

# The Impact of CLIL on FL Grammar and Vocabulary



Macarena Navarro-Pablo

**Abstract** This chapter examines the impact of Content and Language Integrated Learning (CLIL) on 351 Primary and Secondary Education students' English language attainment regarding grammar and vocabulary in seven bilingual public schools and one non-bilingual charter school in the province of Seville. Intergroup comparisons are carried out across the different schools and discriminant analyses are performed with all the intervening variables of the study (setting, socio-economic status, verbal intelligence, motivational variables and extramural exposition) in order to determine whether CLIL is truly responsible for the differences ascertained or whether other variables account for a greater proportion of the variance.

## 1 Introduction

CLIL is nowadays the methodological approach adopted by most European countries in order to meet the challenges posed by the European Union's multilingualism policy (European Union, 2008), which establishes that European citizens should be able to communicate in two languages other than their mother tongue. The implementation of this measure implies the need for greater levels of foreign language (FL) proficiency, which therefore calls for a revision and, probably in many countries, a thorough reform of the approaches to FL teaching and learning. Dalton-Puffer (2014) points out that CLIL is a methodological revolution, not only in the context of FL teaching, but also in the teaching of non-linguistic subjects. All over Europe, CLIL is considered 'a new learning and teaching environment' (Coonan, 2007, p. 625). In the last few decades, interest in CLIL, defined as the 'teaching and learning through a foreign language' (Marsh, 2002, p. 54), has gained momentum. In many countries, this methodology has been adopted in and adapted to different school settings. Pérez-Cañado (2012), who offers an overview of the literature on CLIL since the term was coined in Finland in the 1990s, has pointed out that in

---

M. Navarro-Pablo (✉)  
University of Sevilla, Seville, Spain  
e-mail: [mnp@us.es](mailto:mnp@us.es)

recent years new studies on CLIL have started to emerge, offering sometimes contradictory and conflicting views on this methodological approach. One controversial issue pertains to the categorisation of CLIL. For example, not all researchers agree that this is an altogether new methodological approach, but rather consider it as having evolved from different communicative methodologies already in use in FL teaching (Content-Based Language Teaching, Communicative Language Teaching or the Natural Approach L+1 Hypothesis) (Cenoz et al., 2014).

However, CLIL was not originally intended as a definitive break from preceding methodologies. Rather, as several researchers contend (Lasagabaster & Sierra, 2009; Nikula, 2007), CLIL is used as an ‘an umbrella term covering a dozen or more educational approaches’ (Mehisto et al., 2008, p. 12). Furthermore, a commonality between CLIL and preceding methods for FL teaching, or the fact that CLIL may be considered a continuation thereof, is not necessarily a negative aspect of CLIL. As suggested by Dueñas (2004), CLIL is ‘a flexible operational framework for language instruction’ and this is precisely what allows this approach to be adapted to a great variety of contexts. In this line, Turner (2013, p. 397) indicates that ‘this broad definition of CLIL allows for programmes that existed in different European countries to be redefined as CLIL programmes’. At the same time, there are also authors that identify some distinguishing features in CLIL. For example, Mehisto et al. (2008, p. 12) state that ‘What is new about CLIL is that it synthesizes and provides a flexible way of applying the knowledge learnt from these various approaches’. In this line, Coyle (2008, p. 97) establishes that the distinctiveness of CLIL lies in its ‘integrated approach, where both language and content are conceptualised on a continuum without an implied preference for either’. CLIL gives teachers the opportunity to introduce cross-curricular connections, meaningful interactions, cognitive skills training and a variety of cultural contexts in the subject content classroom, something that also goes beyond traditional non-linguistic content instruction. All these aspects that make CLIL different from other types of instruction need to be considered when analysing CLIL’s potential language benefits for learners.

## 2 Prior Research on CLIL

Research on CLIL began in different European countries in the first decade of the twenty-first century and initially focused on comparative studies of CLIL and non-CLIL learners. More recently, there has been an increasing number of studies concerning the implementation and the effects of CLIL programmes. Research on the assessment of language learning and teaching programmes in CLIL contexts entails certain difficulties due to the variety of factors that influence final learning outcomes, and to the different ways in which this methodology has been implemented in different countries. According to Lasagabaster (2015, p. 19), ‘as CLIL syllabuses are usually developed to meet local needs, there is huge variation in its implementation’. Also, Nikula (2007, p. 208) draws attention to the ‘great deal of variation

in CLIL'. In the Finnish context, Nikula (2007) show how forms of implementation may vary in terms of both depth and breadth.

Table 1 sums up and classifies some of the literature produced on CLIL and allows us to narrow down and select the most relevant sources for the present study. The first section includes research focused on theoretical assumptions concerning CLIL conceptualisation. These studies provide a conceptual framework for bilingual and plurilingual education, bilingual programme design, policy issues from the different countries where CLIL programmes are implemented, and the main challenges to CLIL implementation (see Research type 1 in Table 1). From a more practical standpoint, several studies gather data about the implementation of CLIL programmes

**Table 1** Types of research

Research type	Author/s, year of publication
Research type 1 (RT1): CLIL conceptualisation	Marsh (2002) Dueñas (2004) Nikula (2007) and Nikula et al. (2016) Coyle (2008) Mehisto et al. (2008) Turner (2013) Cenoz et al. (2014) Dalton-Puffer (2011, 2014) Pérez-Cañado (2012) Ruiz de Zarobe (2017)
Research type 2 (RT2): CLIL Implementation. Stakeholders' beliefs and opinions	Coonan (2007) Dalton-Puffer (2011) Pérez-Cañado (2014)
Research type 3 (RT3): CLIL and English learning outcomes	Admiraal et al. (2006) Ruiz de Zarobe (2008) Agustín-Llach (2009) Dalton-Puffer (2009) Fernández Fontecha (2010) Martínez and Juncal Gutiérrez (2009) Moreno (2009) Ojeda (2009) Ruiz de Zarobe and Jiménez-Catalán (2009) San Isidro (2010) Villarreal and García (2009) Navés and Victori (2010) Lasagabaster and Ruiz de Zarobe (2010) Llinares and Dafouz (2010) Breidbach and Viebrock (2012) Agustín-Llach and Canga (2014) Lasagabaster (2015) Juan-Garau et al. (2015) Roquet and Pérez-Vidal (2015) Pérez-Vidal and Roquet (2015) Rumlich (2016)

(continued)

**Table 1** (continued)

Research type	Author/s, year of publication
Research type 4 (RT4): Individual and Contextual variables	Dalton-Puffer (2007) Lasagabaster and Sierra (2009) Lasagabaster (2011) Llinares et al. (2012) Martín del Pozo (2013, 2015) Fernández and Canga (2014) Dafouz and Hibler (2013) Doiz et al. (2014) Arribas (2016) Pfenninger (2016) Lorenzo (2017) Navarro-Pablo and García-Jiménez (2018) Fernández-Sanjurjo et al. (2019) Madrid and Barrios (2018) Rascón and Bretones (2018) Alejo and Piquer (2016) Pavón (2018) Lancaster (2018) Lorenzo et al. (2019)

in Europe and their strengths and weaknesses, either through participants' opinions, classroom observation or both (Research type 2). Most of these are qualitative studies in which interviews, questionnaires or observation tools are used to analyse the effectiveness of CLIL teaching performance or stakeholders' beliefs and opinions on the matter. Their findings usually provide positive views on language learning in CLIL programmes. The third group is made up of studies that analyse students' English learning outcomes, making use of quantitative data from students' results in language tests (Research type 3). Finally, there is an increasing number of studies that focus on individual and contextual variables that affect language learning, such as motivation or extramural exposure (see Research type 4).

