciences in English, which is the subject that is going to be assessed. On the other hand, at first glance students with very different SES attend the programmes, which may rule out elitism associated with CLIL programmes (Bruton, 2013). Finally, the Social Sciences teacher was the same in both soft-CLIL and hard-CLIL programmes, which helped to control individual teaching differences between the CLIL groups. In relation to the contents taught by both teachers, - CLIL and non-CLIL professors-, a certain level of coordination is expected, since all the contents taught are based on the official curriculum. In addition, prior to the content exam, the content seen by the students throughout the course was reviewed with both teachers. They shared what content they had seen in both the CLIL and non-CLIL groups and previous tests taken by the students were revised. Finally,

both teachers and the head of the department agreed on the topic and the questions that would appear in the test, content that had already been explicitly taught in class.

The final groups were composed of (1) 18 non-CLIL students, (2) 18 soft-CLIL students and (3) 24 hard-CLIL students. Thus, the total sample collected consisted of 60 students. Students who were not present on the day of the test were excluded from the study (4 students) as were students whose mother tongue was other than Spanish (2 students) because it was believed that they would not be able to transfer their knowledge to a language they barely know. In addition, a student who joined the school late was also excluded from the study as it was considered that this participant was not on equal conditions regarding the contents seen in class. Finally, the results from 7 students who did not answer the questionnaire or the assessment were also eliminated from the study. So, the final sample size consisted of (1) 14 non-CLIL students, (2) 14 soft-CLIL students and (3) 18 hard-CLIL students, 46 students in total.

### **Instruments, procedure and data collection**

In order to accomplish this study, a mixed methodology was followed. Three different resources were used: a tailored content-subject assessment test, a socio-economic survey and a self-efficacy questionnaire, which were conducted in Spanish, the common language of all the students and in which almost every student holds a native-like level.

### **Social Sciences exam.**

The exam assessing the Social Sciences subject was a test designed together with the two teachers teaching the two groups. According to them, the content included had already been taught in class. The content was taken from the curriculum and the syllabus of the Geography and History subject of the 2<sup>nd</sup> year of ESO. These contents were taken from the Geography section. The type of questions was approved by both teachers and the head of department. The type of questions was chosen in order to make the examination as narrow as possible. A short exam, preferably not lasting the whole hour, was preferred so as not to interfere with the normal flow of the classes.

The test was divided into two parts, and each part was worth five out of 10 points. The first part consisted of five definitions (1 point per definition) which asked about specific concepts seen in the course in both CLIL and non-CLIL programmes. These definitions were related to declarative knowledge of the students, which did not require substantial written production. The other type was more cognitively demanding, with an open-ended question in which students had to produce longer compositions using specific vocabulary and discourse of the subject. This developmental question was divided into three sub-questions, which were worth 1.5, 1 and 2.5 points respectively (Table 1).

**Table 1.** *Social Sciences exam. Translated into English.*

Question	Type	Required Knowledge	Sub-sections
1. Define the following concepts (5pt.)	Definition	Declarative knowledge	<ul style="list-style-type: none"> <li>- Birth rate (1pt.)</li> <li>- Mortality (1pt.)</li> <li>- Population pyramid (1pt.)</li> <li>- Democracy (1pt.)</li> <li>- Republic (1pt.)</li> </ul>
2. Content-development question (5pt.)	Open-ended question	Procedural knowledge	<ul style="list-style-type: none"> <li>- What is population density? (1.5pt.)</li> <li>- What are the most populated cities in the world? (1 pt.)</li> <li>- Why do you think population is unevenly distributed around the world? (2.5pt)</li> </ul>

The assessment was corrected according to the level of the curriculum. A rubric was created to evaluate the exam, based on the usual marking form (Appendix 1) of the teachers. The rubric contained "model" definitions on which the scoring of the students' answers was based. For the development question, which was more open to different answers, the students' reasoning ability was assessed and whether or not their answers were consistent with what they had seen in class (Appendix 2).

### **Socio-economic survey**

The second one was a socio-economic survey based on the one used by Fernández-Sanjurjo et al. (2017). The materials were kindly shared by Dr. Lázaro-Ibarrola, the main researcher from the "Applied Linguistics: Language Acquisition and Language Teaching" research group from the Public University of Navarra.

Students answered questions regarding aspects of their personal, academic and socio-economic cultural dimensions (Appendix 3). From these, mainly two aspects were taken into account: questions related to their socio-economic status and questions related to their self-efficacy. These aspects were chosen because they were considered important variables when comparing academic results between students.

