

The effect of CLIL on productive thematic vocabulary

Dr Leah Geoghegan 

Departamento de Filologías Modernas, Universidad de La Rioja, Logroño, Spain

Correspondence

Leah Geoghegan, Departamento de Filologías Modernas, Universidad de La Rioja, c/ San José de Calasanz s/n, 26004 Logroño, La Rioja, Spain.

Email: legeogw@unirioja.es

Abstract

In recent years, content and language integrated learning (CLIL) has become increasingly popular throughout Spain. One suggested benefit of this approach is that it increases students' exposure to the language and facilitates their acquisition of subject-specific vocabulary (SSV). However, there is a scarcity of research demonstrating this advantage, particularly in terms of lexical availability (LA). This has led to calls to focus on semantic prompts which are linked to the CLIL subject, to better understand the effect of this linguistic exposure. This study investigates the LA of students taking either science or economics through English. Results indicate that CLIL instruction plays an important role in acquiring SSV, revealing that it can help students improve their token production to the extent that they can effectively bridge the gap between them and students with a higher language level and greater token production in other domains.

En los últimos años, el Aprendizaje Integrado de Contenidos y Lenguas Extranjeras (AICLE) ha crecido a gran escala por toda España. Uno de los supuestos beneficios del enfoque es que aumenta las cantidades de exposición a la lengua meta que reciben los alumnos, facilitando en particular la adquisición del vocabulario relacionado con el contenido.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *International Journal of Applied Linguistics* published by John Wiley & Sons Ltd.

Sin embargo, hay una falta de investigación que demuestra esta ventaja, especialmente dentro del campo de la disponibilidad léxica. Esto ha hecho un llamado para centrarse en los centros de interés que están relacionados con la propia asignatura AICLE, para entender mejor su efecto. Con este fin, este estudio investiga la disponibilidad léxica de alumnos que cursan una asignatura de ciencia o economía en inglés. Los resultados indican que la instrucción AICLE juega un importante papel en la adquisición de vocabulario relacionado con el contenido de la clase, y revelan que la instrucción AICLE puede ayudar a que los alumnos mejoren su disponibilidad léxica hasta efectivamente romper la distancia entre ellos y otros alumnos con un nivel de idioma más alto y que producen más vocabulario en otros ámbitos lingüísticos.

KEYWORDS

CLIL, English as a foreign language, lexical availability, subject-specific vocabulary, thematic vocabulary, Disponibilidad léxica, Vocabulario específico, Vocabulario temático, Aprendizaje Integrado de Contenidos y Lenguas Extranjeras, inglés como lengua extranjera

1 | INTRODUCTION

Since the turn of the century, content and language integrated learning (CLIL) has been increasingly implemented throughout Spain. This approach, whereby content lessons are taught through the medium of a target language (TL), was seen as a remedy to previous shortfalls in language teaching: it provided a more authentic learning scenario, gave students a reason to use the language, and economically increased exposure to the language without increasing demands on busy timetables (Dalton-Puffer & Smit, 2007).

An additional purported benefit of a CLIL context is the effect it has on the acquisition of subject-specific vocabulary (SSV) (Heras & Lasagabaster, 2015). This is because by increasing exposure to the TL, CLIL is thought to help students acquire the specific vocabulary which is related to their content classes. However, there is little empirical evidence which supports this claim, especially in terms of lexical availability (LA), which is the retrieval of words prompted by a specific semantic category. For example, when given the prompt *Animals*, participants will likely produce words such as “cat,” “dog,” and so on. However, it remains to be seen what type of vocabulary participants will be able to retrieve in a prompt such as *Economics*, after receiving exposure to related vocabulary in a CLIL economics class.

To this effect, Canga Alonso (2017) has thus called for the inclusion of prompts which are linked to the CLIL subject, so as to establish the extent to which students actually acquire this vocabulary. In addition, he has emphasized the need to incorporate some measure of language proficiency alongside LA, to determine whether there is a relationship between language level and the number of words that students retrieve. This study seeks to address this gap by investigating the language level and token production of thematic and non-thematic prompts in a cohort of Spanish adolescent CLIL students.

2 | VOCABULARY AND CLIL

Within a European context, CLIL teaching arose in the mid-1990s, largely due to European language policy promoting intercultural competence, multiple identities, and multilingual citizens (Dalton-Puffer et al., 2014). It describes “an dual-focused educational context in which an additional language, thus not usually the first language of the learners involved, is used as a medium in the teaching and learning of non-language content” (Marsh, 2002, p. 15). It is thus intended to be an integrated approach where both language and content are seen on a continuum, with neither taking precedence (Coyle, 2007). One of the specific language competencies which has been found to be favorably affected in this context is vocabulary, particularly in the case of technical, semi-technical, and general academic language (Dalton-Puffer, 2008). As indicated by Baten et al. (2020), there are a number of reasons why CLIL education may have a positive effect on lexical development:

1. The combination of explicit and implicit learning conditions
2. Purposeful context and higher prospect of practice
3. Less intimidating atmosphere and error correction
4. Greater student–teacher interaction
5. Extramural language contact

First, it has been suggested that CLIL is ideally suited to learning vocabulary given that it encompasses both explicit and implicit learning conditions (Merikivi & Pietilä, 2014). For example, students may intentionally study target vocabulary which is related to the CLIL class units. In fact, as Matiasek (2005) notes, vocabulary is frequently the only linguistic skill which is treated explicitly in CLIL classrooms. In addition, students will also likely acquire vocabulary implicitly, as Tabuenca Cuevas and Alcaraz Mármol (2014) note, via reading different texts, listening to the teacher’s instructions, and communicating with the teacher or other classmates in the TL.

