

# The Effects of CLIL on FL Learning: A Longitudinal Study



María del Mar Gálvez Gómez

**Abstract** This article reports on a longitudinal study conducted with Content and Language Integrated Learning (CLIL) and English as a Foreign Language (EFL) students in seven public and private schools situated in the Spanish monolingual community of Andalusia, more concretely in the province of Jaén. This research aims at evaluating the impact of CLIL on the development of 223 students' English linguistic competence by assessing their attainment in grammar, vocabulary, reading, listening and speaking. To this end, the bilingual and the non-bilingual strands were matched on a pre-test in terms of English level, verbal intelligence and motivation and they were then administered post- and delayed post-tests to trace their progress from Primary Education to Compulsory Secondary Education to Baccalaureate. Within-cohort and across-cohort comparisons are carried out in order to determine the evolution of both CLIL and non-CLIL students in the different educational levels in terms of the linguistic components and skills. Discriminant analyses are finally carried out taking into account all the intervening variables considered in the study, with the objective of determining if CLIL is actually responsible for the differences detected or if, on the contrary, other variables account for a greater proportion of the variance.

## 1 Introduction

After almost two decades since the appearance of Content and Language Integrated Learning in Europe, this 'major trend' (Fernández Sanjurjo et al., 2019, p. 661) has been implemented in all educational settings throughout the continent (Hüttner & Smit, 2014, p. 160), with the objective of pushing plurilingualism forward and meeting the so-called 'mother tongue + 2 objective' (Pérez Cañado, 2018, p. 52), by means of which all citizens should become proficient in their mother tongue and in two foreign languages at least (European Commission, 1995). Hence, with the support of many European Union Institutions, CLIL has been embraced enthusiastically by

---

M. d. M. Gálvez Gómez (✉)  
University of Jaén, Jaén, Spain  
e-mail: [mmgg0007@red.ujaen.es](mailto:mmgg0007@red.ujaen.es)

© Springer Nature Switzerland AG 2021  
M. L. Pérez Cañado (eds.), *Content and Language Integrated Learning  
in Monolingual Settings*, Multilingual Education 38,  
[https://doi.org/10.1007/978-3-030-68329-0\\_8](https://doi.org/10.1007/978-3-030-68329-0_8)

all the stakeholders (Lasagabaster & Doiz, 2016, p. 1) ‘as a lever for change and success in language learning’ (Pérez Cañado & Ráez Padilla, 2015, p. 1), becoming ‘a well-established part of education systems across Europe’ (Surmont et al., 2016, p. 320).

As a consequence of this widespread implementation of CLIL, the research into the effects of this methodology on students and their skills has also increased. As Jäppinen (2005, p. 149) underscores, CLIL has become ‘an extremely prolific phenomenon’, making foreign language learning more naturalistic (Nieto Moreno de Diezmas, 2016, p. 81) and demonstrating that nowadays ‘multilingualism is the norm while monolingualism is the exception’ (Ouazizi, 2016, p. 113). However, despite the great amount of publications on the effects of CLIL, some of the already existing studies present a series of methodological shortcomings which might endanger the validity of the results (Bruton, 2011, 2013, 2015; Paran, 2013; Pérez Cañado, 2011, 2012). These lacunae, classified by Pérez Cañado and Ráez Padilla (2015) in terms of *variables, research design and statistical methodology*, make us recognise that ‘we simply do not have enough evidence’ (Paran, 2013, p. 331) and that there is still ‘a need for unbiased, unskewed and methodologically sound research to continue driving the CLIL agenda forward’ (Pérez Cañado & Lancaster, 2017, p. 2). In this respect, according to many scholars (Bruton, 2011; Lasagabaster & Ruiz de Zarobe, 2010; Pérez Cañado, 2017a, 2018; Pérez Cañado & Ráez Padilla, 2015; Ruiz de Zarobe, 2011), more importance should be given to longitudinal investigations which can examine the effects of CLIL across the different educational levels.

This is exactly the starting point of the present investigation, which will report on the results of a longitudinal study on the effects of CLIL on foreign language (FL) outcomes across educational levels (Primary, Compulsory Secondary and non-compulsory Secondary Education). As Vez (2009, p. 18) claims, ‘There is not yet empirical evidence from EU countries on which to base definitive claims about the educational (or other) advantages of multilingual education’. Moreover, ‘longitudinal studies with pre-, post-, and follow-up assessments are still rare’ (Piesche et al., 2016, p. 109). Therefore, this investigation seeks to offer a rigorous monitoring of CLIL implementation, which is ‘key for a better understanding of the processes and outcomes of these courses’ (Coyle et al., 2010 cited in Pascual Bajo, 2018, p. 222).

After framing the topic against the backdrop of prior investigations, the research design of the study will be described, reporting on the results obtained within and across cohorts in terms of English as a foreign language competence. The evolution of the bilingual and non-bilingual groups, which were previously matched on a pre-test phase in terms of English level, verbal intelligence and motivation, from Primary Education to Compulsory Secondary Education (CSE) to Bacalaureate, is depicted through the administration of post- and delayed post-tests. Within-cohort comparisons in relation to the intervening variables considered are also presented, together with the discriminant analyses carried out to find out if the independent variable (CLIL) is actually responsible for the differences detected.

## 2 A Critical Reading of Prior Research

CLIL practice has spread rapidly in the past ten years, currently spanning the continent from north to south, and from east to west (Pérez Cañado, 2012, p. 319). However, this interest in CLIL has not always been accompanied by the same amount of publications on the issue, especially when regarding research projects with a longitudinal focus. The existing longitudinal investigations tend to focus on the four fields around which Wolff (2005) considers CLIL investigations should be articulated: the effects of CLIL on FL, L1 and subject content competence and motivational aspects. Some of these investigations have focused on Primary or Secondary Education and those evaluating FL learning have considered both receptive and productive skills, although they generally have not done it simultaneously.

Among these longitudinal studies analysing FL competence, the investigation by Admiraal et al. (2006) in The Netherlands is worthy of mention. After measuring the vocabulary knowledge, the reading comprehension level and the oral proficiency of 1,305 Secondary students who had received four years of CLIL education through English in five Dutch schools, the results revealed higher scores for the oral and reading parts of the investigation, whereas no differences appeared when dealing with vocabulary. No negative effects were found either for subject matter learning or the L1. Nevertheless, the study lacks statistical analyses that confirm the outcomes can be attributed to CLIL (Admiraal et al., 2006, p. 91).

One year later, this time in Switzerland, Serra (2007) conducted a longitudinal study to evaluate German-speaking Primary Education students' FL competence, focusing on their oral and written comprehension in Italian or Romansch as a second language. Similar results were obtained for the experimental and the control groups, suggesting then that both cohorts performed equally well on these skills.