As for rating the SES of the students, 8 questions were chosen, based on the usual questions used to measure students' ESCS (Economic, Social and Cultural Status.) The first seven questions had 5 different options, from the lowest to the highest level of education or from the lowest to the highest employment status. The last one referred to the amount of material goods in the household. This last question consisted of 8 sub-questions with different material goods and the possible answer was either yes or no (Table 2).

To calculate the students' SES, points were assigned from 1 to 5 in ascending order for each of the students' responses (Table 2).

The average of the parents' education was calculated, with a maximum score of 5 points. In the same way, the score was calculated in response to the parents' job, provided that the answer to their work situation was: "Works and earns a salary" for at least one of the parents or guardians. Finally, the number of books at home was also scored from 1 to 5, in ascending order as shown in the table. Therefore, depending on the level of education and work of the parents and the amount of books at home, a score of up to 15 points was assigned.

The last question, referring to material goods, was scored as follows: 5 resources out of the 8 were selected according to the highest variability among all those who have answered the survey correctly. One point was assigned to each item where the participant answered "Yes" (owns this asset), thus scoring a maximum punctuation of 5 points.

**Table 2.** Sample questions to measure socio-economic status. (Translated)

Question	Lázaro-Ibarrola (2022)	Response and punctuation
Highest level of education of your mother or guardian 1.	<i>¿Cuál es el nivel de estudios más alto que ha completado tu madre o tutor 1?</i>	<b>1pt.</b> Did not complete basic education. No education. <b>2pt.</b> Basic education, primary education and/or ESO (Compulsory Secondary Education) <b>3pt.</b> Vocational training or Bachillerato/Bacalaureate <b>4pt.</b> University and/or Master's Degree <b>5pt.</b> Doctorate.
Employment status of your mother or guardian 1.	<i>¿Cuál es la situación laboral de tu madre o tutor 1?</i>	<b>1pt.</b> Works at home, housework <b>2pt.</b> Unemployed (for at least one year) <b>3pt.</b> Is retired <b>4pt.</b> Works and earns a salary (receives money for her work) <b>5pt.</b> Not applicable (in case she is missing)
Actual job of your mother or guardian 1.	<i>¿Cuál es el actual trabajo de tu madre o tutor 1?</i>	<b>1pt.</b> Construction worker... <b>2pt.</b> Administrative... <b>3pt.</b> Administrative manager... <b>4pt.</b> Teacher, Nurse, B-level Civil servant... <b>5pt.</b> Doctor, Lawyer, Teacher...
Amount of books in your house.	<i>Aproximadamente, ¿cuántos libros hay en tu casa? En una balda de un metro caben aproximadamente 40 libros. No incluyas las revistas, los periódicos ni los libros de texto.</i>	<b>1pt.</b> 0 to 10 books <b>2pt.</b> 11 to 50 books <b>3pt.</b> 51 to 100 books <b>4pt.</b> 101 to 200 books <b>5pt.</b> More than 200 books

Personal goods at home (8 items: a) computer/laptop, b) internet, c) reading books, d) encyclopaedias, e) specialised magazines, f) smartphone, g) tablet and h) individual bedroom)	<i>¿Tienes los siguientes recursos en tu casa? (8 elementos: a)ordenador/portátil, b) internet, c) libros de lectura, d) enciclopedias, e) revistas especializadas, f) smartphone, g) tablet y h) dormitorio individual)</i>	<b>1pt.</b> (Yes) Computer <b>1pt.</b> (Yes) Encyclopaedias <b>1pt.</b> (Yes) Specialised magazines <b>1pt.</b> (Yes) Tablet <b>1pt.</b> (Yes) Individual bedroom
---	--	---

### Self-efficacy questionnaire

Regarding the student's self-efficacy questions, was a reduced version of Pastorelli et al's (2001) questionnaire. It included 23 out questions targeting the three dimensions of Self-Efficacy:

1. Perceived Academic Efficacy: 14 items targeted different aspects of academic activities: 4 dealt with student's beliefs in their capability to master different areas of coursework; 8 items measured students' perception of the capacity for regulating their motivation and learning activities and 2 addressed students' beliefs on teachers and parents' expectations.
2. Perceived Social Efficacy: 7 items related to a variety of social domains: 2 related to forming and maintaining social relationships and dealing with conflicts; and 5 measured self-assertive efficacy.
3. Self-Regulatory Efficacy: 2 items measuring the student's capacity to resist peer pressure to engage high-risk activities related to alcohol, drugs and other activities.

Responses were represented with a 5-point-Likert scale from 1- strongly disagree to 5- strongly agree (Appendix 4). Each of the answers was scored from 1 to 5 in ascending order, with a total of 115 points (23 questions with a maximum of 5 points per answer). The score was then adjusted to a scale of 1 to 10.