Research on CLIL characterisation (RT1) helps to understand the differences between CLIL and non-CLIL instruction. Studies concerning CLIL teaching performance and stakeholders' beliefs and opinions (RT2) offer information about the way these programmes are being implemented, and their effect on the participants. Studies that analyse the implementation of CLIL programmes comparing the language results of CLIL and non-CLIL groups (RT3) are especially relevant for this study as they provide pertinent empirical evidence in the field. Finally, the present study also considers the possible influence of different individual and contextual variables (setting, socioeconomic status, verbal intelligence, motivational variables, extramural exposure and language needed and produced in class) on students' linguistic performance. Agreeing with many researchers (RT4) on the importance of moderating variables, this study includes a discriminant analysis in order to determine the influence of these variables on linguistic outcomes.

In general, research provides positive results for CLIL strands, where CLIL learners outperform non-CLIL ones. However, there are also some conflicting and not fully comparable findings, for which some factors are accountable, namely:

- Context (country and region; monolingual or bilingual context) and languages involved: it is necessary to study bilingual and monolingual regions independently, since their experience with bilingual and immersion programmes is so different that it can affect students' results and hence research findings. In Spain, where the present study takes place, it is necessary to bear in mind the local context, as there are bilingual regions, such as the Basque Country or Catalonia, in which the language introduced through CLIL (mostly English) is a third language, whereas in monolingual regions, English tends to be the second language.
- Educational level: results vary depending on whether research takes place in Primary, Secondary or Higher Education.
- Language tests and competence categorisation: when analysing and comparing students' results, the same types of tests need to be used in order to measure the same linguistic aspects. This has not always been the case, and researchers have measured students' linguistic competence using different types of tests, thus placing the focus on some linguistic aspects over others (for example, skills or specific language components, such as grammar or vocabulary).
- Type of study (cross-sectional vs. longitudinal study): studies have tended to focus on students' results at a given point in time. More recently, longitudinal studies are starting to gain ground. In this way, the effectiveness of CLIL programmes can be measured throughout time, offering a wider perspective on CLIL (Dalton-Puffer, 2011).
- Language needed and produced in CLIL classrooms: when analysing students' linguistic outcomes, teachers' metalanguage and classroom discourse functions are essential to the teaching and learning of content curricular subjects. The BICS and CALP distinction made by Cummins (2008), the *Language Triptych* (Language of Learning, Language through Learning and Language for Learning) identified by Coyle et al. (2010) and the *LAC model* (General Language, Academic Language, Subject/Domain Specific Language and Classroom Language) by Gierlinger (2013) provided a theoretical framework for the analysis of the different languages involved in CLIL classroom.

Bearing these factors in mind, this chapter presents the results of a study carried out in a monolingual region (Andalusia) at both Primary and Secondary level. The present analysis focuses on quantitative data obtained from a linguistic test administered to 351 Primary and Secondary Education students (more information on the data collection tools in the Methodology section below). More specifically, it focuses on learners' results regarding two specific linguistic aspects: grammar and vocabulary.

The last decade has seen a proliferation of research that disaggregates and compares CLIL learners' performance in specific skills and/or mastery of linguistic

components, rather than just drawing on their overall achievement levels. Most findings indicate that when specific skills or linguistic aspects are analysed, or other variables (individual, motivational or contextual) are introduced, results are not consistent (Ojeda, 2009; Roquet & Pérez-Vidal, 2015; Ruiz de Zarobe, 2008). This may be due to the fact that these studies measure aspects such as vocabulary or grammar, which are generally emphasised in regular EFL courses, as opposed to the emphasis placed on communication in CLIL. In spite of this, there is a general consensus in the literature regarding CLIL programmes' positive effects on learners' language development over those of regular EFL programmes.

The literature review that is offered in the following lines focuses on two different types of studies: longitudinal and cross-sectional ones. Then, some studies dealing specifically with vocabulary and grammar are considered.

Longitudinal studies on the long-term benefits of CLIL instruction are emerging progressively (cf. also Chapter 8 in this volume). In this line, Pérez-Vidal and Roquet (2015) provide empirical data from two studies where writing, reading and listening skills, together with lexico-grammatical abilities, are examined. Their findings indicate that 'larger relative gains are obtained by the FI + CLIL programme, however not in all domains and to the same degree' (Pérez-Vidal & Roquet, 2015, p. 80). As regards students' lexico-grammatical abilities, CLIL learners show higher relative gains, whereas results regarding only the writing skill indicate that 'the superiority of CLIL cannot be confirmed', since 'although improvement in the case of the FI + CLIL group is shown, results were only significant in the domain of accuracy' (Pérez-Vidal & Roquet, 2015, p. 1). The findings of their quantitative data reveal that CLIL learners' progress in syntactic and lexical complexity as well as fluency is better, although the differences between CLIL and non-CLIL strands are not statistically significant. The same results are found through the qualitative data they collect for grammar and vocabulary. Similarly, Juan-Garau et al. (2015) also research the impact of CLIL programmes by analysing CLIL learners' development of lexico-grammatical accuracy over a period of three years throughout Secondary Education. They conclude that CLIL and non-CLIL learners 'significantly improved their overall longitudinal lexical and grammatical ability' (2015, p. 189) and their results also suggest that lexico-grammatical development is faster in CLIL learners.

Turning now to cross-sectional studies, Llinares and Dafouz (2010) describe the UAM-CLIL project carried out at secondary school level in Madrid, where three types of data were collected (whole class discussion, written composition and oral interview) from a corpus of approximately 40,000 spoken and 25,000 written words. Dealing with both lexical and grammatical competence, and comparing the same students' written and spoken performance, their study indicates that CLIL 'students use appropriate lexis to express content-specific ideas' (2010, p. 106). Similarly, San Isidro (2010) analyses language competence improvement in Secondary schools considering three main variables: student type, gender and school type. It includes specific sections measuring grammar and vocabulary development where 'the examiners assessed the ability to use vocabulary, structure and paraphrase strategies to convey meaning' (2010, p. 67). Findings reveal that CLIL learners 'were able not

only to pass an objective skills-based test but also with much better results' (2010, p. 67).