Turning now to the Spanish scenario, special attention should be paid to the investigation by Ruiz de Zarobe (2008), who compared the oral and written competence of 89 students in their 3rd and 4th year of CSE and at the post-compulsory level. The results were extremely positive since CLIL students obtained better outcomes than the non-CLIL counterparts. Further evidence on the differences between skills and abilities in favour of CLIL was provided, coinciding with previous research (Jiménez Catalán et al., 2006; Ruiz de Zarobe, 2007; Ruiz de Zarobe & Jiménez Catalán, 2009 cited in Lasagabaster & Ruiz de Zarobe, 2010, p. 17).

In 2015, Rallo Fabra and Jacob carried out another longitudinal research in Secondary Education to evaluate the effects of CLIL on pupils' oral skills at the onset of the programme and two years after its implementation. They worked with 43 students from state-run Secondary schools in the Balearic Islands, who were divided into the experimental CLIL group and the control non-CLIL cohort. Special attention was paid to students' fluency and the number of vowel errors in English. Outcomes revealed pupils' pronunciation of English vowels was unaffected by CLIL and there was no significant improvement over the two years considered. Moreover, CLIL students' pronunciation was not better than their EFL peers' and no significant differences were detected in fluency either (Rallo Fabra & Jacob, 2015).

Similarly, in the study by Pladevall-Ballester and Vallbona (2016), the superiority of CLIL is not proved. The project, which was developed over the course of two academic years, focused on analysing students' receptive skills in a minimal CLIL input context at Primary school level. 287 pupils from four different state-funded private schools participated in the investigation, whose results showed an important progress regarding the achievement and development of reading and listening skills in both CLIL and non-CLIL cohorts between the first and the last test. However, the EFL group outperformed the CLIL cohort in relation to listening skills. As for reading, no significant differences were detected.

More recently, Pascual Bajo (2018) developed an investigation aimed at evaluating CLIL from a qualitative and a quantitative point of view within the context of two educational institutions: a public school with CLIL and non-CLIL streams, and a semi-private school with no CLIL provision in the province of Valencia. 63 pupils constituted the sample of the longitudinal quantitative study. Outcomes confirmed the superior English language competence of CLIL pupils. The CLIL cohort outperformed the non-CLIL stream on all the skills and aspects considered, with particularly significant differences in the use of English, vocabulary and reading (Pascual Bajo, 2018, p. 382). Six months after the end of the CLIL programme, in the delayed post-test phase, the outcomes improved for all skills studied, except for that of writing.

To finish, moving on to the Andalusian community, where our study has also been carried out, two longitudinal investigations must be foregrounded: Pérez Cañado and Lancaster's (2017) and Pérez Cañado's (2018). The former is a longitudinal, quantitative, quasi-experimental case study on oral comprehension and production in Andalusia. Their research aimed at determining whether CLIL students acquired greater oral comprehension and production skills in comparison to non-CLIL pupils. Moreover, it tried to find out if the possible differential effects of CLIL continued after the CSE CLIL programme finished. 24 secondary students participated in the investigation and sat two FL competence tests to assess their oral comprehension and production skills. An initial pre-test was used to guarantee the homogeneity of the cohorts in terms of English language proficiency. Outcomes evinced CLIL students had a higher level in English oral comprehension and production when compared to those obtained by the EFL group. It was also found that the effects of CLIL pervaded six months after the programme was discontinued, but only in the case of oral production.

The latter (Pérez Cañado, 2018) focused on the effects of CLIL on L2 learning. This investigation is especially interesting in the context of our study because of its similarities with our study in terms of research questions, instruments and variables. A total of 1,033 students enrolled in a CLIL programme and 991 EFL pupils took part in the project. These learners, who were completing 6th grade of Primary Education or finishing the 4th grade of CSE, came from different public, semi-public and private schools situated in three Spanish monolingual communities: Andalusia, The Canary Islands and Extremadura. Participants were firstly matched in terms of FL proficiency, motivation and verbal intelligence to guarantee the homogeneity of the cohorts. The outcomes revealed the CLIL group at both Primary and Secondary levels had a higher linguistic competence (on grammar, vocabulary, listening, speaking and

reading). The linguistic competence differential was especially marked at Secondary Education, where the CLIL group clearly outperformed its non-CLIL counterpart on all aspects considered. As for the durability of the effects of CLIL, it was shown that outcomes pervaded six months after the programmes were discontinued. However, no statistically significant differences were detected between the EFL semi-private stream and the CLIL groups (public and private) in 1st year of non-compulsory Secondary Education. The delayed post-test phase results coincide, to some extent, with Pladevall-Ballester and Vallbona's findings (2016), according to which CLIL had more positive effects on the productive skills than on the receptive skills. To finish, the discriminant analysis carried out to study the competence differential between the treatment and comparison groups allowed Pérez Cañado (2018) to confirm CLIL is truly responsible for the differences found.

After this brief summary of the most important longitudinal studies developed hitherto in Europe and more concretely in Spain (cf. also Chapter 7 in this volume), let us now turn to our investigation, which strives to provide updated empirical evidence on CLIL practice by overcoming the main shortcomings presented by prior investigations.

### 3 The Study

The present investigation is framed within a broader research project which has developed a large-scale evaluation of CLIL programmes in one of the Spanish monolingual communities with the least tradition in bilingual education: Andalusia. With a mixed-research design, the study examines the effects of CLIL from quantitative and qualitative perspectives. The impact of CLIL on FL learning, content learning (Natural Sciences) and L1 learning of Primary (6th grade) and Secondary (4th grade) Education Students is analysed in the quantitative part of the project, in which participants in the experimental CLIL group are assessed in comparison to a control non-CLIL cohort, in order to find out whether they develop superior language and content skills to those promoted by a traditional EFL programme. Moreover, the study aims to determine if the possible effects exerted by CLIL pervade six months after the programme is discontinued, when the same CSE students are in the first grade of Baccalaureate, or if they gradually disappear. This quantitative part is then completed from a qualitative point of view by means of a SWOT analysis on the satisfaction generated by CLIL programmes among all the agents involved. Stakeholders' perspectives (teachers, students and parents) are collected via questionnaires, and personal and focus-group interviews are carried out with teachers and students. The present study is inserted within the quantitative side of the project and focuses specifically on the effects of CLIL on English as an FL learning through the following research questions.

### ***3.1 Research Questions***

**RQ1.** Do CLIL programmes implemented with Primary and Secondary school students (experimental group) develop superior linguistic competence (in grammar, vocabulary, receptive skills and oral production) to that promoted by EFL programmes with students from the same level (control group)? More simply, is there a linguistic competence differential between CLIL and EFL groups at Primary and Secondary school level in the province of Jaén?

**RQ2.** What is the modulating (differential) effect exerted on the Primary and Secondary students' English language competence by the following intervening variables: type of school (public and private), setting (rural–urban), gender, socioeconomic status (SES), motivation, verbal intelligence and English level?

**RQ3.** Do the possible differential effects exerted by CLIL programmes on English language competence pervade six months after the CLIL programme is discontinued or do they gradually disappear?

**RQ4.** If there is a competence differential between the treatment and comparison groups, is it truly ascribable to language learning based on academic content processing?