The instruments were administered over a period of three weeks. In the first place, the socio-economic survey was uploaded into *Google Forms* format. This survey was shared with the students through a link uploaded onto their *Google Classroom's* platform. Students were able to answer this survey with the help of their teacher or at home with their parents or tutors. Afterwards, a day and time were arranged with the teachers to conduct the content exam. Some of them agreed to take the exam in their own class time and on other occasions another time slot had to be found. The exam lasted an average of 35 minutes depending on the class and the group. The instructions were the same for all groups. Both the assessment and the survey were accomplished individually, following a regular exam procedure for the sake of

students' commitment. The exam was presented as a placement test for the public university of Navarre.

Once all exams and questionnaires were collected, both SES and self-efficacy results were computed and compared across groups. Data obtained from the different instruments was analysed through normality tests (Kolmogorov/Saphiro-Wilkins). Some variables did not meet normality and thus, non-parametric tests were used for all the variables. The average obtained from the test scores was compared between the different groups and methodologies. Then, the mean points obtained in the socio-economic survey were compared with the three groups and the two methodologies, to determine whether there was a noticeable relationship between the results (both SES and Self-Efficacy) and the language programmes, and whether these influenced the students' grades.

## 5. Results

### Social Sciences exam

Table 3 presents the test scores by methodology and Table 4 presents the test scores by linguistic programme.

When CLIL groups were merged irrespective of the intensity of exposure, the mean scores' differences were smaller (Table 3). Non-CLIL methodology students had a slightly lower average (0.46 points of difference) than their CLIL counterparts. Despite this, the standard deviation of the CLIL group was higher, which showed a less homogeneous group. Test results were compared between groups by non-parametric tests (Kruskal-Wallis) and no differences were found ( $>0.05$ ). Therefore, according to the methodology (non-CLIL vs. CLIL) the groups had similar final test results.

**Table 3.** Mean and SD of the assessment's final scores according to methodologies.

methodology	N	Mean	SD
<i>non-CLIL methodology</i>	14	5.48	1.47
<i>CLIL methodology</i>	32	5.94	2.02

The three groups passed the exam, although the scores were not very high. When the total test score was examined, the non-CLIL students obtained the lowest average marks (5,48 points out of 10), closely followed by the soft-CLIL students. As expected, it is the students in the hard-CLIL group who have the highest mean scores, with up to 0.58 and 0.71 points of difference respectively. Despite this, we can observe a higher standard deviation in the hard-CLIL group, which shows that it is a less homogeneous group (Table 4).



When each of the parts was observed, however, results were less clear-cut. Taking into account the separate marks by type of knowledge required, we found that both non-CLIL and soft-CLIL students had better marks in the procedural knowledge part. Hard-CLIL students scored worse results in the procedural knowledge part (open-ended question), although there were no significant differences between the three groups.

**Table 4.** Mean and SD of the assessment's scores for the different linguistic programmes.

group	N	Mean	SD
<b>Declarative knowledge (5pt)</b>			
non-CLIL	14	2.66	1.14
soft-CLIL	14	2.52	1.16
hard- CLIL	18	3.60	1.05
<b>Procedural knowledge (5pt.)</b>			
non-CLIL	14	2.82	1.05
soft-CLIL	14	3.09	.97
hard- CLIL	18	2.60	1.48
<b>Final scores (10pt)</b>			
non-CLIL	14	5,48	1.47
soft-CLIL	14	5,61	1.61
hard- CLIL	18	6.19	2.30

**Table 5.** Kruskal-Wallis test to compare non-CLIL, soft-CLIL and hard-CLIL groups's scores.

	Kruskal-Wallis H	DF	Asymp.Sig.
<b>Declarative knowledge</b>	6.709	2	.035
<b>Procedural knowledge</b>	.874	2	.646
<b>Total scores</b>	1.607	2	.448

a. Kruskal Wallis test

b. Variable: non-CLIL, soft-CLIL, hard-CLIL

Moreover, we found a significant difference (.035) between the three groups in the declarative knowledge part (definitions) (Table 5). Therefore, we contrasted pairs of groups through a series of Mann-Whitney tests and significant differences were found between soft and hard-CLIL groups (.015) (Table 6). Hard-CLIL students showed better skills in defining

specific concepts yet showing a higher standard deviation. In this part, soft-CLIL students were the worst performers.

**Table 6.** Mann-Whitney U test: Soft-CLIL and hard-CLIL groups in the Declarative knowledge part.

	Mann-Whitney U	Wicolxon W	Z	Asymp.Sig. (bilateral)
<b>Declarative knowledge</b>	62.500	167.500	-2.424	.015

a. Variable: soft-CLIL, hard-CLIL

### Socio-Economic Status

As for the socio-economic level of the participants, as expected, we found a difference of up to 3.21 points out of 20 between non-CLIL and CLIL students (Table 7). Significant differences (.010) were found in the total SES score among methodologies, so they were compared by groups.