Examples of both longitudinal and cross-sectional studies carried out in Primary and Secondary Education are provided by Navés and Victori (2010). They present two studies as part of the BAF project that aim at examining the effect of onset age in the acquisition of English as a foreign language. The studies compare students' marks in two Primary and three Secondary schools in Catalonia. The first study focuses on learners' general language proficiency and includes students from Years 5, 7, 8 and 9 in CLIL and non-CLIL strands. Cross-sectional findings reveal that CLIL learners outperform non-CLIL learners on all the tests administered: grammar, cloze, dictation and listening. Furthermore, for all the measures analysed, longitudinal results indicate younger learners in CLIL strands obtain similar results to those of the older EFL students (2010, p. 47). The second study analyses learners' writing skills from Years 5 to 12. Similarly, there were statistical significant differences for each of the areas tested: fluency, accuracy and syntactic and lexical complexity. Again, CLIL learners at lower levels performed better in writing than the older EFL students. Likewise, Ruiz de Zarobe (2008) measures the oral competence of Secondary school students. In order to do so, five categories are established: pronunciation, vocabulary, grammar, fluency and content. On the one hand, the cross-sectional analysis reveals that CLIL students perform significantly better in all the scales. On the other hand, the longitudinal study reveals that, after one year, the CLIL group also outperforms the non-CLIL one in all the categories. After two years, however, differences between both groups are not statistically significant in the vocabulary category.

Finally, we focus on research dealing specifically with vocabulary and grammar. Regarding vocabulary, some studies of this kind are described in Ruiz de Zarobe and Jiménez-Catalán (2009). The sample in all of them is comprised of learners in Year 6 of Primary Education. As regards lexical competence, Jiménez-Catalán and Ruiz de Zarobe aim at establishing connections between the type of instruction and its effect on FL vocabulary acquisition, based on the hypothesis that 'the type of language instruction relates positively to vocabulary knowledge' (2009, p. 82). The study contrasts the results of CLIL and non-CLIL groups in two receptive vocabulary tests, and concludes that, in both tests, CLIL learners outperform their non-CLIL counterparts. A further study is provided by Moreno (2009), who analyses the results of a free word association task as a means to explore learners' mental lexicon, assuming that this type of data can 'complement and corroborate findings that emerge from analyses of other lexical data' (2009, p. 93). One of the main purposes of this study is to describe the characteristics of the productive lexical profile of EFL Spanish learners in Primary Education, comparing both CLIL and non-CLIL instructional models. Findings show that CLIL learners produce more tokens and types than non-CLIL learners, which is indicative that they have a higher proficiency level. Findings also reveal that CLIL learners exhibit 'a slightly higher productive vocabulary size' and 'more lexical richness in the word association test [. . .] by recalling a higher number of infrequent words' (2009, p. 100). Thus, the study shows that statistically significant differences between CLIL and non-CLIL groups apply to both vocabulary size and vocabulary depth. However, the author also contends that CLIL learner results

'are not so overwhelming if we compare the great difference of formal instruction exposure between groups' (2009, p. 106).

In a similar line, the role of the L1 (Spanish) in CLIL and non-CLIL learners' FL vocabulary use is analysed by Agustín-Llach (2009). Taking into account variables such as students' proficiency levels, amount of FL exposure and instructional approach, the aim of the study is to reach conclusions regarding which group of students (CLIL or non-CLIL) has more transfer episodes from Spanish to English. In order to do so, three categories are established: borrowings, coinages and calques. Findings reveal that non-CLIL learners transfer from their L1 more frequently than their CLIL counterparts, showing a higher number of lexical errors for all categories distinguished. In another study, Agustín-Llach and Canga (2014) perform a cross-sectional and longitudinal analysis of Primary school learners' FL receptive vocabulary size and lexical growth. Whereas in the cross-sectional analysis, CLIL learners have a slightly larger vocabulary size, the longitudinal study reveals that differences between groups are not significant in the early years, and become progressively significant in the later years. Therefore, the study points to growing differences in favour of CLIL students as learners' educational and proficiency levels increase.

Also, Ojeda (2009) provides an interesting study on vocabulary and themes, drawing attention to the importance of considering the differences between the CLIL methodology and regular FL approaches, with special emphasis on CLIL and non-CLIL learners' view of the target language. In this regard, learners following an FL programme conceive the target language 'as a single object for language learning' (2009, p. 130) where the language is organised around linguistic components (grammar and vocabulary among others), whereas in CLIL language instruction is organised around non-linguistic topics and lessons, being the FL a vehicular language used for communication and for conveying meaning. The study analyses learners' written compositions in order to compare 'the vocabulary most frequently implemented by the two samples of participants' (2009, p. 132). The author draws upon the similarities and differences found in a total of 60 comparable texts following a lexical field theory where taxonomy serves to classify and organise the lexis found in the corpus. Findings are mixed concerning the similarities and discrepancies of both groups, depending on the lexical field analysed. In the interpretation of the results, the author suggests that the 'CLIL sample seems to have a slight tendency to use a wider range of types including both colloquial and even sophisticated words' (2009, p. 137) and refers to the 'non-CLIL sample's greater difficulties to express abstract ideas that entail a higher degree of complexity' (2009, p. 152). At the same time, Ojeda points to the non-CLIL group's 'higher lexical reiteration' (2009, p. 140) and CLIL learners' 'higher lexical variation' (2009, p. 153).

There are also some qualitative studies where stakeholders confirm the results reported in quantitative research concerning vocabulary improvement in CLIL programmes. The results in Pérez-Cañado (2014) are largely consistent with findings reported in Juan-Garau et al. (2015). Pérez-Cañado analyses data gathered from questionnaires for a European study in which in-service teachers across Europe provide their perceptions on teacher training needs for bilingual education. One of the thematic blocks considered in the study is related to participants' current level of



linguistic and intercultural competence. Findings reveal that ‘all the items comprised within linguistic and intercultural competence are invariably considered to be appropriately mastered’ and, more interestingly for our study, that ‘this is especially the case for accurate pronunciation and knowledge of specialized academic vocabulary (within linguistic competence)’ (2014, p. 11). Although studies carried out so far generally provide quite positive results for CLIL learners’ gains, research also points to the impossibility of ascertaining that findings are due to the type of instruction, rather than, for example, to the increased number of hours that CLIL programmes imply (Juan-Garau et al., 2015; Ruiz de Zarobe, 2017).

However, not all the studies show such satisfactory results. For instance, Admiraal et al. (2006), who carry out an evaluation of bilingual education at Secondary school level in The Netherlands, administer a test that specifically measures receptive vocabulary, where findings show that there are no significant differences between bilingual and non-bilingual groups. Similarly, Fernández Fontecha (2010) provides results from a lexical availability task administered to Year 6 Primary school learners in which non-CLIL students outperform CLIL ones. The findings are interpreted as due to factors such as ‘the type of test used, which requires that the learners produce types in a limited amount of time not in a communicative interaction, which is more typical of a CLIL environment [ . . . ] or the early stage of CLIL instruction at which learners had been tested’ (2009, p. 87).