### ***3.2 Research Design***

The quantitative side of the broader study is an example of applied, primary, quasi-experimental research, with a pre-test/post-test control group design, to which a delayed post-test has also been added. Thus, as Rossell and Baker (1996), together with Cummins (1979 cited in Lancaster, 2015, p. 137) specified, four benchmarks are necessary for studies to be methodologically sound, and this study meets them:

1. Studies must compare students in a bilingual programme to a control group of similar students.
2. The design must ensure that initial differences between treatment and control groups are controlled statistically.
3. Results must be based on standardised test scores.
4. Differences between the scores of treatment and control groups must be determined by means of appropriate statistical tests.

### ***3.3 Sample***

The final sample that took part in the quantitative part of the investigation was made up of 223 students from public and private centres in the Andalusian province of Jaén, situated in the north-eastern part of the autonomous community, in the south of Spain. Most of the participants are studying the fourth grade of Compulsory Secondary

Education (60.1%), being the rest Primary Education students (39.9%). In the same vein, the majority of the cohort belong to a public school (76.2%) where CLIL and EFL branches co-exist, while 23.8% are enrolled in a bilingual private school (23.8%). 76.2% of these schools are located in urban settings, while the remaining 23.8% are placed in rural areas. Regarding gender, practically equal percentages are found, with a slightly higher number of female participants (50.7%). Finally, a significantly higher percentage of CLIL pupils have participated in the study (81.3%), with the rest of pupils being enrolled in the traditional EFL programme (18.7%).

The homogeneity of the sample, both in the experimental and the control group, has been guaranteed from the beginning of the research. In fact, the first year was entirely devoted to matching students within schools in terms of verbal intelligence, motivation and English level. Pupils were administered initial motivation and verbal intelligence tests that would allow us to select the really comparable groups. Moreover, the English grades of these students were collected to compare the results obtained by CLIL and non-CLIL pupils. Those schools whose outcomes evinced the greatest homogeneity comprise the final sample of the investigation.

### 3.4 Variables

Three different types of variables have been incorporated in the study: dependent, independent and moderating.

- The *dependent variable* is the students' English language (FL) competence (grammar, vocabulary, receptive skills and oral productive skill).
- The *independent variable* corresponds to the CLIL programme implemented in the different types of schools (public and private).
- Finally, the *moderating variables* are the following:
  - Verbal intelligence
  - Motivation
  - Socioeconomic status (SES)
  - Gender
  - Type of School
  - Setting

### 3.5 Instruments

For the collection of data, four different instruments were used, depending on the stage of the investigation. An initial questionnaire was firstly administered to students. It comprised personal information and data on their parents' age and educational level, which was taken as a proxy for SES. Moreover, verbal intelligence and motivation tests were employed in this initial phase, together with the English

language tests. All of these were already existing and validated instruments which belong to language teaching and psychology research areas.

The verbal intelligence test, which was part of the Evaluación Factorial de las Aptitudes Intelectuales (EFAI), designed by Santamaría et al. (2016), and the motivation test (Pelechano Barberá, 1994) were applied in each of the schools over the course of an hour almost at the end of the academic year 2014–2015, after exactly ten years of CLIL implementation in the community. Two different versions of the verbal intelligence tests were applied, adapted to the sixth grade of Primary Education and fourth grade of Compulsory Secondary Education. The former version comprises 26 items, while the latter is reduced to 23. In both cases, pupils had to choose from four multiple-choice options involving analogies, antonyms and odd-one-out and they had five minutes to complete as many items as they could. In turn, Pelechano's MA test (1994) comprises a total of 35 items aimed at measuring students' motivation, and it isolates four motivational factors of achievement and anxiety: (i) vain desire to work and self-esteem (10 items); (ii) anxiety when facing exams (9 items); (iii) lack of interest in studying (9 items); and (iv) realistic personal self-demand (7 items).

Finally, the English competence tests applied were originally devised for the project (Madrid et al., 2018) and incorporated three different batteries of six tests each (grammar, vocabulary, reading, writing,<sup>1</sup> listening and speaking) which corresponded to the levels at which our study has been developed, namely, 6th grade of PE, 4th grade of CSE and 1st grade of Baccalaureate. A rubric was also designed and validated for the assessment of oral production, following five main criteria: grammatical accuracy, lexical range, fluency and interaction, pronunciation and task fulfilment (Pérez Cañado & Lancaster, 2017).

### 3.6 Data Analysis

The data obtained from all the tests has been analysed statistically with the aid of the SPSS program, in its 23.0 version. To guarantee the homogeneity and comparability of the sample, participants have been matched for verbal intelligence, motivation and English level through the ANOVA and the T-test. Moreover, in order to determine the existence of any statistically significant differences within and across groups in terms of the different identification variables considered, the ANOVA, the T-test, the Mann–Whitney U test and Tukey's HSD test have been employed. To calculate effect sizes, Cohen's *d* and eta squared have been used. Lastly, to address RQ4, successive discriminant analyses have been carried out to establish which variable(s) are truly responsible for the differences detected.

---

<sup>1</sup>The results corresponding to the writing skill are not included in this article since they are still under analysis.



## 4 Results and Discussion

Taking into consideration the research questions set out at the beginning of the investigation, a detailed examination of the FL level attained by both CLIL and non-CLIL students will be offered in this section, with a special focus on productive and receptive skills, the effects of the different intervening variables considered on the students' FL proficiency, and lastly, the durability or medium-term effects of the CLIL and EFL programmes on FL competence.

### 4.1 Across-Cohort Comparison

After the initial overall comparison was carried out, statistically significant differences were detected in favour of the experimental group ( $p < 0.001$ ). High confidence levels were then found on most of the aspects analysed, with listening being the only skill for which no differences between the CLIL and the non-CLIL group were observed, coinciding with Serra (2007). However, CLIL students clearly outperformed their non-CLIL counterparts since the means obtained were 7.99 and 6.56, respectively (cf. Table 1).

Focusing on speaking, statistically significant differences in favour of the CLIL group were detected ( $p = .0240$ ). Moreover, more differences emerged when attending to the five subspects mentioned above. Thus, CLIL students outperformed their non-CLIL counterparts in their knowledge and use of grammar ( $p = .0330$ ) and in their pronunciation in the FL ( $p = .0050$ ). Nevertheless, no statistically significant differences appeared when analysing students' lexical range, their fluency in English and their adaptability to the task provided ( $p = .0840$ ,  $.0580$  and  $.0590$ , respectively). However, the means obtained by the CLIL group in each one of these subspects were

**Table 1** Foreign language competence: Across-cohort comparison

Educational Level	Skills	Group	Mean	Standard Deviation	Cohen's d	p-value
General	Use of English	Non-CLIL	13.34	8.323	-0.995	<0.001
		CLIL	23.23	10.263		
	Vocabulary	Non-CLIL	7.98	4.150	-0.807	<0.001
		CLIL	10.75	3.250		
	Listening	Non-CLIL	6.56	4.707	-0.301	0.084
		CLIL	7.99	4.761		
	Reading	Non-CLIL	3.15	2.963	-0.730	<0.001
		CLIL	5.83	3.810		
Total	Non-CLIL	31.02	15.244	-1.231	<0.001	
	CLIL	47.79	13.223			