Second, CLIL provides a more meaningful learning scenario in which to learn vocabulary (Ruiz de Zarobe, 2011). This is because students need the language at hand to fulfil real purposes and are thus given a more contextualized opportunity to practice the TL (Rodríguez Bonces, 2012). This in turn is suggested to be optimal for vocabulary retention, as by receiving more meaningful learning opportunities and becoming more involved, it is more likely that vocabulary will be retained (Tabuenca Cuevas & Alcaraz Mármol, 2014).

Third, CLIL may provide a less threatening learning environment, given its focus on meaningful communication as opposed to linguistic accuracy. In other words, as learning the TL is not the sole focus of the lesson, students may feel less anxious when using the language and making mistakes (Baten et al., 2020).

Fourth, one potential benefit of a CLIL approach is the quantity of student–teacher interaction, which may lead to higher frequency of lexical encounters. This, however, is by no means a guarantee; while Nikula (2010) reports a higher amount of student–teacher interaction, Lo and Macaro (2015) report little interaction in a CLIL approach. The amount of interaction and consequentially the exposure to vocabulary is thus evidently subject to variation from one class to another.

Finally, researchers such as Sylvén (2019) have found that CLIL students may come into contact with the TL to a larger degree than non-CLIL students. Vocabulary gains, however, may not be attributable to the approach itself; as this author notes, CLIL students also outperformed non-CLIL students in a receptive vocabulary test before CLIL instruction had actually started. Thus, it is rather that students enrolled in CLIL classes also reported more extramural language contact and achieved better results on the vocabulary test.

In practice, Spanish CLIL research focusing on vocabulary has largely found that CLIL students have higher levels of receptive vocabulary (e.g., Diéguez & Martínez-Adrián, 2017; Agustín Llach & Canga Alonso, 2016) than non-CLIL students. Advantages have also been found in terms of productive vocabulary, particularly in adolescent learners (e.g., Canga Alonso & Arribas García, 2015; Jiménez Catalán & Agustín Llach, 2017). In addition to this benefit of

a CLIL approach on general vocabulary, recent research by Vraciu and Marcol (2023) and Rieder-Bünemann et al. (2022) has also highlighted the benefit of CLIL in fostering the acquisition of SSV. The former investigated the SSV of primary CLIL students after a 4-week program using tests catered to CLIL input. Findings revealed a significant improvement in participants' receptive and productive knowledge of SSV. The authors also highlight the need to use input profiled assessment methods when evaluating vocabulary gains in this context. The latter study analyzed the spontaneous oral classroom productions of secondary school students in Austria, who were studying European economics and politics through the medium of English. Their findings indicate that, although there was considerable variation between different students, there was substantial and active use of subject-specific terminology.

This kind of research is clearly of key importance in CLIL instruction, as the ability to comprehend and use such vocabulary evidently enables or prevents students from understanding the content they are studying. However, there remains a surprising lack of studies addressing SSV in this context, particularly in terms of LA. Rieder-Bünemann et al. (2022) also highlight a lack of research in terms of the range of subjects which are investigated and the systematic specification of SSV in the CLIL subject. Regarding the latter, while most research carried out on terminology in larger corpora is limited to individual words, technical vocabulary often goes beyond individual word boundaries (Nation, 2016). It is thus imperative that research take multi-unit words and phrases into account. One advantage of LA research is that it allows for this need, as the learner is not confined to producing individual words, but rather can retrieve any words or phrases which come to mind. In addition, it provides teachers with an efficient way to explore the most readily available and practical SSV which needs to be taught, as well as a way to assess it (Geoghegan & Agustín Llach, 2023).

3 | CLIL AND LEXICAL AVAILABILITY STUDIES

Lexical availability is understood to be both a concept and a data collection task (Jiménez Catalán, 2023). While the former is concerned with the cognitive process of retrieving words from the mental lexicon, the latter refers to the predefined semantic categories used to prompt the generation of words stored in the mental lexicon (Jiménez Catalán, 2023; Hernández-Muñoz et al., 2006). This is assessed using time-controlled tasks, such as the lexical availability task (LAT), which involve the activation of vocabulary in response to a particular stimulus, measuring the number of words retrieved (Fernández-Fontecha et al., 2021). Within a CLIL context, such LATs provide a convenient way to explore and assess the most readily available vocabulary used in content classes (Geoghegan & Agustín Llach, 2023). However, despite the value of LATs as a means of measuring students' lexical resources and the importance of acquiring SSV in CLIL, there is a surprisingly dearth of published research investigating LA in this context. In addition, the little research that has been carried out in Spain provides very disparate results.

On the one hand, in terms of young learners, Jiménez Catalán and Ojeda Alba (2009) investigated the language level and LA of CLIL ($n = 42$) and non-CLIL ($n = 44$) learners in sixth grade, using 15 different prompts. Results showed that there was a statistically significant difference between the two groups in a language placement test, to the advantage of the non-CLIL participants. In terms of LA, while the non-CLIL participants produced a higher number of tokens, there was no statistically significant difference between the two groups. On the other hand, in terms of adolescents, Jiménez Catalán and Agustín Llach (2017) analyzed the LA of 24 CLIL and 26 non-CLIL 8th grade students and 19 non-CLIL students in 10th grade. Notably, the 10th grade non-CLIL participants had received an equal number of hours of exposure to the TL as the 8th grade CLIL group, thus allowing the researchers to disentangle whether gains were attributable to the context itself or to the number of hours of exposure. Results revealed that the 8th grade CLIL group outperformed both their non-CLIL peers and the 10th grade non-CLIL group, suggesting that the advantage seen is indeed a result of CLIL instruction.