Turning now to research specifically focused on grammar development, Breidbach and Viebrock (2012) review recent CLIL research in Germany. Especially relevant for the present study is Berenbröker, as described by Breidbach and Viebrock (2012), where a comparative study between 195 CLIL and non-CLIL learners over a period of two years shows that, whereas CLIL has a very positive influence on FL competence in general, as regards grammar, differences are less accentuated. According to the study, this is due to the fact that ‘regular foreign language teaching is often more concerned with an explicit focus on grammar, whereas CLIL is more concerned with implicit grammatical knowledge, which is acquired in the process of exchanging subject-specific information’ (2000, p. 7). In turn, tense and agreement morphology in Secondary Education is analysed through a collection of oral narrations by Villarreal and García (2009). They compare affixal forms against suppletive forms. They find that the omission rate is very high across both groups of learners (CLIL and non-CLIL), which implies a parallel behaviour of the groups taken independently. However, when their overall performance is contrasted, the CLIL group outperforms the non-CLIL group in the production of affixal morphemes. Quite similarly, Martínez and Gutiérrez (2009) study the acquisition of syntax, also through the analysis of Secondary students’ oral narrations. In order to do so, they select several morpho-syntactic features, null subjects, production of placeholders, negation and production of null objects, and conclude, ‘CLIL learners significantly outperform non-CLIL learners only in the use of placeholders’ (2009, p. 193).

As has been shown, findings regarding specific linguistic aspects are not clear-cut. Indeed, Juan-Garau et al. (2015) state that ‘no conclusive results have so far been obtained regarding the development of lexico-grammatical competence in CLIL

contexts' (2015, p. 182). At the same time, Pérez-Vidal and Roquet (2015, p. 81) contend that:

[G]eneral results [concerning the linguistic benefits of CLIL] seem to be by and large positive, although there are aspects which are either unaffected by CLIL or for which research is in-existent or inconclusive, namely syntax, productive vocabulary, written accuracy, discourse skills and pragmatic efficiency (see Llinares et al. 2012), and pronunciation, that is, degree of foreign accent. Such a positive impact has generally been attributed to higher quantity and quality of exposure. However, methodological issues are still unresolved in CLIL research and subject to debate.

The following conclusions may be extracted from this literature review:

- More research is needed in order to shed light on findings that are apparently contradictory or inconclusive. Any teaching approach needs time and fine-tuned research so that its theoretical bases can feed the results of research and produce better learning outcomes.
- CLIL is an approach that seeks contextualised learning based on meaning construction through the use of the target language. It is also a methodology that is eminently grounded on interaction and communication. Therefore, research instruments that measure learners' mastery of specific linguistic systems (typically trained in regular EFL courses) are likely to yield less positive results in CLIL groups. In view of this, specific instruments measuring learners' communicative competence and contextualised learning should be used.
- In order to design accurate research tools, it is also necessary to take into account the characteristic tasks of each instructional approach. In this sense, CLIL learners' vocabulary seems to be best measured by means of integrative vocabulary tests and word association tests than by discrete decontextualised receptive vocabulary tests.
- Different personal and contextual factors must be considered in any assessment of CLIL, as these have an effect on and may help explain learners' results.

### 3 The Study

**Research Questions** The aims of the present study are: (1) to examine the impact of CLIL on the English grammar and vocabulary of 351 Primary and Secondary school students and (2) to investigate the relationship between individual difference and contextual variables in order to determine which of them has a stronger influence on students' linguistic outcomes. The effects of the following intervening variables are analysed: setting, socioeconomic status, verbal intelligence, motivational variables and extramural exposition. Considering these aims, this study seeks to answer the following research questions:

- RQ1. Are there statistically significant differences between the achievement levels of CLIL and non-CLIL learners concerning grammar and vocabulary? If so, what is the effect size?

- RQ2. Are there statistically significant differences between the achievement levels of CLIL and non-CLIL learners concerning individual differences and contextual variables? If so, which variable has a stronger influence on students' linguistic outcomes?

**Scope** In order to guarantee the homogeneity of the sample, three actions were undertaken: first, the researchers contacted the provincial coordinator of bilingual programmes in order to request a list of the state schools with English bilingual programmes. The schools selected for the sample have both CLIL and non-CLIL (or regular EFL) groups, which acted as experimental and control groups, respectively. Thus, participants in this study are streamed into two different instruction types: students enrolled in CLIL programmes and students who follow an EFL approach. Secondly, verbal intelligence and motivation tests were applied to each group. Finally, information regarding students' socioeconomic status, their English grades and their extramural exposure to English were also collected. The results of these actions allowed us to match students and ensure that these factors did not interfere with the results of the study. At the same time, this also allowed us to determine whether the differences in language attainment could be ascribed to the programme implementation rather than to any other factor related to the students' initial capacities, motivation or any other contextual variables. A total of eight schools participated in the study (four Primary schools and four Secondary schools), of which seven are state schools, and one is a charter school. From the overall total of 351 students, 193 are in Year 6 of Primary Education (in the age range of 11–12) and 158 are in the final year of Compulsory Secondary Education.

**Instruments** A total of four instruments were administered to the students: three tests (English, verbal intelligence and motivation tests) and a questionnaire. The language tests were designed following the *Common European Framework of Reference for Languages (CEFR)*, the national Decrees and the regional Orders which establish the official curriculum for the educational stages assessed. The sections of the tests measuring learners' grammatical competence, both at Primary and Secondary level, combine traditional formal activities with exercises that require understanding meaning within a context. Similarly, the tests designed to measure students' lexical competence combine activities focused on form with exercises based on texts in which meaning has to be reconstructed from the context. Thus, in both instances, tests are suitable to assess lexico-grammatical competences in methodological approaches, CLIL and regular EFL courses. Verbal intelligence was measured by means of two different adapted versions (one for Primary students and one for Secondary students) of the EFAI (*Evaluación Factorial de las Aptitudes Intelectuales*) test (Santamaría, 2018). Both versions include analogies, antonyms and odd-one-out questions, of which students have to answer as many as possible in five minutes. In order to measure motivation, Pelechano's (1994) MA test was used. This test is composed of 35 items that isolate four motivational factors related to achievement and anxiety: (1) desire to work and self-esteem (comprising 10 items); (2) anxiety in the face of exams (composed of 9 items); (3) lack of interest in studying (made up of 9 elements); and (4) realistic personal self-demand (comprising

7 elements). An initial questionnaire provided personal data and information on students' socioeconomic status and extramural exposure to English.

**Data analysis** The data collected have been statistically analysed using SPSS 24.0. In order to answer RQ1, the t-test was used to identify significant differences between the two groups under study. Also, the effect size was calculated through Cohen's d coefficient. Finally, a discriminant analysis has been applied in order to address RQ2, as it can be considered a powerful technique for examining differences between the two groups with respect to several variables simultaneously. The grouping variable selected for the data analysis is the type of instructional programme followed by the students (CLIL vs. non-CLIL). The dependent variable is the students' results in both the grammar and vocabulary tests. The independent variables are setting, socioeconomic status, verbal intelligence, motivational variables and extramural exposure.

## 4 Results

Regarding the first research question, Table 2 presents the means scores, standard deviations and Cohen's d coefficient with the effect size for the grammar and vocabulary variables. Findings show that there are significant differences between CLIL and non-CLIL groups, both at Primary and Secondary level, in favour of CLIL groups. In Primary Education, the effect size is small (Cohen's d = 0.336) for grammar and medium (Cohen's d = 0.504) for vocabulary. Cohen's d sizes indicate that there are standard deviations of 0.336 and 0.504 between CLIL and non-CLIL groups, respectively. Both are considerably higher in Secondary Education, where both grammar and vocabulary differences show a large effect size (Cohen's d = 1.150 and 0.858), with standard deviations of 1.150 and 0.588 for CLIL and non-CLIL groups, respectively.