Within the three groups, hard-CLIL students scored the highest socio-economic level with a total of 13.28 points out of 20 (Table 8). The difference, however, was not significant compared to the soft-CLIL students. The difference between non-CLIL and CLIL students was more pronounced (.044 between non- and soft-CLIL; and .003 between non- and hard-CLIL).

**Table 7.** SES results according to methodologies.

Methodology	N	Mean	SD
<i>non-CLIL methodology</i>	14	9.68	2.97
<i>CLIL methodology</i>	32	12.89	2.79

**Table 8.** SES results according to groups.

### Self-efficacy levels

Group	N	Mean	SD
<i>non-CLIL</i>	14	9.68	2.97
<i>soft-CLIL</i>	14	12.39	2.77
<i>hard-CLIL</i>	18	13.28	2.81

Regarding the methodology, we find a difference of 8.12 out of 115 points between groups (Table 9). Despite this, we found no significant differences according to the methodology. In any case, the scores of the three groups were generally high. SD shows that the results of the non-CLIL group are more homogeneous than those of the CLIL group.

**Table 9.** *Self-efficacy levels according to methodologies.*

<b>Methodology</b>	<b>N</b>	<b>M</b>	<b>SD</b>
<b><i>non-CLIL methodology</i></b>	14	81.93	13.44
<b><i>CLIL methodology</i></b>	14	73.81	22.04

As for the differences according to groups, we find that students in the soft-CLIL group scored the lowest levels of self-efficacy, followed by hard-CLIL students. Students in the non-CLIL group scored highest (Table 10) and further, evidenced higher Self-Regulatory Efficacy levels. Accordingly, Self-Regulatory Efficacy levels were precisely the lowest in the soft-CLIL group (Table 11).

**Table 10.** *Self-efficacy levels according to groups.*

<b>Group</b>	<b>N</b>	<b>M</b>	<b>SD</b>
<b><i>non-CLIL</i></b>	14	81.93	13.44
<b><i>soft-CLIL</i></b>	14	69.00	24,27
<b><i>hard-CLIL</i></b>	18	77.24	20.04

**Table 11.** *Self-Efficacy levels according to categories.*

	<b>Perceived Academic Efficacy</b>	<b>Perceived Social Efficacy</b>	<b>Self-Regulatory Efficacy</b>
<b><i>non-CLIL</i></b>	3.55	3.55	3.82
<b><i>soft-CLIL</i></b>	3.07	2.97	2.61
<b><i>hard-CLIL</i></b>	3.35	3.40	3.44

### Relationship between CLIL, SES and Self-Efficacy

Various Spearman correlations were carried out to see whether there was a link between the academic results of the students, their SES and their Self-Efficacy levels. According to the methodology (non-CLIL vs. CLIL), no relation between marks, SES and Self-Efficacy was found.

Considering the different groups, we found that there was also no correlation between the aforementioned variables in the non-CLIL group. Nevertheless, in the soft-CLIL group, a strong relationship was found between the academic scores, and the Self-Efficacy levels (-.622). As for the hard-CLIL group, we found that there was a strong relationship between the test scores and Self-Efficacy levels (.543). Surprisingly, the correlation of the soft-CLIL group was negative while that of the hard-CLIL group was positive (Table 12).

**Table 12.** Spearman Correlation between final scores and self-efficacy levels.

	Self-efficacy levels		
	<i>N</i>	Correlation coefficient	Sig. (2-tailed)
<i>soft-CLIL final scores</i>	14	-.622	.013
<i>hard-CLIL final scores</i>	18	.543	.016

## 6. Discussion

This study aimed to measure students' academic performance in L1 and compare it with the type of methodology (non-CLIL vs. CLIL methodologies) in which their classes are conducted. For this purpose, we measured three different groups in which the exposure to English was different, but the methodology of two of the groups was CLIL (soft-CLIL and hard-CLIL groups), with a focus on the subject of Social Sciences. In addition, we wanted to take into account two other variables, which were the socio-economic level of the students and their level of self-efficacy, and compare them with the different groups to see if they were correlated.