Regarding this research, there are two important points that must be made. First, while at primary level it appears that non-CLIL students perform somewhat better than CLIL students, at secondary level the opposite has been found, with CLIL students outperforming non-CLIL students. However, it is unclear whether these results are attributable to the age of the participants or to other factors which may be at play. Second, students in both studies have been assessed using general prompts which are not strictly related to the content which they are exposed to. This is evidently problematic, as there is little reason to believe that by studying a subject such as arts and crafts or economics through English, students will increase their vocabulary in domains such as *Animals*. Given the aforementioned research which highlights the benefit of CLIL in terms of vocabulary, and precisely SSV, it is evidently necessary for LA studies to explicitly target content-related lexical domains alongside language proficiency (Canga Alonso, 2017). This would allow us to better assess and understand students' gains in this learning context, by comparing the vocabulary retrieved to the target vocabulary in the CLIL syllabus, and to determine whether advantages in LA may simply be attributed to language level.

In order to address these issues, the current study seeks first to compare the language proficiency and LA of CLIL students who are enrolled in the same program but who at times take different classes and, second, to better link the LA prompts used to students' target CLIL vocabulary. This would allow us to ensure that participants have received the same amount of exposure to the TL via their participation in the same CLIL program, and to answer Canga Alonso's (2017) call for the inclusion of both a measurement of proficiency and semantic categories which are actually related to the vocabulary studied in the students' content classes. To this effect, the study at hand set out to answer the following research questions:

RQ1: Is there a quantitative difference in the L2 proficiency of adolescent CLIL students taking science or economics?

RQ2: Is there a quantitative difference in the production of non-thematic and thematic vocabulary of adolescent CLIL students taking science or economics?

RQ3: Are there longitudinal differences in the production of non-thematic and thematic vocabulary of adolescent CLIL students taking science or economics?

As has been suggested in previous research, there is a tendency for more advanced learners to produce a higher number of words (van Ginkel & van der Linden, 1996, as cited in Schmitt, 2000). It is thus suggested that language level will play an important role in the results, and that students with a higher language level will also perform better in terms of LA. However, considering the assumption that a CLIL approach has a positive effect on students' SSV (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015; Rieder-Bünemann et al., 2022), it is hypothesized that when the task at hand is related to the participants' content classes, there will be differences in the number of tokens produced depending on the exposure they have received to this vocabulary in their CLIL classes. In other words, economics students are expected to produce more tokens in a prompt related to economics than science students.

4 | METHODOLOGY

4.1 | Research design

The purpose of this study is to determine whether there is a quantitative difference in the language level and productive thematic and non-thematic vocabulary of adolescent students taking different CLIL classes, and whether this difference evolves over time. To this end, language level and LA were analyzed in students who were enrolled in two English CLIL classes, one compulsory and one optional. All participants studied physical education (PE) in English, while some participants took a science subject and others took economics. Data were gathered when students were in 10th grade and then 1 year later when they were in 11th grade to determine whether there were differences

TABLE 1 Research design.

Data collection 1		Data collection 2	
10th Grade		11th Grade	
Science	Economics	Science	Economics

TABLE 2 Participants' English content and language integrated learning classes.

	10th Grade (n = 41)	11th Grade (n = 40)
Male	<ol style="list-style-type: none"> 1. Physics and chemistry (n = 10) or Economics (n = 5) 2. PE (n = 15) 	<ol style="list-style-type: none"> 1. Economics (n = 10)
Female	<ol style="list-style-type: none"> 1. Biology (n = 16) or Economics (n = 10) 2. PE (n = 26) 	N/A

in L2 proficiency or non-thematic and CLIL thematic vocabulary (Table 1). While the majority of the participants remained the same across the two grades, some participants took part in only one data collection. When conducting the longitudinal analysis, only students who took part in both data collections and studied the same CLIL subject from one year to the next were included in the analysis.

4.2 | Participants

The participants in this study were Spanish native speakers who attended two semi-private sister schools, one for males and one for females. These schools conduct content and language classes in the students' native language, Spanish, as well as two TLs, English and French, as part of their plurilingualism project. This approach is adopted throughout early childhood education, primary education, and obligatory secondary education (ages approximately 3 to 16), and in general ceases when students advance to high school. In the program, classes are organized in such a way that each vehicular language takes up a third of the students' school day and, in some cases, students also study a fourth language such as Latin or Greek. In terms of language level, students have reached a B2 level according to the school curriculum, which is geared toward a B2 level in 10th grade and advances to a C1 level in 11th grade.

In terms of their English CLIL classes, 10th grade students (n = 41) take two CLIL subjects through the medium of English. While all students take PE, they also choose to study either a science subject (n = 26, physics and chemistry, or biology) or economics (n = 15). In this grade, these subjects are studied only in English, and not in Spanish. While these CLIL classes generally cease to be taught in English in 11th grade, male students enrolled in economics continue to study this subject through the medium of English, whereas female students who choose to study economics do so through Spanish. This exposure is summarized in Table 2.

In order to compare the differences in types of CLIL exposure, the participants were organized into two groups: those who studied a science subject and those who studied economics. These participants were assessed both in 10th grade and 11th grade (Table 3).

In 10th grade, while all students studied PE, 26 studied science and 15 studied economics. The following year, science students no longer took content classes in English, while 9 female participants studied economics through Spanish and 10 male participants studied economics through English.

While the participants in each grade were largely the same (n = 32), a total of 17 students—9 in 10th grade and 8 in 11th grade—took part in only one data collection. This was largely due to the fact that the change from 10th to 11th grade entails a changeover in school stage (from obligatory secondary education to high school). As a result,

TABLE 3 Subgroups of students taking different content and language integrated learning subjects.