If we now take a closer look at the exercise type students were asked to answer, the t-test was used in order to determine whether learners' results vary depending on the task they had to carry out. In Primary Education (see Table 3), CLIL learners

**Table 2** Means, standard deviations, Cohen's d coefficient and effect size

Educational Level	Independent variables	Non-CLIL		CLIL		Cohen's d	Effect size
		Mean	SD	Mean	SD		
Year 6 Primary Education	Grammar	9.084	5.600	10.900	5.208	0.336	0.166
	Vocabulary	5.930	3.309	7.567	3.180	0.504	0.245
Year 4 Secondary Education	Grammar	16.857	10.587	28.712	10.0104	1.150	0.499
	Vocabulary	6.825	3.420	9.661	3.187	0.858	0.394

*SD* standard deviation

**Table 3** T-test exercise type. Primary Education

Exercises	Group	N	Mean	Sig. (bilateral)
A. Write questions for these answers	CLIL	35	2.89	.045 <sup>a</sup>
	Non-CLIL	59	2.19	
B. Look at the picture on the right. Read and complete. Use: ON, IN, BY, WITH, UNDER	CLIL	59	2.44	.024 <sup>a</sup>
	Non-CLIL	115	1.90	
C. Match the numbers with the letters	CLIL	61	3.00	.175
	Non-CLIL	121	2.61	
D. Complete using: MY, ITS, HIS, THEIR	CLIL	58	2.55	.006 <sup>a</sup>
	Non-CLIL	120	1.89	
E. Complete using: DO, DOES, DID, AREN'T, ISN'T, IS	CLIL	56	3.27	.007 <sup>a</sup>
	Non-CLIL	117	2.45	
F. Complete using: FEEDS, CHASES, DRIVES, GETS UP	CLIL	49	2.96	.046 <sup>a</sup>
	Non-CLIL	116	2.50	
G. Write the parts of the body	CLIL	34	4.62	.099
	Non-CLIL	26	3.96	
H. Read and complete. Use: A GIRL, RIGHT, HERE, LOVELY, NAME	CLIL	53	3.96	.002 <sup>a</sup>
	Non-CLIL	102	3.10	

(<sup>a</sup>p = 0.05)

outperform their peers on all exercises. From the eight exercises testing lexical and grammatical competence, there are six in which statistically significant differences are found, and only two where differences are not statistically significant. Interestingly, neither of these involves the use of contextualised vocabulary within a meaningful text.

In Secondary Education, our results for grammar are consistent with those obtained at Primary level, as CLIL learners also show better results, with statistically significant differences on all the activities (see Table 4). However, regarding vocabulary, although CLIL learners also outperform non-CLIL learners, differences between both groups are not statistically significant.

Turning now to research question 2, Table 5 shows the results of the test of equality of CLIL and non-CLIL group means in Primary Education. It allows us to examine whether significant differences exist between the groups, in terms of predictor variables. Wilks' lambda reveals that the discriminant function is statistically significant only for three variables: setting ( $F = 10.897$  and  $p$ -value = 0.001), vocabulary ( $F = 5.279$  and  $p$ -value = 0.024) and verbal intelligence ( $F = 4.069$  and  $p$ -value = 0.046). In the case of these variables, the null hypothesis is rejected ( $p$ -value < 0.05). The setting and verbal intelligence variables show higher means for non-CLIL groups; however, the results for grammar and vocabulary are better for CLIL groups. There are no statistically significant differences for the rest of the variables. Results from Box's test of Equality of Covariance Matrices ( $F = 1.398$ ,  $p$ -value = 0.012 < 0.05)

**Table 4** T-test exercise type. Secondary Education

Exercises	Group	N	Mean	Sig. (bilateral)
A. Rewrite these questions using the correct word order	CLIL	61	2.1311	.001 <sup>a</sup>
	Non-CLIL	81	1.3951	
B. Change these sentences into the active. Example: Pictures are transmitted by television. Television transmits picture	CLIL	51	2.2353	.044 <sup>a</sup>
	Non-CLIL	57	1.6491	
C. Complete the text with the verbs in the past tense	CLIL	54	8.1296	.003 <sup>a</sup>
	Non-CLIL	66	6.7879	
D. Complete these sentences with: anything/something; tell/say (in the right form); bring/take (in the right form)	CLIL	51	6.4706	.000 <sup>a</sup>
	Non-CLIL	56	4.2321	
E. Underline the right words in brackets	CLIL	64	6.3438	.001 <sup>a</sup>
	Non-CLIL	89	5.3034	
F. Complete this text with the correct forms—the PAST (for example, went) or the PRESENT PERFECT (for example, has gone)—of the verbs in brackets	CLIL	46	5.2391	.026 <sup>a</sup>
	Non-CLIL	61	4.3115	
G. Complete this text using one of the following words and expressions: ON YOUR OWN, OVERCOME, QUIT, CUT DOWN, IMPROVE, HEALTH, WILL POWER, HARM	CLIL	46	3.3261	.086
	Non-CLIL	50	2.6200	
H. Match symbol and text	CLIL	61	7.3279	.000 <sup>a</sup>
	Non-CLIL	80	6.3250	

(<sup>a</sup>p = 0.05)

**Table 5** Tests of equality of group means. Primary Education

Variables	Wilks' Lambda	F	Df1	Df2	Sig.	
Setting	.901	10.897	1	99	.001 <sup>a</sup>	
Socioeconomic status	.999	.141	1	99	.709	
Verbal intelligence	.961	4.069	1	99	.046 <sup>a</sup>	
Motivation	Desire to work and self-esteem	.990	.980	1	99	.325
	Anxiety	.996	.411	1	99	.523
	Lack of interest	.978	2.254	1	99	.136
	Self-demand	.995	.457	1	99	.501
Extramural exposition	.988	1.237	1	99	.269	
Language test	Grammar	.977	2.308	1	99	.132
	Vocabulary	.949	5.279	1	99	.024 <sup>a</sup>
	Listening	.998	.246	1	99	.621
	Reading	.997	.299	1	99	.586
Total	.983	1.693	1	99	.196	

(<sup>a</sup>p = 0.05)

**Table 6** Classification results. Year 6 Primary Education

Group	Predicted group		Total
	Non-CLIL	CLIL	Non-CLIL
Non-CLIL	55	16	71
CLIL	9	21	30
Non-CLIL	77.5	22.5	100.0
CLIL	30.0	70.0	100.0

established that equal matrices of variances are rejected; therefore, the groups do not have the same variance matrix.

The discriminant analysis shows a canonical correlation of 0.507 with an eigenvalue of 0.346 and a statistical Wilks' lambda of 0.743, with 12 degrees and  $p$ -value = 0.006 ( $< 0.05$ ), which leads us to reject the null hypothesis of equality of means and indicates the existence of a discriminant function that separates CLIL and non-CLIL groups significantly and accounts for 25% of the variance observed in their scores.