According to RQ1, the academic results of the non-CLIL and CLIL methodologies did not show significant differences. Despite the CLIL group being a more heterogeneous group in terms of their results, they had a slightly higher average. In line with previous studies (Elorza & Muñoa, 2008), this could mean that the methodology followed in these groups would not be detrimental to the acquisition of the content. Moreover, as both CLIL groups are in the early stages of the CLIL methodology, they may gradually outperform their non-CLIL counterparts in subsequent years. This possible outperformance would be explained due to the fact that CALP acquisition develops years later, thus strategies related to academic competence in

general may arise later as well. In addition, we found that, although without significant differences, hard-CLIL students' group was the one that obtained the best scores, coincidentally being the group with the greatest number of hours of exposure to the CLIL methodology. Nevertheless, as mentioned above, as a heterogeneous group in terms of their academic scores, individual differences that were not taken into account in the present study may influence students' results as well.

Considering each part of the test separately (declarative knowledge and procedural knowledge parts), we found significant differences between non-CLIL and CLIL students in the declarative knowledge part (definitions). From these results, it is clear that CLIL students outperformed non-CLIL students in both parts, in line with the study of Elorza and Muñoa (2008). However, when comparing the results with those of the aforementioned study, we found dissimilarities. Whereas in Elorza and Muñoa's (2008) CLIL students scored massively higher in the procedural knowledge part, CLIL students in the present study did better in the declarative knowledge part. Despite this, it is worth noting the analysis by separate groups. Significant differences were found among soft and hard-CLIL students in the declarative knowledge part (definitions). These results suggest that hard-CLIL students performed better at defining concepts, whereas soft-CLIL students were the worst in this aspect. However, soft-CLIL students scored the best results in the procedural knowledge part. Therefore, these differences in performance among the test parts may be related to more specific variables that have not been taken into account in the present study. The fact that hard-CLIL students scored worse in the procedural knowledge part, may be due to the fact that they are able to express themselves better in English than in Spanish. This may be a result of the hours of exposure to this language (75% of the teaching hours, as previously mentioned), and that is why they are not able to give answers as accurate as those given in English. On the other hand, it is the soft-CLIL students who scored better results in procedural knowledge. This may be due to the fact that they maintain a balance between the CLIL methodology in the Social Sciences subject and exposure to their L1 in other subjects (45% of the teaching hours). Therefore, academic language in Spanish is probably more frequent, so it can be suggested that they are able to express themselves correctly in both languages.

Regarding RQ2, we found that the results were in line with those of Fernández-Sanjurjo et al. (2017). The socio-economic status of non-CLIL students was significantly lower than that of CLIL students. This could be due to the tendency of families with higher SES to include their children in this type of programme (Bruton, 2013). In terms of language programmes, no significant differences were found between the soft- and hard-CLIL groups, which could mean that these groups were similar with respect to their SES. Therefore, it could be said that the

SES levels of the students might be more related, if it is the case, to the methodology rather than to the programme itself.

In tune with RQ3, it was found that the students' SES (both non-CLIL and CLIL) was not related with their academic results. Therefore, it could be said that the SES variable did not influence the academic results of the participants, and it is possible that the subtle difference in the final scores of the groups is due to the methodology employed rather than other variables. These results would demonstrate that the fact that the best students are those who access to CLIL programmes (Bruton, 2019) does not necessarily have to do with their SES, at least in a public secondary school such as the one in the present study.

With reference to RQ4, it was found that the levels of self-efficacy of the non-CLIL methodology were higher than those of CLIL. Although these differences were not significant, the low levels of Self-Efficacy of CLIL learners may be due to the cognitive burden they face within these programmes (Jaekel, 2020). These students may also face higher demands from their parents and tutors, due to the prestige attributed to these types of programmes (Bruton, 2013). If we take into account the different categories of self-efficacy, self-regulatory efficacy was the dimension that differed the most between programmes. This result highlights the fact that non-CLIL students answered positively to the questions "How well can you resist peer pressure to do things in school that can get you into trouble?" and "How well can you stand firm to someone who is asking to do something unreasonable or inconvenient?" (Appendix 4). On the other hand, soft-CLIL students were those who responded more negatively, thus suggesting that they are less determined to resist peer pressure to engage in high-risk activities. The fact that this category is the most irregular one was also reported in Pastorelli et al.'s (2001) study. According to Pastorelli et al. (2001), gender, variable that was not taken into account in the present study, may influence self-regulatory efficacy.

Concerning RQ5, for both the non-CLIL and CLIL methodologies, no relationship was found between students' grades and their level of Self-Efficacy. However, taking into account the different groups, we found a strong relationship between soft-CLIL students' Self-Efficacy and their test scores. Their levels of Self-Efficacy were quite low in relation to their marks. These results may suggest that soft-CLIL learners show higher levels of self-demand, which could lead to higher levels of stress (Olhberger and Wegner, 2019). These results are contradictory to those of the hard-CLIL group, in which we found that Self-Efficacy levels are positively correlated with academic results. The moderate relationship between hard-CLIL students' grades and self-efficacy levels suggests that, in relation to studies such as Jaekel's (2020), a higher level of self-efficacy among students has a positive effect on student grades. This contradictory distinction between the two CLIL groups may be due to school-specific variables. As some teachers at the school explain, students in the hard-CLIL group enjoy a

higher prestige not only among teachers, but also among students. This is why soft-CLIL students may have a poorer self-perception of themselves, being too self-demanding that they do not trust their own abilities to achieve their desired objectives. Despite the possible consequences of these results, low levels of self-efficacy among soft-CLIL students did not have an impact on their academic performance.