	10th Grade (CLIL)	11th Grade (CLIL/non-CLIL)
Science	26 (English)	21 (Spanish)
Economics	15 (English)	19 (10 English, 9 Spanish)
Total	41	40

*Sample C-test***School for heroes**

The internationally acclaimed Fire Services College in Gloucestershire is widely regarded as the best fire college in the world. The site, a former airfield which once launched Wellington bombers for raids on Germany, was recently transferred from the Home Office to the Fire Services. And such is its reputation that people come from as far away as China and Trinidad to sharpen up their leadership skills. Even fire chiefs from the other side of the world have attended as College students.

FIGURE 1 Sample C-test. [Color figure can be viewed at wileyonlinelibrary.com]

some students no longer attended the participating schools in 11th grade, while others enrolled in 11th grade. In other cases, this was due to absences on the day of one data collection. In addition, while most students who had taken a science subject or economics in 10th grade continued with this subject in 11th grade some students chose to change subjects in 11th grade, when starting high school. When conducting the longitudinal analysis, only students who studied the same CLIL subject in both 10th and 11th grades were included: science ($n = 18$) and economics ($n = 10$).

4.3 | Instruments

4.3.1 | Language level C-test

The language level of the participants in this study was measured by means of a C-test, a text completion test in which “proficiency can be measured via the rate of successful restorations of the missing message elements” (Grujić & Danilović, 2012, p. 2). C-tests have been widely used as a means of testing language proficiency (Daller et al., 2003), given the ample research which has proven their ability to tap macro-level skills and processing (McManus, 2011). The C-test in this study was adopted from Daller and Phelan (2006) and consisted of four texts taken from an online news website. A sample text was provided before the exercise which the researcher used to explain the task (Figure 1). Each text contained a total of 20 gaps; from the second sentence onward, the second half of each second word was deleted and replaced with a blank space.

4.3.2 | Lexical availability task

To assess the participants' LA, they were administered a paper-and-pencil LAT, where they were presented with a total of four prompts and told to write down any words that came to mind. Each prompt was presented on a separate page

TABLE 4 Language level in science and economics students.

English C-test	<i>M</i>		<i>SD</i>		<i>p</i>
	Science	Economics	Science	Economics	
	10th Grade	38.00	27.20	10.42	
11th Grade	44.10	30.32	10.70	10.36	<0.001

with numbered lines, and participants were allowed 2 min to complete each prompt. The prompts chosen for the task included two general prompts and two prompts selected specifically to tap into students' English SSV:

1. Animals
2. Food & Drink
3. Sport & Physical Activity
4. Economy & Money

The first two general prompts, *Animals* and *Food & Drink*, were included as they had been used in previous LA research with Spanish adolescents (Canga Alonso, 2017; Fernández Orío & Jiménez Catalán, 2015) and had been found to be extremely productive, meaning that participants are generally able to produce a large number of tokens in these areas. The other two prompts were selected to tap into the SSV of two of the students' English CLIL classes: *Sport & Physical Activity* (PE) and *Economy & Money* (economics). While these prompts were tailored to the specific content the participants were exposed to, similar prompts had also been used in previous research, for example, *Sports and Hobbies* (Agustín Llach & Fernández-Fontecha, 2014) and *The Economy* (Neilson Parada, 2016).

Due to time restrictions, it was not possible to include a larger number of prompts. Consequently, those chosen include two general prompts, one content-related prompt related to a subject studied by all participants, and one content-related prompt related to a subject studied by a subgroup of participants. Economics was chosen for this purpose due to the fact that the science subject taken by participants varied depending on school and consequently gender, and so while some participants studied physics and chemistry, others studied biology. In addition, economics continued to be studied through English by a subgroup of students in 11th grade, whereas science subjects were at this point taught through Spanish.

In the lemmatization of students' responses, marking criteria were adopted from Jiménez Catalán and Agustín Llach (2017). This included (a) correction of spelling mistakes, (b) deletion of unintelligible and Spanish words, (c) including words repeated in the same prompt only once, (d) lexical phrases and compound words being lemmatized as one lexical unit, (e) deletion of proper nouns, and (f) changing plural words to the singular, unless the word is always plural.

5 | RESULTS AND DISCUSSION

5.1 | English language proficiency

In order to address RQ1, participants' English L2 proficiency as measured by the C-test was assessed, comparing those who studied a science subject with those who studied economics. This was done first for 10th grade and then for 11th grade and the results of each grade were then compared.

First, results of the independent sample *t*-tests revealed that the science students had a statistically significant higher level of English than the economics students both in 10th grade and in 11th grade (Table 4).

TABLE 5 Longitudinal differences in language level.

	M		F	p	η_p^2
CLIL group	Science	Economics	14.83	0.001	0.363
	41.55	25.50			
Time	10th Grade	11th Grade	4.51	0.043	0.148
	32.01	35.04			
CLIL group × time	10th Grade	11th Grade	0.197	0.661	0.008
	Science	Economics			
	Science	Economics			
	39.72	24.30			

To determine whether there were longitudinal differences between the science ($n = 18$) and economics students ($n = 10$) from 10th to 11th grade, two-way mixed ANOVAs were conducted, investigating the effect of time on CLIL group in terms of performance on the C-test in 10th and 11th grades (Table 5). As previously stated, only participants who were enrolled in the same subject in each grade were included, thus excluding the four longitudinal participants who had changed from studying science in 10th grade to studying economics in 11th grade or vice versa.

The results showed that there was a significant main effect of CLIL group on the C-test at each time, where the science group outperformed the economics group overall. There was also a significant main effect of time, with participants improving from 10th grade to 11th grade. However, no statistically significant interaction was found between time and CLIL group; both groups improved from 10th to 11th grade. There thus appears to be a key difference in the language level of the two groups, with science students performing substantially better in the C-test than economics students, however, both groups improved across time and this improvement does not appear to be related to CLIL group; in other words, as would be expected, studying a science subject or economics does not affect improvement across time.