**Table 2** Foreign language competence. General speaking analysis

	Group	Mean	Standard Deviation	Rosenthal's R	p-value
Speaking_Total	Non-CLIL	5.806	2.3772	.7577	.0240
	CLIL	7.296	1.7279		
Grammatical	Non-CLIL	1.028	.5278	.6799	.0330
	CLIL	1.370	.4722		
Lexical_Range	Non-CLIL	1.194	.5461	.4465	.0840
	CLIL	1.426	.3008		
Fluency_Interaction	Non-CLIL	1.111	.6077	.5377	.0580
	CLIL	1.444	.4237		
Pronunciation	Non-CLIL	1.222	.3919	1.1575	.0050
	CLIL	1.556	.3203		
Task_Fulfilment	Non-CLIL	1.250	.4926	.5295	.0590
	CLIL	1.500	.3669		

always higher than the non-CLIL group. These results corroborate Pérez Cañado and Lancaster's outcomes (2017), according to which CLIL students outperformed their EFL counterparts in oral production (cf. Table 2).

When analysing the data from Primary and Secondary students separately, the superior linguistic competence of CLIL pupils was again confirmed. When comparing 4th year CSE EFL students with CLIL learners at the same educational level, it is proved that the latter outperformed the former on all the skills and aspects sampled. On the contrary, when dealing with 6th year Primary Education students, no differences appeared in terms of speaking, in contrast to the tendency found in previous research at Primary Education level (Madrid & Barrios, 2018; Nieto Moreno de Diezmas, 2016; Ruiz de Zarobe, 2008), according to which the CLIL group was significantly superior in terms of oral production. However, Primary school students belonging to CLIL are found to be superior on all skills analysed ( $p = 0.931$ ) (cf. Table 3).

## 4.2 Differential Effect of Intervening Variables on FL Competence

To fully understand the effects of CLIL and EFL programmes on English proficiency, an analysis of the data in terms of the different intervening variables considered will now be included, addressing our RQ2.

Taking into account the gender of the participants, in the initial overall analysis, no statistically significant differences were detected between female and male participants in any of the skills evaluated, coinciding with the results obtained by Heras and Lasagabaster (2015) and Pascual Bajo (2018). However, in a deeper analysis carried

**Table 3** FL competence comparison per group (T-test)

Educational level	Skill	Group	Mean	Standard Deviation	Cohen's d	p-value
Primary Education	Use of English	Non-CLIL	10.94	7.197	-0.777	0.006
		CLIL	15.99	6.339		
	Vocabulary	Non-CLIL	11.13	2.964	-0.575	0.040
		CLIL	12.75	2.803		
	Listening	Non-CLIL	11.88	2.247	-0.665	0.018
		CLIL	13.29	2.098		
	Reading	Non-CLIL	5.25	3.357	-1.072	<0.001
		CLIL	9.14	3.679		
	Speaking	Non-CLIL	6.083	2.354	-0.077*	0.931
		CLIL	5.800	2.049		
Total	Non-CLIL	39.19	14.520	-0.901	0.002	
	CLIL	51.16	13.015			
Compulsory Secondary Education	Use of English	Non-CLIL	14.88	8.76	-1.438	<0.001
		CLIL	28.27	9.43		
	Vocabulary	Non-CLIL	5.96	3.52	-1.156	<0.001
		CLIL	9.35	2.78		
	Listening	Non-CLIL	3.16	1.70	-0.761	0.004
		CLIL	4.30	1.46		
	Speaking	CLIL	5.667	2.480	0.434*	0.011
		Non-CLIL	7.636	1.497		
	Reading	Non-CLIL	1.80	1.66	-1.118	<0.001
		CLIL	3.52	1.51		
	Total	Non-CLIL	25.80	13.52	-1.508	<0.001
		CLIL	45.45	12.91		

\*The analyses for speaking are done by means of Rosenthal's R

out within each separate cohort, the T-test evinced women in the CLIL group have a higher level of vocabulary in the FL and generally outperform their male peers. Regarding the non-CLIL group, no statistical confirmation of clear differences could be reported between girls and boys (cf. Table 4).

Following with the variable of setting, our first evaluation detected statistically significant differences in favour of those students belonging to an urban centre, both generally and in their level of use of English, although pupils in the rural schools outperformed their counterparts on the oral receptive skill. Similar results appeared when analysing each cohort separately, as differences were also found in favour of the urban school pupils within the CLIL group. However, the rural school students outperformed their counterparts on all skills tested except for use of English within the non-CLIL group (cf. Table 5).

**Table 4** Comparison per gender (T-test)

Group	Skills	Gender	Mean	Standard Deviation	Cohen's d	p-value	
General	Use of English	Male	20.68	10.75	-0.111	0.408	
		Female	21.86	10.45			
	Vocabulary	Male	9.68	3.89	-0.257	0.056	
		Female	10.61	3.32			
	Listening	Male	7.63	4.70	-0.002	0.988	
		Female	7.64	4.87			
	Reading	Male	4.88	3.68	-0.197	0.144	
		Female	5.63	3.90			
	Total	Male	42.87	15.26	-0.189	0.160	
		Female	45.73	15.02			
Non-CLIL	Use of English	Male	15.25	9.94	0.454	0.161	
		Female	11.52	6.13			
	Vocabulary	Male	8.10	4.71	0.058	0.854	
		Female	7.86	3.65			
	Listening	Male	7.10	4.54	0.222	0.481	
		Female	6.05	4.91			
	Reading	Male	3.40	3.45	0.166	0.599	
		Female	2.90	2.47			
	Speaking	Male	6.250	2.699	.1683*	0.515	
		Female	5.450	2.166			
	Total	Male	33.85	17.59	0.363	0.252	
		Female	28.33	12.46			
	CLIL	Use of English	Male	22.15	10.64	-0.207	0.170
			Female	24.26	9.83		
Vocabulary		Male	10.21	3.49	-0.329	0.030	
		Female	11.26	2.92			
Listening		Male	7.92	4.74	-0.028	0.850	
		Female	8.05	4.81			
Reading		Male	5.34	3.65	-0.248	0.100	
		Female	6.29	3.92			
Speaking		Male	7.036	1.759	.1742*	0.375	
		Female	7.577	1.718			
Total		Male	45.62	13.65	-0.325	0.032	
		Female	49.87	12.53			