Given these results, it could be said that both SES and self-efficacy variables have no significant effects on students' academic results. Thus, although there were no significant differences in the academic outcomes of non-CLIL and CLIL students, we found that the latter group slightly outperformed the former. This fact suggests that the CLIL methodology, in tune with the results of Elorza and Muñoa's (2008), may have positive effects not only on language performance but on content performance as well. Learners not only do not experience a decrease in their levels of content acquisition, but also obtain (subtly) better scores. That is, they are able to transfer their content knowledge from English (FL) to Spanish L1, express themselves clearly and accurately and develop a written topic successfully in their L1.

## **7. Conclusion, pedagogical implications and further studies**

This paper contributes to developing the body of knowledge of content acquisition in CLIL settings. Comparing the content acquisition of CLIL and non-CLIL learners, we observe that there are no significant differences between the groups, although CLIL learners perform slightly better. In terms of exposure to the CLIL methodology, we find that students enrolled in a hard-CLIL programme perform better than the soft-CLIL group. These results suggest that an increase in the number of CLIL subjects in the students' syllabus would not be detrimental, and would even be beneficial for content acquisition.

This study can be useful for the school, as it provides information regarding the different groups, in relation to their academic results in Spanish L1. This transfer of knowledge from one language to another will be important for students who study through CLIL programmes, since it is in higher levels of education (Baccalaureate and University) where subjects are generally taught in L1 Spanish, and the acquisition of content should be necessary regardless of the language in which it is taught. In addition, students should not only have the knowledge acquired but also be able to express it in one language or another, so that emphasis should be placed on written expression regardless of the language used.

Further studies may focus on a longitudinal study of students throughout the secondary education stage, in terms of their acquisition of second language content and its possible transfer to the mother tongue, and compare this with their counterparts studying in their mother

tongue. In this way, it could be examined if CLIL learners gradually outperform their non-CLIL counterparts in the field of content acquisition throughout the years.

Moreover, future research may consider focusing on the written expression differences between CLIL and non-CLIL students. In this way, it could be explored whether a high exposure to a second or foreign language has detrimental effects on written expression in the mother tongue. Moreover, the importance of written expression in a foreign language may be examined, in order to confirm if it influences the ability to write more efficiently in the L1 in academic contexts.

## **8. Limitations**

Although this study contributes to the investigation of the acquisition of content through the CLIL methodology, several limitations can be found. Firstly, the sample size was small, so conclusions may not be generalizable to other settings. In addition, the conditions under which both the test and the questionnaire were administered could have been more homogenised, so that all students took the test at the same time, to avoid possible variables such as the time at which the test was administered and the exact required time for taking the test. Furthermore, by preparing this study in advance, a total homogenisation of the contents taught in class could be carried out, so that all groups would dedicate exactly the same number of hours to the teaching of the same contents.

## **9. Acknowledgments**

I would like to express my sincere gratitude to IES Basoko, to its students, to my internship tutor and to the teachers involved in this study. In addition, thanks to the supervisor of this paper, Izaskun Ibarrola, for having guided and helped me during the elaboration of the present study.