5.2 | Token production

To address RQ2, participants' English LA in terms of token production was assessed, again comparing science students and economics students, first in 10th grade and then in 11th grade (Table 6).

As shown above, results found that the science students produced a statistically significant higher number of words in the overall LAT, as well as in three out of the four prompts, again in both grades. However, regarding the prompt *Economy & Money*, while the economics students had a slightly higher mean in 10th grade and a slightly lower mean in 11th grade, there was no statistically significant difference between the two groups in either grade. This suggests that, despite the fact that the science students exhibited a higher language level and produced a statistically significant higher number of tokens overall and in all other prompts, by studying economics through English, the economics students were not surpassed by their science peers in this prompt. In other words, the exposure these students have received in their economics classes has potentially allowed them to bridge the gap between them and the science students in this specific lexical domain, as although they do not outperform them, they produce a similar number of tokens in each grade.

5.3 | Token production across time

To address RQ3, which dealt with longitudinal differences from 10th to 11th grade, two groups of analyses were carried out which again focused on the 28 participants who had taken the same content classes across the two grade

TABLE 6 Token production in science and economics students.

10th Grade					
Category	M		SD		p
	Science	Economics	Science	Economics	
Animals	20.08	15.00	4.78	2.90	<0.001
Food & Drink	22.92	19.33	5.42	4.33	0.035
Sport & PA	18.38	13.53	4.18	3.50	0.001
Economy & Money	13.85	14.13	4.94	4.67	0.856
Mean Tokens	19.13	15.49	3.71	3.50	0.004
11th Grade					
Category	M		SD		p
	Science	Economics	Science	Economics	
Animals	23.43	16.16	4.60	4.07	<0.001
Food & Drink	25.52	19.68	6.58	6.79	0.009
Sport & PA	19.29	15.21	4.54	3.83	0.004
Economy & Money	14.10	13.53	5.72	3.99	0.720
Mean Tokens	20.71	15.96	4.42	4.01	0.001

levels. First, two-way mixed ANOVAs were run to compare the effect of time on the two CLIL groups at hand, and second, a descriptive analysis was carried out on the non-thematic prompts *Animals* and *Food & Drink* and the thematic prompts *Sport & Physical Activity* and *Economy & Money* of each group across time, focusing on the number of tokens produced by each group and the increases or decreases observed.

First, the results of the two-way mixed ANOVAs revealed key differences between the two CLIL groups in terms of LA (Table 7).

As shown, the results found a significant main effect of CLIL group on the overall LAT at each time, where the science group produced more words than the economics group overall. There was no statistically significant main effect of time, with only a slight increase in the means from 10th grade to 11th grade. There was also no significant interaction between time and CLIL group; both groups improved from 10th to 11th grade.

In terms of the four individual prompts of the LAT, the results revealed a statistically significant main effect of CLIL group in the first three prompts at each time: in all cases, the science students produced more tokens than economics students. However, of extreme interest is the fact that for the prompt *Economy & Money*, there was no statistically significant main effect of CLIL group. This suggests that although the science students outperformed the economics students from 10th to 11th grade in all other categories, both groups performed similarly on this particular prompt. As was the case for the overall LAT, the results found no statistically significant effect of time on any of the individual prompts and no statistically significant interaction between time and CLIL group for any of the individual prompts.

Second, in terms of the non-thematic and thematic prompts, there was also a notable difference between the two groups. Table 8 below provides the data discussed above, comparing each group in terms of non-thematic and thematic prompts.

For science students, a much greater increase was found in the non-thematic prompts *Animals* and *Food & Drink*, with students producing an average additional 2.28 and 1.56 tokens, respectively. However, for this group the number of tokens produced in the thematic prompts *Sport & Physical Activity* and *Economy & Money* actually slightly decreased from 10th to 11th grade by an average of 0.39 and 0.56 tokens, respectively. This implies that, while science students continued to improve their general English vocabulary across time, vocabulary specific to the domains of sport

TABLE 7 Longitudinal differences in token production.

CLIL group	M		F	p	η_p^2		
	Science	Economics					
	Animals	22.02				15.25	20.48
Food & Drink	24.94	18.15	10.56	0.003	0.289		
Sport & PA	18.91	13.50	13.90	0.001	0.348		
Economy & Money	14.16	14.20	0.000	0.986	0.000		
Mean Tokens	20.01	15.27	10.35	0.003	0.285		
Time							
	M		F	p	η_p^2		
	10th Grade	11th Grade					
	C-Test English	32.01				35.04	4.51
Animals	17.99	19.28	2.83	0.104	0.098		
Food & Drink	21.23	21.86	0.322	0.575	0.012		
Sport & PA	16.30	16.11	0.076	0.785	0.003		
Economy & Money	14.02	14.34	0.158	0.694	0.006		
Mean Tokens	17.38	17.90	0.837	0.369	0.031		
CLIL group \times time							
	M				F	p	η_p^2
	10th Grade		11th Grade				
	Science	Economics	Science	Economics			
	Animals	20.89	15.10	23.17			
Food & Drink	24.17	18.30	25.72	18.00	0.703	0.409	0.026
Sport & PA	19.11	13.50	18.72	13.50	0.076	0.785	0.003
Economy & Money	14.44	13.60	13.89	14.80	1.17	0.289	0.043
Mean Tokens	19.65	15.12	20.37	15.42	0.143	0.709	0.005

and economics do not improve. This is very likely due to the fact that in 11th grade these students no longer study PE in English and do not study economics in either 10th or 11th grade. Conversely, for economics students, very little difference is observed across time in the non-thematic prompts *Animals* and *Food & Drink*, which respectively each increase and decrease by an average of 0.30. As for the thematic prompts, *Sport & Physical Activity* remains constant, with students producing the same number of tokens in each year, while *Economy & Money* increases slightly by an average of 1.20 tokens. This minor increase in this thematic prompt again suggests that there is a difference between the two groups in terms of their exposure to SSV in the TL. However, it is necessary to carry out classroom observations to gain greater understanding of these differences. In addition, by including additional, science-related prompts, it would be possible to determine whether there are also differences in terms of the vocabulary to which science students alone are exposed.