The standardised coefficients indicate that vocabulary is the best discriminating variable ( $-0.928$ ), followed by the variable 'setting' (0.567).

The analyses also evaluate the accuracy of the classification. Table 6 shows that measures resulted in a fairly positive classification for students belonging to their corresponding groups. 75.2% of original grouped cases are correctly classified.

As regards Secondary Education, Table 7 shows the results of discriminant function analyses where Wilks' lambda reveals that the discriminant function is statistically significant to classify results obtained regarding the following six variables:

**Table 7** Tests of equality of group means. Secondary Education

Variables		Wilk's Lambda	F	Df1	Df2	Sig.
Setting		.997	.403	1	120	.527
Socioeconomic status		.983	2.055	1	120	.154
Verbal intelligence		.957	5.358	1	120	.022 <sup>a</sup>
Motivation	Desire to work and self-esteem	.978	2.674	1	120	.105
	Anxiety	.999	.075	1	120	.785
	Lack of interest	.957	5.350	1	120	.022 <sup>a</sup>
	Self-demand	.988	1.441	1	120	.232
Extramural exposition		.992	.994	1	120	.321
Language test	Grammar	.749	40.261	1	120	.000 <sup>a</sup>
	Vocabulary	.843	22.367	1	120	.000 <sup>a</sup>
	Listening	.881	16.276	1	120	.000 <sup>a</sup>
	Reading	.868	18.259	1	120	.000 <sup>a</sup>
Total		.746	40.923	1	120	.000 <sup>a</sup>

(<sup>a</sup> $p = 0.05$ )

**Table 8** Classification results. Year 4 Secondary Education

Group	Predicted group		Total
	Non-CLIL	CLIL	Non-CLIL
Non-CLIL	48	15	63
CLIL	18	41	59
Non-grouped cases	1	0	1
Non-CLIL	76.2	23.8	100.0
CLIL	30.5	69.5	100.0
Non-grouped cases	100.0	.0	100.0

verbal intelligence ( $F = 5.358$  and  $p$ -value = 0.022); lack of interest ( $F = 5.350$  and  $p$ -value = 0.022); grammar ( $F = 40.261$  and  $p$ -value = 0.000); vocabulary ( $F = 22.367$  and  $p$ -value = 0.000); listening ( $F = 16.276$  and  $p$ -value = 0.000); and reading ( $F = 18.259$  and  $p$ -value = 0.000). Results from Box's test of Equality of Covariance Matrices ( $F = 0.987$ ,  $p$ -value = 0.510) indicate that the null hypothesis is accepted; therefore, equal matrices of variances are assumed. There is a canonical correlation of 0.521 with an eigenvalue of 0.373 and a statistical Wilks' lambda (= 0.728) with 12 degrees of freedom of Chi-square and a critical significance level of 0.000 ( $< 0.05$ ). This allows us to reject the null hypothesis and accept that at least one discriminant axis is significant. In this case, the standardised coefficients of the canonical discriminant functions indicate that the variable that has the greatest influence on the calculation of the function is grammar (0.787).

Finally, in the same way as with Primary Education, Table 8 shows a positive classification, as 73% of the original grouped cases are correctly classified.

## 5 Discussion

Positive results for Primary Education in favour of CLIL learners are in line with those presented in Navés and Victori (2010), where there are statistically significant differences between both groups on all the tests performed. In Secondary Education, our results coincide with Ruiz de Zarobe (2017), in which positive findings report greater lexical and syntactic complexity in CLIL learners. However, those results only coincide partially with Pérez-Vidal and Roquet (2015), in which CLIL learners outperform non-CLIL learners but without showing statically significant differences. With respect to the differences found between Primary and Secondary Education, our results are consistent with Agustín-Llach and Canga (2014), Garau et al. (2015) and Navés and Victori (2010). Their longitudinal studies show that, over time, CLIL learners improve their competence, giving rise to progressively larger differences between both groups. Although our study is not longitudinal, the fact that the effect size increases from Primary to Secondary Education can be related to the results presented in these longitudinal studies.



With respect to the differences found regarding the type of exercise included in the tests, in Primary Education our results are consistent with Jiménez-Catalán and Ruiz de Zarobe (2009), where CLIL learners also perform better than non-CLIL learners. However, in their study, there are statistically significant differences on all the tests performed, whereas in our study there are two exercises for which differences between both groups are not statistically significant. This may be due to the fact that these are exercises that do not require the use of lexico-grammatical elements within a meaningful context. The main implication of these results is that the type of instruction plays a role in the results. It is necessary to bear in mind that Jiménez-Catalán and Ruiz de Zarobe's study only focuses on receptive vocabulary and that their tests are different from ours, which may also account for the discrepancies in the results. There are also parallelisms between our study and Moreno's one (2009), in which there are statistically significant differences between CLIL and non-CLIL groups which apply not only to vocabulary size but also to vocabulary depth, pointing to the type of instruction as responsible for qualitative (and not only quantitative, as previous studies contend) differences in lexical competence. Also, our results may be related to the ones provided in Agustín-Llach (2009), where CLIL learners transfer from their L1 less frequently than their non-CLIL counterparts, showing a lower number of lexical errors for all the categories analysed. Agustín-Llach explains these differences alluding to the role of the target language and the way learners perceive it. Thus, the instructional programme is responsible for the differences insofar as it makes learners perceive the target language as a means of communication (CLIL instruction) or as merely a school subject (non-CLIL instruction): for CLIL students, 'the text becomes an exercise of communication rather than a language task' (124). In this sense, Ojeda (2009) also suggests the importance of the instructional programme in relation to how it makes students perceive their learning. Finally, our results confirm those in Admiraal, Westhoff and de Bot (2006), where the type of test administered is presented as accountable for CLIL's negative results: we agree with Admiraal, Westhoff and de Bot that the exercises used in tests should be adapted to both the instructional approach and the activities that are used in the classroom. Regarding Secondary Education, this study coincides with Ruiz de Zarobe (2008), where CLIL learners outperform non-CLIL learners in all scales measured. However, our results regarding vocabulary are not as satisfactory, due to the fact that differences between both groups are not statistically significant in any of the exercises. Again, regarding vocabulary, these results contradict those found in San Isidro (2010), where CLIL students show more strategies to convey meaning, and are more consistent with Ruiz de Zarobe (2008), where vocabulary is the only category in which Secondary students do not improve over time, and the only category for which differences between groups are not statistically significant. On the other hand, regarding grammar, our results do not coincide with Breidbach and Viebrock (2012), in which grammatical differences are less pronounced. Both groups of CLIL learners show a better grammatical competence with significant differences in both the global findings.

## 6 Conclusions

This study offers new empirically grounded insights into the current state of CLIL implementation and the effects of CLIL on students' language attainment. As regards RQ1, the results obtained complement previous research by offering CLIL outcomes regarding the impact of CLIL on the English grammar and vocabulary of 351 Primary and Secondary school students. In this respect, this study has shown that both at Primary and Secondary levels, there are statistically significant differences between CLIL and non-CLIL learners, in favour of the CLIL groups. Results on vocabulary and grammar show different effect sizes in Primary Education, being small for grammatical competence and medium for vocabulary. These differences in effect size increase in Secondary Education, which is indicative of students' improvement over time.