## Bibliography

- Ball, P., Kelly, K., & Clegg, J. (2015). *Putting CLIL into practice: Oxford handbooks for language teachers*. Oxford University Press.
- Baranova, T. A., Kobicheva, A. M., & Tokareva, E. Y. (2019). *Does CLIL work for Russian higher school students? The Comprehensive analysis of Experience in St-Petersburg Peter the Great Polytechnic University*. 2019 7th International Conference on Information and Education Technology (pp. 140-145).  
<https://doi.org/10.1145/3323771.3323779>
- Bruton, A. (2013). CLIL: Some of the reasons why... and why not. *System*, 41(3), 587-597.  
<https://doi.org/10.1016/j.system.2013.07.001>
- Bruton, A. (2019). Questions about CLIL which are unfortunately still not outdated: A reply to Pérez-Cañado. *Applied Linguistics Review*, 10(4), 591-602.  
<https://doi.org/10.1515/applirev-2017-0059>
- Cenoz, J. (2015). Content-based instruction and content and language integrated learning: the same or different? *Language, Culture and Curriculum*, 28(1), 8–24.  
<https://doi.org/10.1080/07908318.2014.1000922>
- Coyle, D. (2012). *C.L.I.L. Content and Language Integrated Learning* (1st ed.). Cambridge University Press.
- Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. *Encyclopedia of language and education*, 2(2), 71-83.  
[https://doi.org/10.1007/978-0-387-30424-3\\_36](https://doi.org/10.1007/978-0-387-30424-3_36)
- Dewaele, J. M. (2012). Personality: Personality traits as independent and dependent variables. *Psychology for language learning* (pp. 42-57). Palgrave Macmillan, London.  
 DOI: [10.1057/9781137032829\\_4](https://doi.org/10.1057/9781137032829_4)
- Elorza, I., & Muñoa, I. (2008). Promoting the Minority Language Through Integrated Plurilingual Language Planning: The Case of the Ikastolas. *Language, Culture and Curriculum*, 21(1), 85–101.  
<https://doi.org/10.2167/lcc345.0>
- Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J. M. (2017). Analysing students' content-learning in science in CLIL vs. non-CLIL programmes: empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism*, 22(6), 661–674.  
<https://doi.org/10.1080/13670050.2017.1294142>

- Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J. M. (2018). Assessing the influence of socio-economic status on students' performance in Content and Language Integrated Learning. *System*, 73, 16-26.  
<https://doi.org/10.1016/j.system.2017.09.001>
- Hughes, S. P., & Madrid, D. (2019). The effects of CLIL on content knowledge in monolingual contexts. *The Language Learning Journal*, 48(1), 48–59.  
<https://doi.org/10.1080/09571736.2019.1671483>
- Jaekel, N. (2020). Language learning strategy use in context: the effects of self-efficacy and CLIL on language proficiency. *International Review of Applied Linguistics in Language Teaching*, 58(2), 195-220.  
<https://doi.org/10.1515/iral-2016-0102>
- Kim, D. H., Wang, C., Ahn, H. S., & Bong, M. (2015). English language learners' self-efficacy profiles and relationship with self-regulated learning strategies. *Learning and individual differences*, 38, 136-142.  
<https://doi.org/10.1016/j.lindif.2015.01.016>
- Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, 1(1).  
 DOI: [10.2174/1874913500801010030](https://doi.org/10.2174/1874913500801010030)
- Martínez Agudo, J. D. D. (2020). The impact of CLIL on English language competence in a monolingual context: a longitudinal perspective. *The Language Learning Journal*, 48(1), 36–47.  
<https://doi.org/10.1080/09571736.2019.1610030>
- Ohlberger, S., & Wegner, C. (2019). CLIL modules and their affective impact on students with high English anxiety and low self-efficacy. *Apples-Journal of Applied Language Studies*, 13 (3), 1-15.  
<https://doi.org/10.17011/apples/urn.201906253409>
- Pastorelli, C., Caprara, G. V., Barbaranelli, C., Rola, J., Rozsa, S., & Bandura, A. (2001). The structure of children's perceived self-efficacy: A cross-national study. *European Journal of Psychological Assessment*, 17(2), 87.  
 DOI: [10.1027//1015-5759.17.2.87](https://doi.org/10.1027//1015-5759.17.2.87)
- Pérez Cañado, M. L. (2016). From the CLIL craze to the CLIL conundrum: Addressing the current CLIL controversy. *Bellaterra Journal of Teaching and Learning Language and Literature*, 9(1), 0009-31.  
<https://doi.org/10.5565/rev/jtl3.667>

Piacentini, V. (2021). CLIL and Science education. A review for a Language focus in Science teaching. *Ricerche di Pedagogia e Didattica. Journal of Theories and Research in Education*, 16(3), 113-131.

<https://doi.org/10.6092/issn.1970-2221/12646>

Şenay, H. H. (2018). *The Relationships among academic self-efficacy, academic optimism, family income and academic achievement* (Master's thesis, Middle East Technical University).