\*The analyses for speaking are done by means of Rosenthal's R

**Table 5** Cohort comparison regarding setting

Group	Skills	Setting	Mean	Standard Deviation	Cohen's d	p-value
General	Use of English	Rural	15.32	8.57	-0.776	<0.001
		Urban	23.14	10.49		
	Vocabulary	Rural	10.13	3.73	-0.007	0.963
		Urban	10.16	3.62		
	Listening	Rural	9.23	4.06	0.444	0.002
		Urban	7.14	4.89		
	Reading	Rural	5.26	3.61	0.001	0.99
		Urban	5.26	3.88		
Total	Rural	39.94	14.18	-0.383	0.016	
	Urban	45.69	15.25			
Non-CLIL	Use of English	Rural	12.50	8.71	-0.179	0.573
		Urban	14.00	8.14		
	Vocabulary	Rural	10.67	3.16	1.402	<0.001
		Urban	5.87	3.61		
	Listening	Rural	11.11	3.08	3.401	<0.001
		Urban	3.00	1.65		
	Reading	Rural	5.17	3.17	1.516	<0.001
		Urban	1.57	1.50		
Total	Rural	39.44	14.32	1.119	0.001	
	Urban	24.43	12.68			
CLIL	Use of English	Rural	16.77	8.25	-0.822	<0.001
		Urban	24.81	10.11		
	Vocabulary	Rural	9.86	4.01	-0.343	0.133
		Urban	10.97	3.01		
	Listening	Rural	8.26	4.19	0.070	0.711
		Urban	7.92	4.90		
	Reading	Rural	5.31	3.86	-0.167	0.377
		Urban	5.95	3.80		
Total	Rural	40.20	14.32	-0.743	<0.001	
	Urban	49.65	12.30			

Socioeconomic status (SES) was also incorporated in our investigation taking into account the educational level of parents. Hence, three different groups were established according to their educational attainment: high (Tertiary Education), medium (vocational training or Secondary) and low (school qualifications or no studies). As a result, after applying the ANOVA test, statistically significant differences were detected on all skills analysed in favour of those students having a high socioeconomic status, except for use of English (cf. Table 6). These results coincide with

**Table 6** Comparison among cohorts regarding SES

Group	Skills	SES	Mean	Standard Deviation	Eta Squared	<i>p</i> -value
General	Use of English	Low	21.13	10.25	0.001	0.872
		Medium	21.81	11.13		
		High	21.00	10.18		
	Vocabulary	Low	8.28	3.26	0.150	0.000
		Medium	9.54	3.48		
		High	11.71	3.22		
	Listening	Low	5.19	3.18	0.236	0.000
		Medium	5.97	3.77		
		High	10.32	4.90		
	Reading	Low	3.23	1.97	0.197	0.000
		Medium	4.23	2.99		
		High	7.18	4.24		
	Total	Low	37.83	13.04	0.120	0.000
		Medium	41.56	14.51		
		High	50.20	14.19		

Anghel et al. (2016) and Fernández Sanjurjo et al. (2019) outcomes, according to which CLIL programmes tend to have a negative effect on students from a low socioeconomic background. However, recent investigations have proved that CLIL has cancelled out differences among social classes (Lorenzo, 2019; Pavón Vázquez, 2018; Pérez Cañado, 2017b; Rascón Moreno & Bretones Callejas, 2018). No statistically significant differences appeared among non-CLIL learners in any of the skills evaluated, except for their knowledge of vocabulary.

When analysing the cohorts separately, the same tendency was repeated among CLIL students, where those pupils with a high socioeconomic level are revealed to also have the highest level in all skills tested but use of English. In the case of non-CLIL students, statistically significant differences in favour of those having a high socioeconomic level exist only when dealing with their knowledge of vocabulary (cf. Table 7).

Valuable results were also obtained when taking into consideration type of school. In general terms, statistically significant differences were found for all skills considered in favour of those students belonging to a private bilingual school, with the exception of use of English, for which pupils from public bilingual schools were revealed to obtain better outcomes (cf. Table 8).

Different results are obtained within cohorts. In the case of Primary Education students, the same pattern is repeated, showing that private bilingual school students obtain higher results in vocabulary, use of English, listening and reading. However, when dealing with Secondary Compulsory Education pupils, results vary, as no private bilingual schools were considered in our study. Thus, statistically significant differences were found on all skills analysed in favour of public bilingual school

**Table 7** Comparison within cohorts in terms of SES (ANOVA)

Group	Skills	SES	Mean	Standard Deviation	Eta Squared	<i>p</i> -value
Non-CLIL	Use of English	Low	14.82	7.26	0.021	0.684
		Medium	12.70	7.15		
		High	12.00	9.43		
	Vocabulary	Low	6.00	3.32	0.177	0.030
		Medium	8.25	4.15		
		High	10.88	3.27		
	Listening	Low	5.45	4.32	0.098	0.157
		Medium	6.35	4.66		
		High	9.50	4.78		
	Reading	Low	2.18	1.54	0.046	0.427
		Medium	3.50	3.59		
		High	3.75	2.76		
	Speaking	Low	3.333	1.6073	0.235	0.102
		Medium	6.182	2.3798		
		High	6.625	1.9738		
Total	Low	28.45	8.69	0.036	0.518	
	Medium	30.80	15.99			
	High	36.13	16.52			
CLIL	Use of English	Low	23.06	10.33	0.021	0.161
		Medium	25.28	10.48		
		High	21.93	9.98		
	Vocabulary	Low	8.97	2.95	0.143	0.000
		Medium	10.05	3.17		
		High	11.96	2.96		
	Listening	Low	5.11	2.82	0.275	0.000
		Medium	5.95	3.47		
		High	10.58	4.82		
	Reading	Low	3.56	1.99	0.214	0.000
		Medium	4.58	2.72		
		High	7.62	4.18		
	Speaking	Low	6.692	1.5075	0.125	0.141
		Medium	7.688	1.6677		
		High	8.083	2.0595		
Total	Low	40.69	12.87	0.116	0.000	
	Medium	45.86	11.65			
	High	52.09	12.89			

**Table 8** Comparison across cohorts in terms of type of school (ANOVA)

Group	Skills	Type of school	Mean	Standard Deviation	Cohen's d	p-value
General	Use of English	Public school non-CLIL	13.34	8.32	0.252	0.000
		Public school CLIL	25.88	10.62		
		Private school CLIL	16.98	5.72		
	Vocabulary	Public school non-CLIL	7.98	4.15	0.271	0.000
		Public school CLIL	9.65	3.06		
		Private school CLIL	13.34	1.98		
	Listening	Public school non-CLIL	6.56	4.71	0.556	0.000
		Public school CLIL	5.46	3.08		
		Private school CLIL	13.96	1.66		
	Reading	Public school non-CLIL	3.15	2.96	0.455	0.000
		Public school CLIL	4.14	2.50		
		Private school CLIL	9.81	3.39		
	Total	Public school non-CLIL	31.02	15.24	0.249	0.000
		Public school CLIL	45.12	13.41		
		Private school CLIL	54.09	10.41		

students, in line with previous investigations which attest to the superiority of CLIL over EFL programmes (Lasagabaster & Ruiz de Zarobe, 2010; Lasagabaster, 2009; Pascual Bajo, 2018) (cf. Table 9).