The longitudinal results thus again show a clear advantage for science students, as there was a statistically significant main effect of CLIL group on the C-test, the overall LAT and three out of the four individual prompts of the LAT at each time to the advantage of participants studying a science subject. It is unclear what has caused this advantage. Given the fact that all participants received an equal number of hours of intramural exposure to English, and the same number of CLIL classes, differences cannot be attributed to number of hours of exposure and CLIL exposure in

TABLE 8 Average tokens produced by content and language integrated learning groups across time.

	Science		+/-	Economics		+/-
	10th	11th		10th	11th	
Non-thematic						
Animals	20.89	23.17	+2.28	15.10	15.40	+0.30
Food & Drink	24.17	25.72	+1.56	18.30	18.00	-0.30
Thematic						
Sport & PA	19.11	18.72	-0.39	13.50	13.50	0.00
Economy & Money	14.44	13.89	-0.56	13.60	14.80	+1.20

school, as is often the case in CLIL research comparing CLIL and non-CLIL participants. One possible cause could be the different exposure within the one CLIL class which differed: economics or science. However, as previously noted, it would be necessary to analyze the specific input received in each class to determine whether this is indeed a factor. Another possible cause is that the science students simply had a higher language level even before CLIL began, and that perhaps those who were attracted to the science track were also more proficient English learners. For example, Rumlich (2017) highlights that there may be large differences between CLIL learners in difference streams even before classes start. Further longitudinal analysis would be required to confirm this, including a pre-test prior to CLIL classes. In addition, it would be beneficial to include other language testing measures in order to determine whether the C-test and LAT were simply more suited to the science participants. Although other data from the study indicates a female advantage, it is unlikely that this has had an effect here, as the groups were similarly distributed in terms of gender. Future research should endeavor to clarify what causes these differences and, more importantly, attempt to compare more evenly matched groups in terms of language level. This would allow us to better compare the groups in terms of LA and the SSV which is of interest.

Despite the science students' advantage in terms of language level and three out of four prompts, there is again similar performance between the groups in the content-relevant prompt, *Economy & Money*, which may well be attributed to the economics students' exposure to SSV in their CLIL class. These findings thus support the supposition that CLIL instruction has a positive effect on students' SSV (Dalton-Puffer, 2008; Heras & Lasagabaster, 2015) and are consistent with the aforementioned research by Rieder-Bünemann et al. (2022), which found a positive effect of CLIL in acquiring SSV.

Additional analysis of token production in thematic and non-thematic prompts also indicates that the science students show greater improvement in terms of more general as opposed to the SSV under analysis. While students in both groups attended the same English foreign language classes, it should be recalled that science students performed significantly better than economics students in terms of language proficiency. Regarding the relationship between language level and productive vocabulary growth, previous CLIL research has, however, indicated that lower-level students may improve their vocabulary to a greater extent. For example, one study by Alejo González and Piquer Píriz (2016) found that while there was a statistically significant difference in the productive vocabulary of lower-level students, there was no such difference for higher-level students. If it is the case that there is some kind of ceiling effect at play, and that lower-level students may have more room for improvement in terms of vocabulary growth, it is surprising that in this study it is the higher-level science students who produce more general vocabulary at the second data collection. Another possible explanation could be the students' extramural English language contact, as previous research has generally found a positive relationship between language contact outside the classroom and vocabulary knowledge (Warnby, 2022). While the extramural language contact of the students in this study was not addressed, this would be a very interesting area to address in future research, so as to better understand whether this has an impact on students' non-thematic and thematic vocabulary.

6 | CONCLUSION

The results of this study offer interesting insights into CLIL instruction in terms of how teaching economics through English affects the prompt *Economy & Money*. Of particular interest is the fact that a CLIL approach appears to alleviate the linguistic gap between the two groups of students under analysis in this specific lexical domain. Despite the fact that science students had a higher language level and performed better in other lexical domains, by studying economics through English the economics students have effectively evened the playing field in this content-related prompt, *Economy & Money*.

These results showcase the ability of the LAT to assess SSV in an extremely convenient manner. However, there are a number of limitations which should be addressed in future research. First, regarding the prompts under analysis, while this study has taken some preliminary steps toward answering Canga Alonso's (2017) call for using content-related prompts, it is suggested that future research would benefit from linking prompts to the individual units which have been studied in the students' CLIL classes, so as to home in on the specific target vocabulary in the participants' curriculum and target particular topics within the broader domains of science and economics. This could be done by performing an analysis on the participants' textbooks and catering the task to the findings. In addition, the present study included just two prompts related to the participants' English CLIL subjects (*Sport & Physical Activity* and *Economy & Money*); however, it would evidently be useful to include prompts which address other subjects studied through English. In particular, including prompts related to the participants' science subjects would have allowed for a more robust comparison of the groups under analysis.

Second, regarding the participants themselves, the study included a relatively small sample size from just two semi-private schools. This may have affected the results and explain why no statistically significant differences were observed in some of the comparisons across time. Future research is clearly needed across a wider sample to test the findings and to confirm whether these results may be generalized to other populations. In particular, as the participants came from semi-private schools, it is also possible that this had an effect on the results, given the link between socio-economics status and L2 achievement (Ellis, 1994).