## Appendix

### Appendix 1. Rubric for the Social Sciences exam.

Definition	Example of correct answer	1pt.	0.75pt.	0.50pt.	0pt.
<b>Birth Rate</b>	<i>Total number of people born in a given place and time period in relation to the total population.</i>	Defines the concept clearly and accurately	Defines the concept clearly enough/ Accurately enough	Defines the concept clearly and accurately but lacks some elements.	The definition is not understood or is not accurate.
Open-ended question	Example of correct answer	1.5pt.	1pt.	0.5pt.	0pt.
<b>Population Density</b>	<i>Number of people living in a given space.</i>	Defines the concept clearly and accurately.	Defines the concept clearly enough/ Accurately enough.	Defines the concept clearly and accurately but lacks some elements.	The definition is not understood or is not accurate.
<b>Most populated cities</b>	<i>Tokio, Beijing, New Delhi</i>	<b>1pt.</b>	<b>0.75pt.</b>	<b>0.50pt.</b>	<b>0pt.</b>
		At least 3 capital cities/densely populated cities.	At least 2 capital cities/densely populated cities.	At least 1 densely populated city.	Irrelevant cities, names of countries instead of cities.
<b>Why is the population unevenly distributed?</b>	<i>Physical (climate, hydrography or mountain relief), or human factors (economic development) that condition the habitability of the geographical space and its proper explanation.</i>	<b>2.5pt.</b>	<b>1.75pt.</b>	<b>1pt.</b>	<b>0pt.</b>
		Explains at least one factor in a clear and detailed way, provides ideas or opinions related to the topic he/she is explaining..	Explains at least one factor in a more or less clear and detailed way, gives some ideas or opinions related to the topic he/she is explaining.	Explains at least one factor in a sufficiently clear and detailed way, provides some ideas or opinions that may not be related to the topic he/she is explaining.	Does not explain at least one factor in a sufficiently clear and detailed way, does not provide ideas or opinions, or these have nothing to do with the topic he/she explains.

**Appendix 2.** *Example of responses given by students.*

Question	Example	
	Correct	Incorrect
<b>1. Define the following concept: Birth rate.</b>	<i>The birth rate is the number of people born in a country or region over the course of a year.</i>	<i>Nation you come from. Country where you were born or live.* * This participant mixed up the concepts of Birth Rate and Nationality (Spanish Natalidad vs. Nacionalidad).</i>
<b>2. Content-development question.</b>	<i>Population density represents the number of people living in a given space, measured in people per square kilometre. The most densely populated cities may be capital cities such as Tokyo or Beijing. The population is unevenly distributed because in developed countries there are more job opportunities [...], better life quality [...], due to emigration...</i>	<i>Where more people are concentrated. The most populated countries are poorer and habitants die due to lack of resources.</i>

**Appendix 3.** *Dimensions questioned within the socio-economic survey. Translated into English.*

Item questioned	Observations
<b>Student's profile</b>	<i>Linguistic programme, age and sex</i>
<b>Language(s) spoken at home</b>	<i>Languages used to communicate at home, level of these languages, Use of languages other than English or Spanish.</i>
<b>Study of English and other languages</b>	<i>Extracurricular foreign language hours, appreciation of foreign language (English)</i>
<b>Stay in English-speaking countries</b>	<i>Stay in English-speaking countries and length of stay</i>
<b>Use of English in everyday life</b>	<i>Use of English and language preferences</i>
<b>Socioeconomic status</b>	<i>Parents' level of education, parents' jobs, number of books at home, resources at home</i>
<b>Self-Perception</b>	<i>Attitude towards English, Importance of the English language, Academic self-perception, Self-Efficacy</i>

**Appendix 4. Example of Self-Efficacy questions.**

<b>Pastorelli et al., (2001)</b>	<b>Self-Efficacy dimension</b>	<b>Adaptation</b>	<b>Response</b>
<b>22. How well can you resist peer pressure to do things in school that can get you into trouble?</b>	Self-regulatory efficacy	<i>No me dejo llevar por compañeros que insisten en hacer cosas que pueden meterme en problemas.</i>	1- Totalmente en desacuerdo 2- Bastante en desacuerdo 3- Ni de acuerdo ni en desacuerdo 4- Bastante de acuerdo 5- Totalmente de acuerdo
<b>26. How well can you stand firm to someone who is asking to do something unreasonable or inconvenient?</b>	Self-regulatory efficacy	<i>Soy capaz de enfrentarme a alguien que me pide que haga algo irracional o inconveniente.</i>	1- Totalmente en desacuerdo 2- Bastante en desacuerdo 3- Ni de acuerdo ni en desacuerdo 4- Bastante de acuerdo 5- Totalmente de acuerdo
<b>27. How well can you live up to what your parents expect of you?</b>	Perceived Academic Efficacy	<i>Estoy a la altura de lo que mis padres esperan de mí.</i>	1- Totalmente en desacuerdo 2- Bastante en desacuerdo 3- Ni de acuerdo ni en desacuerdo 4- Bastante de acuerdo 5- Totalmente de acuerdo
<b>28. How well can you live up to what your teachers expect of you?</b>	Perceived Academic Efficacy	<i>Estoy a la altura de lo que mis profesores esperan de mí.</i>	1- Totalmente en desacuerdo 2- Bastante en desacuerdo 3- Ni de acuerdo ni en desacuerdo 4- Bastante de acuerdo 5- Totalmente de acuerdo