**Table 9** Comparison in terms of type of school (ANOVA)

Group	Skills	Type of school	Mean	Standard Deviation	Cohen's d	p-value
Primary Education	Use of English	Public school non-CLIL	10.94	7.20	0.131	0.002
		Public school CLIL	13.35	7.26		
		Private school CLIL	16.98	5.72		
	Vocabulary	Public school non-CLIL	11.13	2.96	0.138	0.002
		Public school CLIL	11.20	3.94		
		Private school CLIL	13.34	1.98		
	Listening	Public school non-CLIL	11.88	2.25	0.273	0.000
		Public school CLIL	11.50	2.12		
		Private school CLIL	13.96	1.66		
	Reading	Public school non-CLIL	5.25	3.36	0.213	0.000
		Public school CLIL	7.35	3.90		
		Private school CLIL	9.81	3.39		
	Total	Public school non-CLIL	39.19	14.52	0.206	0.000
		Public school CLIL	43.40	16.06		
		Private school CLIL	54.09	10.41		
Compulsory Secondary Education	Use of English	Public school non-CLIL	14.88	8.762	0.246	0.000
		Public school CLIL	28.27	9.434		
		Private school CLIL				
	Vocabulary	Public school non-CLIL	5.96	3.518	0.174	0.000
		Public school CLIL	9.35	2.784		

(continued)

**Table 9** (continued)

Group	Skills	Type of school	Mean	Standard Deviation	Cohen's d	p-value
		Private school CLIL				
	Listening	Public school non-CLIL	3.16	1.700	0.084	0.001
		Public school CLIL	4.30	1.455		
		Private school CLIL				
	Reading	Public school non-CLIL	1.80	1.658	0.165	0.000
		Public school CLIL	3.52	1.513		
		Private school CLIL				
	Total	Public school non-CLIL	25.80	13.515	0.264	0.000
		Public school CLIL	45.45	12.915		
		Private school CLIL				

### 4.3 Durability of Effects of the CLIL Programme on FL Competence

The across- and within-group comparisons presented above are also complemented with general and group-focused analyses in order to confirm if the effects of CLIL remained once the programme was discontinued or if, on the contrary, they gradually disappeared. Hence, vis-à-vis RQ3, a comparison between the results obtained in the post-test phase and the outcomes of the delayed post-tests sat by the same students six months later was carried out.

Starting with the analysis of the CLIL group, our findings indicate that students obtained slightly better results on most skills tested, although no statistically significant differences were detected except for listening ( $p = < 0.001$ ). Slightly worse results were found for use of English in the delayed post-test phase, although still not significant (cf. Table 10).

Turning now to the analysis of the non-CLIL students, no statistically significant differences were detected between the two phases except for the reading skill ( $p = 0.007$ ). Moreover, slightly worse results were detected in the delayed post-tests for use of English and listening, something which clearly differs from the results obtained in the experimental group. Non-CLIL pupils were also found to have a

**Table 10** Post- to delayed post-test comparison of CLIL cohort's skills

Group	Skills	Post/Delayed	Mean	Standard Deviation	Cohen's d	p-value
CLIL	Use of English	Post	33,51	9,456	0.031	0.687
		Delayed	33,22	8,792		
	Vocabulary	Post	10,43	3,373	-0.216	0.097
		Delayed	11,08	2,637		
	Listening	Post	4,71	1,683	-0.393	<0.001
		Delayed	5,35	1,535		
	Reading	Post	4,20	1,472	-0.208	0.213
		Delayed	4,51	1,474		
Total	Post	52,86	13,519	-0.100	0.106	
	Delayed	54,16	12,686			

greater knowledge of vocabulary, although no statistically significant differences were ascertained for this skill (cf. Table 11).

After having expounded on the results obtained in the post-test phase by both CLIL and non-CLIL students, let us now analyse if statistically significant differences emerge across the cohorts. Clear-cut tendencies were discerned, pointing to the supremacy of the bilingual programme, as statistically significant differences emerged on all skills considered in favour of CLIL pupils.

As can be observed in the table below, CLIL students outperformed their non-CLIL peers on use of English ( $p = < 0.001$ ) and on their knowledge of vocabulary ( $p = < 0.001$ ) in the FL. In the same vein, the means obtained for both listening and reading skills were significantly higher for CLIL students (5.35 and 4.51, respectively), coinciding with previous studies by Pladevall-Ballester and Vallbona (2016) and Pascual Bajo (2018). However, these results must be interpreted with caution, since the large effect sizes shown by Cohen's  $d$  have to be taken into consideration in our assessment (cf. Table 12).

**Table 11** Post- to delayed post-test comparison of non-CLIL cohort's skills

Group	Skills	Post/Delayed	Mean	Standard Deviation	Cohen's d	p-value
Non-CLIL	Use of English	Post	15.58	5.551	0.031	0.899
		Delayed	15.42	5.248		
	Vocabulary	Post	5.33	2.348	-0.682	0.056
		Delayed	6.83	2.038		
	Listening	Post	3.00	1.348	0.306	0.422
		Delayed	2.58	1.379		
	Reading	Post	1.33	1.303	-1.201	0.007
		Delayed	2.83	1.193		
	Total	Post	25.25	7.124	-0.334	0.281

**Table 12** Comparison per group (T-test)

Group	Skills	Group	Mean	Standard Deviation	Cohen's d	p-value
General	Use of English	Non-CLIL	15.42	5.25	-2.159	<0.001
		CLIL	33.22	8.79		
	Vocabulary	Non-CLIL	6.83	2.04	-1.675	<0.001
		CLIL	11.08	2.64		
	Listening	Non-CLIL	2.58	1.38	-1.834	<0.001
		CLIL	5.35	1.54		
	Reading	Non-CLIL	2.83	1.19	-1.176	<0.001
		CLIL	4.51	1.47		
Total	Non-CLIL	27.67	7.36	-2.231	<0.001	
	CLIL	54.16	12.69			

#### 4.4 *Language Competence Differential: Discriminant Analyses*

Finally, the effect of the different intervening variables considered is quantified. Successive discriminant analyses have helped us determine which variables are the most significant in explaining the differences detected between the CLIL and non-CLIL strands.

As for the differences found between the experimental and the control group in English, it can be clearly seen that the independent variable (Group—CLIL) together with the moderating variables of socioeconomic status (SES) and verbal intelligence are the ones which display the greatest significance ( $p = 0.000$ ). Accordingly, these variables were later used in a discriminant function, which proved that all of them were significant, as the  $p$ -value obtained was 0,000. Thus, we can confirm that the CLIL programme has the greatest weight in explaining the language competence differential between the experimental and control groups, mirroring Pascual Bajo's (2018) and Pérez Cañado's (2018) outcomes (cf. Tables 13 and 14).