The results reported in this chapter also indicate that the difference between both groups of informants lies not only in language proficiency as reported by better overall results regarding lexical and grammatical competence, but also in the type of instruction as indicated by the comparison of results obtained in each exercise of the tests. Thus, this chapter has drawn attention to the central importance of considering the type of test administered in connection with the type of instruction implemented.

As far as RQ2 is concerned, this study has also investigated second-order interactions of individual difference variables and linguistic and contextual variables. The discriminant analyses evince different discriminant functions depending on the educational level under analysis. Therefore, it seems that, in general, it is important to contextualise findings, since individual and contextual variables do not have the same influence in Primary and Secondary Education.

On the one hand, in Primary Education, setting, verbal intelligence and vocabulary are the variables that display the greatest significance in the test of equality of group means. Vocabulary is the variable that best explains the statistically significant differences found between the groups. On the other hand, in Secondary Education, results show that, as it happens at Primary level, verbal intelligence carries a significant weight in explaining the differences between the groups. However, there are other variables that display significance in the test of equality of group means: lack of interest, vocabulary, grammar, listening and reading. At Secondary level, grammar is the variable that best explains the differences between the groups.

Taken together, these results suggest that, over the rest of the variables considered in this study, vocabulary and grammar are the variables that have the greatest influence on the calculation of the discriminant function. One of the most significant findings of this study is the fact that it confirms the effectiveness of the CLIL approach as far as students' language outcomes are concerned, providing better results even for the development of vocabulary and grammar, in spite of the importance that is traditionally given to them in FL instructional programmes.

**Acknowledgements** This work was supported by the Spanish Ministry of Economy and Competitiveness [research grant number FFI2012-32221] and the Government of Andalusia [research grant number HUM 2348] (Project: *The Effects of Content and Language Integrated Learning in Monolingual Communities: A Large-Scale Evaluation*).

## References

- Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual Secondary Education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation*, 12(1), 75–93.
- Agustín-Llach, M. P. (2009). The role of Spanish L1 in the vocabulary use of CLIL and non-CLIL EFL learners. In Y. Ruiz de Zarobe & R. Jiménez-Catalán (Eds.), *Content and Language Integrated Learning: Evidence from research in Europe* (pp. 112–129). Bristol: Multilingual Matters.
- Agustín-Llach, M. P., & Canga, A. (2014). 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.
- Alejo, R., & Piquer, A. (2016). Urban vs. rural CLIL: An analysis of input-related variables, motivation and language attainment. *Language, Culture and Curriculum*, 29(3), 245–262.
- Arribas, M. (2016). Analysing a whole CLIL school: Students' attitudes, motivation, and receptive vocabulary outcomes. *Latin American Journal of Content and Language Integrated Learning*, 9(2), 267–292. <https://doi.org/10.5294/la-clil.2016.9.2.2>.
- Breidbach, S., & Viebrock, B. (2012). CLIL in Germany—Results from recent research in a contested field of education. *International CLIL Research Journal*, 1(4), 5–16.
- Cenoz, J., Genesee, F., & Gorter, D. (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics*, 35(3), 243–262.
- Coonan, C. (2007). Insider views of the CLIL class through teacher self-observation-introspection. *International Journal of Bilingual Education and Bilingualism*, 10(5), 625–646.
- Coyle, D. (2008). Content and Language Integrated Learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism*, 10(5), 543–562.
- Coyle, D., Hood, P., & Marsh, D. (2010). *Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (pp. 71–83). New York: Springer Science+Business Media LLC.
- Dafouz, E., & Híbler, A. (2013). 'Zip your lips' or 'keep quiet': Main teachers' and language assistants' classroom discourse in CLIL settings. *The Modern Language Journal*, 97(3), 655–669.
- Dalton-Puffer, C. (2007). *Discourse in Content and Language Integrated Learning (CLIL) classrooms*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Dalton-Puffer, C. (2009). Communicative competence and the CLIL lesson. In Y. Ruiz de Zarobe & R. Jiménez-Catalán (Eds.), *Content and Language Integrated Learning: Evidence from research in Europe* (pp. 197–214). Bristol: Multilingual Matters.
- Dalton-Puffer, C. (2011). Content and Language Integrated Learning: From practice to principles? *Annual Review of Applied Linguistics*, 31, 182–204. <https://doi.org/10.1017/S0267190511000092>.
- Dalton-Puffer, C. (2014). "You can stand under my umbrella": Immersion, CLIL and bilingual education. A response to Cenoz, Genesee, and Gorter. 2013. *Applied Linguistics*, 35(2), 213–218.
- Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014). CLIL and motivation: The effect of individual and contextual variables. *Language Learning Journal*, 42, 209–224.
- Dueñas, M. (2004). The whats, whys, hows and whos of content-based instruction in second/foreign language education. *International Journal of English Studies*, 4(1), 73–96.
- European Union. (2008). Council conclusions of 22 May 2008 on multilingualism. *EUR\_Lex Official Journal of the European Union*, 2008/C 140/10. <https://eur-lex.europa.eu/oj/direct-access.html>. Accessed 22 November 2020.
- Fernández, A., & Canga, A. (2014). A preliminary study on motivation and gender in CLIL and non-CLIL types of instruction. *International Journal of English Studies*, 14(1), 21–36.

- Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J. M. (2019). 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>.
- Fontecha, A. F. (2010). First steps of CLIL in Spanish monolingual community: The case of La Rioja. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds.), *Spain: Implementation, results and teacher training* (pp. 79–94). Newcastle: Cambridge Scholars.
- Gierlinger, E. (2013). *The languages for CLIL model*. <http://clilingmesoftly.wordpress.com/clil-teachers-tl-competence/>. Accessed 22 November 2020.
- Jiménez-Catalán, R., & Ruiz de Zarobe, Y. (2009). The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL instruction. In Y. Ruiz de Zarobe & R. Jiménez-Catalán (Eds.), *Content and Language Integrated Learning: Evidence from research in Europe* (pp. 81–92). Bristol: Multilingual Matters.
- Juan-Garau, M., Prieto-Arranz, J., & Salazar, J. (2015). Lexico-grammatical development in secondary education CLIL learners. In M. Juan-Garau & J. Salazar (Eds.), *Content-based language learning in multilingual educational environments* (pp. 179–195). Educational Linguistics. Springer International Publishing Switzerland. [https://doi.org/10.1007/978-3-319-11496-5\\_11](https://doi.org/10.1007/978-3-319-11496-5_11).
- Lancaster, N. (2018). Extramural exposure and language attainment: The examination of input-related variables in CLIL programmes. *Porta Linguarum*, 29, 91–114.
- Lasagabaster, D. (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in Language Learning and Teaching*, 5(1), 3–18.
- Lasagabaster, D. (2015). Different educational approaches to bi-or multilingualism and their effect on language attitudes. In M. Juan-Garau & J. Salazar (Eds.), *Content-based language learning in multilingual educational environments*. Educational Linguistics. Springer International Publishing Switzerland. [https://doi.org/10.1007/978-3-319-11496-5\\_2](https://doi.org/10.1007/978-3-319-11496-5_2).
- Lasagabaster, D., & Sierra, J. M. (2009). Language attitudes in CLIL and traditional EFL classes. *International CLIL Research Journal*, 1(2), 4–17.
- Lasagabaster, D., & Ruiz de Zarobe, Y. (Eds.). (2010). *CLIL in Spain: Implementation, results and teacher training*. Newcastle: Cambridge Scholars.
- Llinares, A., & Dafouz, E. (2010). Content and language integrated programmes in the Madrid region: Overview and research findings. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds.), *Spain: Implementation, results and teacher training* (pp. 95–114). Newcastle: Cambridge Scholars.
- Llinares, A., Morton, T., & Whittaker, R. (2012). *The roles of language in CLIL*. Cambridge University Press.
- Lorenzo, F. (2017). Historical literacy in bilingual settings: Cognitive academic language in CLIL history narratives. *Linguistic and Education*, 37, 32–41.
- Lorenzo, F., Granados, A., & Ávila, I. (2019). The development of cognitive academic language proficiency in multilingual education: Evidence of a longitudinal study on the language of history. *Journal of English for Academic Purposes*, 41. <https://doi.org/10.1016/j.jeap.2019.06.010>.
- Madrid, D., & Barrios, E. (2018). A comparison of students' educational achievement across programmes and school types with and without CLIL provision. *Porta Linguarum*, 29, 29–50.
- Marsh, D. (Ed.). (2002). *CLIL/EMILE. The European dimension: Actions, trends, and foresight potential*. Jyväskylä: University of Jyväskylä.
- Martín del Pozo, M. A. (2013). Formación lingüística del profesorado universitario para la docencia en inglés. *Revista de Docencia Universitaria. REDU*, 11(3), 197–218.
- Martín del Pozo, M. A. (2015). Signposts for comprehensive knowledge of CLIL contexts. *Procedia—Social and Behavioral Sciences*, 212, 99–105.
- Martínez, M., & Juncal Gutiérrez, M. (2009). The acquisition of English syntax by CLIL learners in the Basque Country. In Y. Ruiz de Zarobe & R. Jiménez-Catalán (Eds.), *Content and Language Integrated Learning: Evidence from research in Europe* (pp. 176–196). Bristol: Multilingual Matters.

- Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL: Content and Language Integrated Learning in bilingual and multilingual education*. Oxford: Macmillan Education.
- Moreno, S. (2009). Young learners' L2 word association responses in two different learning contexts. In Y. Ruiz de Zarobe & R. Jiménez-Catalán (Eds.), *Content and Language Integrated Learning: Evidence from research in Europe* (pp. 93–111). Bristol: Multilingual Matters.
- Navarro-Pablo, M., & García-Jiménez, E. (2018). Are CLIL students more motivated? An analysis of affective factor and their relation to language attainment. *Porta Linguarum*, 29, 71–90.
- Navés, T., & Victori, M. (2010). CLIL in Catalonia: An overview of research studies. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds.), *Spain: Implementation, results and teacher training* (pp. 31–54). Newcastle: Cambridge Scholars.
- Nikula, T. (2007). Speaking English in Finnish content-based classroom. *World Englishes*, 26(2), 206–223.
- Nikula, T., Daufouz, E., Moore, P., & Smit, U. (2016). *Conceptualising integration in CLIL and Multilingual Education*. Bristol: Multilingual Matters.
- Ojeda, A. (2009). Themes and vocabulary in CLIL and non-CLIL instruction. In Y. Ruiz de Zarobe & R. Jiménez-Catalán (Eds.), *Content and Language Integrated Learning: Evidence from research in Europe* (pp. 130–156). Bristol: Multilingual Matters.
- Pavón, V. (2018). Learning outcomes in CLIL programmes: A comparison of results between urban and rural environments. *Porta Linguarum*, 29, 9–28.
- Pelechano, V. (1994). Prueba MA. *Análisis y Modificación de la Conducta*, 20, 71–72.
- Pérez-Cañado, M. (2012). CLIL research in Europe: Past, present and future. *International Journal of Bilingual Education and Bilingualism*, 15(3), 315–341. <https://doi.org/10.1080/13670050.2011.630064>.
- Pérez-Cañado, M. (2014). Teacher training needs for bilingual education: In-service teacher perceptions. *International Journal of Bilingual Education and Bilingualism*, 19(3), 266–295. <https://doi.org/10.1080/13670050.2014.980778>.
- Pérez-Vidal, C., & Roquet, H. (2015). The linguistic impact of a CLIL Science programme: An analysis measuring relative gains. *System*, 54, 80–90.
- Pfenninger, S. (2016). All good things come in threes: Early English learning, CLIL and motivation in Switzerland. *Cahiers de I/LSL*, 48, 119–147.
- Rascón, D., & Bretones, C. M. (2018). Socioeconomic status and its impact on language and content attainment in CLIL contexts. *Porta Linguarum*, 29, 115–135.
- Roquet, H., & Pérez-Vidal, C. (2015). Do productive skills improve in Content and Language Integrated Learning contexts? The case of writing. *Applied Linguistics*, 38(4), 489–511. <https://doi.org/10.1093/applin/amv050>.
- Ruiz de Zarobe, Y. (2008). CLIL and foreign language learning: A longitudinal study in the Basque Country. *International CLIL Research Journal*, 1(1), 61–72.
- Ruiz de Zarobe, Y. (2017). Language awareness and CLIL. In J. Cenoz, D. Gorter, & S. May (Eds.), *Language awareness and multilingualism* (pp. 1–11). Springer, Cham. [https://doi.org/10.1007/978-3-319-02325-0\\_13-2](https://doi.org/10.1007/978-3-319-02325-0_13-2).
- Ruiz de Zarobe, Y., & Jiménez-Catalán, R. (2009). *Content and Language Integrated Learning: Evidence from research in Europe*. Bristol: Multilingual Matters.
- Rumlich, D. (2016). *Evaluating Bilingual Education in Germany: CLIL students' general proficiency, EFL self-concept and interest*. Berlin: Peter Lang.
- San Isidro, X. (2010). An insight into Galician CLIL: Provision and results. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds.), *Spain: Implementation, results and teacher training* (pp. 55–78). Newcastle: Cambridge Scholars.
- Santamaría, P. (2018). *EFAI: Evaluación Factorial de las Aptitudes Intelectuales*. Tea.
- Turner, M. (2013). CLIL in Australia: The importance of context. *International Journal of Bilingual Education and Bilingualism*, 16(4), 395–410. <https://doi.org/10.1080/13670050.2012.691086>.

Villarreal, I., & García, M. del P. (2009). Tense and agreement morphology in the interlanguage of Basque/Spanish bilinguals: CLIL versus non-CLIL. In Y. Ruiz de Zarobe & R. Jiménez-Catalán (Eds.), *Content and Language Integrated Learning: Evidence from research in Europe* (pp. 157–175). Bristol: Multilingual Matters.