A final issue concerning the participants is the differences observed in the language level of the science and economics students. As previously mentioned, it is unclear whether these differences would have been found prior to CLIL, or whether the specific CLIL classes taken played a role. It may also be possible that the instruments used were better suited to those in the science group, perhaps due to different learning styles. It is important that this is taken into account in future research, so as to disentangle the multiple factors which may be affecting the results. It is also important to attempt to compare groups with similar language proficiencies in future research.

As has been highlighted by Nation (2001), there is a complementary relationship between vocabulary knowledge and language use: vocabulary knowledge allows learners to use the language, while language use allows students to increase their vocabulary knowledge. It can be suggested that this is all the more relevant in a CLIL context, given that the TL is used to explain key concepts related to the content which is being studied. As a result, studying content through the TL has the potential to expand subject-related vocabulary, but it is also necessary for students to know this vocabulary in order to participate fully and thrive in a CLIL learning environment. It is thus essential that policy makers understand how to best aid students in acquiring this vocabulary and to have resources available to them to monitor and test this acquisition in a convenient way. This study shows that the LAT provides a useful means of achieving this goal and clearly shows the advantage that CLIL has in fostering the acquisition of SSV.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Dr Leah Geoghegan  <https://orcid.org/0000-0003-3084-0927>

REFERENCES

- Agustín Llach, M. P., & Canga Alonso, A. (2016). Vocabulary growth in young CLIL and traditional EFL learners: Evidence from research and implications for education. *International Journal of Applied Linguistics*, 26(2), 211–227.
- Agustín Llach, M. P., & Fernández-Fontecha, A. (2014). Lexical variation in learners' responses to cue words: The effect of gender. In R. M. J. Catalán (Ed.), *Lexical availability in English and Spanish as a second language* (pp. 69–81). Springer.
- Alejo González, R., & Piquer Piriz, A. M. (2016). Measuring the productive vocabulary of secondary school CLIL students: Is Lex30 a valid test for low-level school learners? *Vigo International Journal of Applied Linguistics*, 13, 31–54.
- Baten, K., van Hiel, S., & de Cuypere, L. (2020). Vocabulary development in a CLIL context: A comparison between French and English L2. *Studies in Second Language Learning and Teaching*, 10(2), 307–336. <https://doi.org/10.14746/ssllt.2020.10.2.5>
- Canga Alonso, A. (2017). Spanish L1 speakers' and EFL learners' available lexicon. *Anuario de estudios filológicos*, 40, 5–23. <https://doi.org/10.17398/2660-7301.40.5>
- Canga Alonso, A., & Arribas García, M. (2015). The benefits of CLIL instruction in Spanish students' productive vocabulary knowledge. *Encuentro: Revista de investigación e innovación en la clase de idiomas*, 24, 15–31.
- Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism*, 10(5), 543–562. <https://doi.org/10.2167/beb459.0>
- Daller, H., & Phelan, D. (2006). The C-test and TOEIC® as measures of students' progress in intensive short courses in EFL. In R. Grotjahn (Ed.), *Der C-Test: Theorie, Empirie, Anwendungen/The C-test: Theory, empirical research, applications* (pp. 101–119). Peter Lang.
- Daller, H., van Hout, R., & Daller-Treffers, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24, 197–222. <https://doi.org/10.1093/applin/24.2.197>
- Dalton-Puffer, C. (2008). Outcomes and processes in content and language integrated learning (CLIL): Current research from Europe. In W. Delanoy & L. Volkman (Eds.), *Future perspectives for English language teaching* (pp. 139–157). Carl Winter.
- Dalton-Puffer, C., Llinares, A., Lorenzo, F., & Nikula, T. (2014). "You can stand under my umbrella": Immersion, CLIL and bilingual education. A response to Cenoz, Genesee & Gorter (2013). *Applied Linguistics*, 35(2), 213–218. <https://doi.org/10.1093/applin/amu010>
- Dalton-Puffer, C., Nikula, T., & Smit, U. (2010). *Language use and language learning in CLIL classrooms: Current findings and contentious issues*. John Benjamins.
- Dalton-Puffer, C., & Smit, U. (2007). Introduction. In C. Dalton-Puffer & U. Smit (Eds.), *Empirical perspectives on CLIL classroom discourse* (pp. 7–23). Peter Lang.
- Diéguez, K., & Martínez-Adrián, M. (2017). The influence of CLIL on receptive vocabulary: A preliminary study. *Journal of English Studies*, 15, 107–134.
- Dimitrijević, N. (1969). *Lexical availability*. Julius Gross Verlag.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University.
- Fernández Orío, S., & Jiménez Catalán, R. M. (2015). Lexical availability of EFL learners at the end of Spanish Secondary Education: The effect of language program and prompt. *ES Revista de Filología Inglesa*, 36, 103–127. <http://uvadoc.uva.es/handle/10324/22681>
- Fernández-Fontecha, A. (2021). The role of learner creativity in L2 semantic fluency. An exploratory study. *System*, 103, 102658.
- Fernández-Fontecha, A., Jiménez Catalán, R. M., & Ryan, J. (2021). Lexical production and organisation in L2 EFL and L3 EFL learners: A distributional semantic analysis of verbal fluency. *International Journal of Multilingualism*. <https://doi.org/10.1080/14790718.2021.2016770>
- Geoghegan, L., & Agustín Llach, M. P. (2023). Theoretical exploration of the lexical availability task as a tool for the selection of vocabulary to be taught in class. *International Journal of TESOL Studies*, 5(2), 5–18. <https://doi.org/10.58304/ijts.2023.0202>
- Grotjahn, R., & Stemmer, B. (2002). C-tests and language processing. In J. A. Coleman, R. Grotjahn, & U. Raatz (Eds.), *University language testing and the C-test* (pp. 115–130). AKS-Verlag.
- Grujić, T., & Danilović, J. (2012). *The value of C-tests in English language testing and teaching*. Paper presented at Multidisciplinary Conferences Language, Literature, Values, Faculty of Philosophy, University of Niš, Serbia.
- Heras, A., & Lasagabaster, D. (2015). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*, 19(1), 70–88. <https://doi.org/10.1177/1362168814541736>
- Hernández-Muñoz, N., Izura, C., & Ellis, A. W. (2006). Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology*, 18(5), 730–755.
- Jiménez Catalán, R. M. (2023). Exploring the productive vocabulary of EFL learners through lexical availability. *International Journal of TESOL Studies*, 5(2), 49–63. <https://doi.org/10.58304/ijts.20230205>