## 5 Conclusion

This study has addressed one of the most important current areas of interest in Second Language Acquisition (SLA) research: the analysis of how CLIL is playing out in a Spanish monolingual region which lacks a firm tradition for foreign language learning. More concretely, the investigation has been developed in the province of Jaén, an area where little research on the topic has been published so far. In order to overcome the main lacunae presented by previous investigations into the topic in terms of homogeneity, variables or statistical analyses, the project has worked

**Table 13** Test of equality group means

	Wilks' Lambda	F	df1	df2	Sig.
Exposure to English	.991	2.013	1	214	.157
SES	.854	36.672	1	214	.000
Verbal Intelligence	.837	41.827	1	214	.000
Goodwill	.999	.138	1	214	.711
Anxiety	.992	1.660	1	214	.199
Lack of interest	.954	10.400	1	214	.001
Selfcommitment	.986	3.050	1	214	.082
Type	.864	33.604	1	214	.000
Group	.940	13.670	1	214	.000
Gender	.998	.425	1	214	.515

**Table 14** Canonical discriminant functions

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.362 <sup>a</sup>	100.0	100.0	.515
<b>Test of Functions</b>	<b>Wilks' Lambda</b>	<b>Chi-square</b>	<b>Df</b>	<b>Sig.</b>
1	.734	65.586	3	.000

with students from three different educational levels (Primary Education, Compulsory Secondary Education and Baccalaureate), divided into two different cohorts according to the educational programme they are following (experimental CLIL group and control non-CLIL group) and taking into account different moderating variables.

Regarding RQ1, our outcomes have allowed us to confirm the superior linguistic competence in English of those learners following the CLIL programme. As detailed in Sect. 1 in Chapter “Are CLIL Settings More Conducive to the Acquisition of Digital Competences? A Comparative Study in Primary Education”, the CLIL cohort outperformed the non-CLIL stream on all skills analysed, being listening the only skill for which no statistically significant differences were detected. The oral production of CLIL and non-CLIL students was also analysed, taking into account their use of English (grammar and vocabulary), their pronunciation and fluency and their adequacy to the task. The outcomes obtained attest to the superiority of CLIL students, especially in their knowledge and use of grammar and in their pronunciation in the FL (Madrid & Barrios, 2018; Nieto Moreno de Diezmas, 2016; Pascual Bajo, 2018; Ruiz de Zarobe, 2008). However, no statistically significant differences appeared when analysing learners' lexical range, fluency or their adaptability to the task presented.

A total confirmation of these results was provided when evaluating students separately according to their educational stage. Thus, when comparing 4th year of CSE students following a traditional EFL programme with their CLIL peers, it was proved

that the latter outperformed the former on all the skills and aspects considered. On the contrary, slightly more negative results were obtained regarding the oral production of 6th year Primary Education students, since no differences appeared between cohorts.

In line with RQ2, the data obtained by means of the English tests was also analysed taking into consideration the different intervening variables considered. Valuable conclusions can be drawn in this respect, as numerous differences arose across and within cohorts. From a global perspective, no differences were detected in terms of gender, while the variable of setting offered statistically significant differences in favour of students studying in an urban context. However, students from rural centres obtained better results when analysing their receptive skills. Regarding socioeconomic status, statistically significant differences arose in favour of students with a high socioeconomic status on all skills considered, except for use of English, corroborating Anghel, Cabrales and Carro's (2016) and Fernández Sanjurjo et al. (2019) outcomes, according to which bilingual programmes affected negatively those students whose socioeconomic status was lower. No statistically significant differences appeared among non-CLIL learners in any of the skills evaluated, except for their knowledge of vocabulary. Our outcomes thus run counter to the tendency detected by some of the most recent research (Lorenzo, 2019; Pavón Vázquez, 2018; Pérez Cañado, 2017b; Rascón Moreno & Bretones Callejas, 2018), where CLIL has cancelled out differences among social classes.

In the overall comparisons, the type of school variable yielded differences on all skills considered in favour of those students belonging to a private bilingual school, except for use of English, in which learners from public bilingual schools obtained better outcomes. The same tendency was repeated when analysing the results from Primary schools, since students from private centres obtained better results in vocabulary, use of English, listening and reading. In Secondary Education, the situation changed, as no private bilingual schools were considered in our research. Consequently, statistically significant differences were found on all skills tested in favour of the CLIL group from public centres, corroborating Pérez Cañado's (2018) outcomes.

Vis-à-vis RQ3, the results obtained in the post-test phase were compared with the outcomes of the delayed post-test phase. No statistically significant differences were found for the CLIL group for any of the skills tested, except for listening. That is, although CLIL students obtained slightly better results in the delayed post-tests on all the skills considered, the differences were only significant for their oral receptive skill. In the case of non-CLIL students, no statistically significant differences appeared between the two phases, with the exception of reading. Differences in means were also detected for vocabulary, in favour of the delayed post-tests, although they did not reach statistical significance. However, slightly worse results were obtained in the delayed-post phase for use of English and listening, something which clearly differs from CLIL group's results.

In the across-cohort comparisons, the results obtained attest to the supremacy of the bilingual programme, since significant differences arose for all skills considered in favour of CLIL pupils. Coinciding with the results of previous studies (Lorenzo, 2019; Pascual Bajo, 2018; Pladevall-Ballester & Vallbona, 2016), CLIL students

outperform their EFL counterparts on the receptive skills as well as in use of English and knowledge of vocabulary.

Finally, regarding the last RQ, the successive discriminant analyses performed have confirmed that CLIL programmes, together with motivation and SES, are the variables which best explain the differences detected.

As a conclusion, we can affirm our data point to the general improvement of skills in the FL. Hence, our results support the continuity of CLIL programmes in post-Secondary stages, something which would help to consolidate their positive effects among students. In line with Pérez Cañado (2018), our results indicate that, although many of the effects of bilingual education remain, these can gradually disappear if the programmes are discontinued. That is why it is highly recommendable to promote their continued implementation.

**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. *English Educational Research and Evaluation*, 12(1), 75–93.
- Anghel, B., Cabrales, A., & Carro, J. M. (2016). Evaluating a bilingual education programme in Spain: The impact beyond foreign language learning. *Economic Inquiry*, 54(2), 1202–1223.
- Bruton, A. (2011). Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore (2010). *Applied Linguistics*, 32(2), 236–241. <https://doi.org/10.1093/applin/amr007>.
- Bruton, A. (2013). CLIL: Some of the reasons why... and why not. *System*, 41(3), 587–597. <https://doi.org/10.1016/j.system.2013.07.001>.
- Bruton, A. (2015). CLIL: Detail matters in the whole picture: More than a reply to J. Hüttner and U. Smit (2014). *System*, 53, 119–128. <https://doi.org/10.1016/j.system.2015.07.005>.
- European Commission. (1995). *White paper on education and training: Teaching and learning: Towards the learning society*. [http://europa.eu/documents/comm/white\\_papers/pdf/com95\\_590\\_en.pdf](http://europa.eu/documents/comm/white_papers/pdf/com95_590_en.pdf). Accessed 20 October 2017.
- 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>.
- 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>.
- Hüttner, J., & Smit, U. (2014). CLIL (Content and Language Integrated Learning): The bigger picture: A response to: A. Bruton. 2013. CLIL: Some of the reasons why...and why not. *System*, 44, 160–167.