- Jiménez Catalán, R. M., & Ojeda Alba, J. (2009). Disponibilidad léxica en inglés como lengua extranjera en dos tipos de instrucción. *Lenguaje y Textos*, 30, 167–176.
- Jiménez Catalán, R. M., & Fernández-Fontecha, A. (2019). Lexical availability output in L2 and L3 EFL learners: Is there a difference? *English Language Teaching*, 12(2), 77–87. <https://doi.org/10.5539/elt.v12n2p77>
- Jiménez Catalán, R. M., & Fitzpatrick, T. (2014). Frequency profiles of EFL learners' lexical availability. In R. M. J. Catalán (Ed.), *Lexical availability in English and Spanish as a second language* (pp. 83–100). Springer.
- Jiménez Catalán, R. M., & Agustín Llach, M. P. (2017). CLIL or time? Lexical profiles of CLIL and non-CLIL EFL learners. *System*, 66, 87–99. <https://doi.org/10.1016/j.system.2017.03.016>
- Lo, Y., & Macaro, E. (2015). Getting used to content and language integrated learning: What can classroom interaction reveal? *The Language Learning Journal*, 43(3), 239–255. <https://doi.org/10.1080/09571736.2015.1053281>
- Marsh, D. (2002). *CLIL-EMILE the European dimension: Actions, trends and foresight potential*. European Commission.
- Matiasek, S. (2005). *English as the language of instruction in Austrian chemistry lessons: Instances of explicit language teaching* (Unpublished master's thesis). University of Vienna, Austria.
- McManus, K. (2011). *The development of aspect in a second language* (Unpublished doctoral dissertation). Newcastle University.
- Merikivi, R., & Pietilä, P. (2014). Vocabulary in CLIL and in mainstream education. *Journal of Language Teaching and Research*, 5(3), 487. <https://doi.org/10.4304/jltr.5.3.487-497>
- Nation, I. S. P. (2016). *Making and Using Word Lists for Language Learning and Testing*. John Benjamins.
- Nikula, T. (2010). Effects of CLIL on a teacher's classroom language use. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 105–124). John Benjamins.
- Neilson Parada, M. N. (2016). *Lexical availability in diaspora Spanish: A cross-generational analysis of Chilean Swedes* (Unpublished doctoral dissertation). University of Illinois at Chicago.
- Payne, C. (2016). Lexical availability. *English Teaching Professional*, 102, 18–20.
- Rieder-Bünemann, A., Hüttner, J., & Smit, U. (2022). 'Who would have thought that I'd ever know that!': Subject-specific vocabulary in CLIL student interactions. *International Journal of Bilingual Education and Bilingualism*, 25, 3184–3198. <https://doi.org/10.1080/13670050.2021.2020211>
- Rodríguez Bonces, J. (2012). Content and language integrated learning (CLIL): Considerations in the Colombian context. *GIST: Education and Learning Research Journal*, 6, 177–189.
- Ruiz de Zarobe, Y. (2011). Which language competencies benefit from CLIL? An insight into applied linguistics research. In Y. Ruiz de Zarobe, J. M. Sierra, & F. Gallardo del Puerto (Eds.), *Content and foreign language integrated learning: Contributions to multilingualism in European contexts* (pp. 129–154). Peter Lang.
- Rumlich, D. (2017). CLIL theory and empirical reality – Two sides of the same coin? A quantitative-longitudinal evaluation of general EFL competence and affective-motivational dispositions in CLIL students at German secondary schools. *Journal of Immersion and Content-Based Language Education*, 5, 110–134. <https://doi.org/10.1075/jicb.5.1.05rum>
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
- Sylvén, L. K. (2019). Extramural English. In L. K. Sylvén (Ed.), *Investigating content and language integrated learning: Insights from Swedish high schools* (pp. 152–167). Multilingual Matters.
- Tabuenca Cuevas, M. F., & Alcaraz Mármol, G. (2014). From EFL to CLIL teaching and learning vocabulary in the primary classroom. In J. R. Calvo-Ferrer & M. A. Campos-Pardillos (Eds.), *Investigating lexis vocabulary teaching, ESP, lexicography and lexical innovation* (pp. 27–37). Cambridge Scholars Publishing.
- van Ginkel, C. I., & van der Linden, E. H. (1996). Word associations in foreign language learning and foreign language loss. In K. Sajavaara & C. Fairweather (Eds.), *Approaches to second language acquisition* (pp. 25–35). University of Jyväskylä.
- Vraciu, A., & Marsol, A. (2023). Content-specific vocabulary in CLIL: Exploring L2 learning outcomes in a primary education programme in Catalonia. *Language Teaching Research*. <https://doi.org/10.1177/13621688231170073>
- Warnby, M. (2022). Receptive academic vocabulary knowledge and extramural English involvement—Is there a correlation? *ITL-International Journal of Applied Linguistics*, 173(1), 120–152. <https://doi.org/10.1075/itl.21021.war>

How to cite this article: Geoghegan, L. (2023). The effect of CLIL on productive thematic vocabulary. *International Journal of Applied Linguistics*, 1–15. <https://doi.org/10.1111/ijal.12506>