- Jäppinen, A. K. (2005). Thinking and content learning of Mathematics and Science as cognitive development in Content and Language Integrated Learning (CLIL): Teaching through a foreign language in Finland. *Language and Education*, 19(2), 147–168.
- Jiménez Catalán, R. M., Ruiz de Zarobe, Y., & Cenoz, J. (2006). Vocabulary profiles of English foreign language learners in English as a subject and as a vehicular language. *VIEWZ (Vienna English Working Papers)*, 15(3), 23–27.
- Lancaster, N. K. (2015). *The effects of Content and Language Integrated Learning on the oral skills of Compulsory Secondary Education students: A longitudinal study*. Jaén: Universidad de Jaén.
- Lasagabaster, D. (2009). The implementation of CLIL and attitudes towards trilingualism. *ITL, International Journal of Applied Linguistics*, 159, 23–45.
- Lasagabaster, D., & Doiz, A. (2016). CLIL students' perceptions of their language learning process: Delving into self-perceived improvement and instructional preferences. *Language Awareness*. <https://doi.org/10.1080/09658416.2015.1122019>.
- Lasagabaster, D., & Ruiz de Zarobe, Y. (2010). Ways forward in CLIL: Provision issues and future planning. In D. Lasagabaster & Y. Ruiz de Zarobe (Eds.), *CLIL in Spain: Implementation, results and teacher training* (pp. 278–295). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Lorenzo, F. (2019). *Educación Bilingüe en Andalucía: Informe de Gestión, Competencias y Organización*. Junta de Andalucía: Consejería de Educación.
- 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.
- Madrid, D., Bueno, A., & Ráez, J. (2018). Investigating the effects of CLIL on language attainment: Instrument design and validation. In M. L. Pérez Cañado (Eds.), *Content and Language Integrated Learning in monolingual settings: New insights from the Spanish context*. Amsterdam: Springer.
- Nieto Moreno de Diezmas, E. (2016). The impact of CLIL on the acquisition of language competences and skills in L2 in primary education. *International Journal of English Studies*, 16(2), 81–101.
- Ouazizi, K. (2016). The effects of CLIL education on the subject matter (mathematics) and the target language (English). *Latin American Journal of Content and Language Integrated Learning*, 9(1), 110–137. <https://doi.org/10.5294/laclil.2016.9.1.5>.
- Paran, A. (2013). Content and language integrated learning: Panacea or policy borrowing myth? *Applied Linguistics Review*, 4(2), 317–342. <https://doi.org/10.1515/applirev-2013-0014>.
- Pascual Bajo, C. (2018). *The impact of Content and Language Integrated Learning (CLIL) on Valencian Secondary Education students' English language acquisition and academic achievement: A case study*. Valencia: Universidad de Valencia.
- Pavón Vázquez, V. (2018). Learning outcomes in CLIL programmes: A comparison of results between urban and rural environments. *Porta Linguarum*, 29, 9–28.
- Pelechano Barberá, V. (1994). Prueba MA. *Análisis y Modificación de la Conducta*, 20, 71–72.
- Pérez Cañado, M. L. (2011). The effects of CLIL within the APPP: Lessons learned and ways forward. In R. Crespo & M. A. García de Sola (Eds.), *Studies in honour of Ángeles Linde López* (pp. 389–406). Granada: Universidad de Granada.
- Pérez Cañado, M. L. (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. L. (2017a). Stopping the 'pendulum effect' in CLIL research: Finding the balance between Pollyanna and Scrooge. *Applied Linguistics Review*, 8(1): 79–99. doi.org/<https://doi.org/10.1515/applirev-2016-2001>.
- Pérez Cañado, M. L. (2017b). The evolution of bilingual education in monolingual settings: An Andalusian case study. In M. Jedyneck & P. Romanowski (Eds.), *The many faces of bilingualism-living with two languages* (pp. 207–241). Amsterdam: Springer.
- Pérez Cañado, M. L. (2018). CLIL and educational level: A longitudinal study on the impact of CLIL on language outcomes. *Porta Linguarum*, 29, 51–70.



- Pérez Cañado, M. L., & Lancaster, N. K. (2017). The effects of CLIL on oral comprehension and production: A longitudinal case study. *Language, Culture and Curriculum*, 30(3), 300–316. doi.org/<https://doi.org/10.1080/07908318.2017.1338717>.
- Pérez Cañado, M. L., & Ráez Padilla, J. (2015). Introduction and overview. In D. Marsh, M. L. Pérez Cañado, & J. Ráez Padilla (Eds.), *CLIL in action: Voices from the classroom* (pp. 1–12). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Piesche, N., Jonkmann, K., Fiege, C., & Kebler, J. (2016). CLIL for all? A randomized controlled field experiment with sixth-grade students on the effects of content and language integrated science learning. *Learning and Instruction*, 44, 108–116.
- Pladevall-Ballester, E., & Vallbona, A. (2016). CLIL in minimal input contexts: A longitudinal study of primary school learners' receptive skills. *System*, 58, 37–48.
- Rallo Fabra, L., & Jacob, K. (2015). Does CLIL enhance oral skills? Fluency and pronunciation errors by Spanish-Catalan learners of English. In M. Juan-Garau & J. Salazar Noguera (Eds.), *Content-based language learning in multilingual educational environments*, 63–177. Berlin: Springer.
- Rascón Moreno, D., & Bretones Callejas, C. M. (2018). Socioeconomic status and its impact on language and content attainment in CLIL contexts. *Porta Linguarum*, 29, 115–136.
- Rossell, C., & Baker, K. (1996). The educational effectiveness of bilingual education. *Research in the Teaching of English*, 30(1), 7–74.
- Ruiz de Zarobe, Y. (2007). CLIL in a bilingual community: Similarities and differences with the learning of English as a foreign language. *Vienna English Working Papers*, 16(3), 47–53.
- Ruiz de Zarobe, Y. (2008). CLIL and foreign language learning: A longitudinal study in the Basque Country. *International CLIL Research Journal*, 1, 60–73.
- Ruiz de Zarobe, Y. (2011). Which language competencies benefit from CLIL? An insight into applied linguistic research. In Y. Ruiz de Zarobe, J. Sierra, & F. Gallardo del Puerto (Eds.), *Content and foreign language integrated learning* (pp. 129–153). Bern: Peter Lang.
- Santamaría, P., Arribas, D., Pereña, J., & Seisdedos, N. (2016). *EFAI: Evaluación Factorial de las Aptitudes Intelectuales*. Madrid: TEA Ediciones.
- Serra, C. (2007). Assessing CLIL at primary school: A longitudinal study. *International Journal of Bilingual Education and Bilingualism*, 10(5), 582–602.
- Surmont, J., Struys, E., Van Den Noort, M., & Van De Craen, P. (2016). The effects of CLIL on mathematical content learning: A longitudinal study. *Studies in Second Language Learning and Teaching*, 6(2), 319–337.
- Veiz, J. M. (2009). Multilingual education in Europe: Policy developments. *Porta Linguarum*, 12, 7–24.
- Wolff, D. (2005). Approaching CLIL. In D. Marsh (Eds.), *The CLIL quality matrix: Central workshop report*, coord (pp. 10–26). [http://www.ecml.at/mtp2/CLILmatrix/pdf/wsrepD3E2005\\_6.pdf](http://www.ecml.at/mtp2/CLILmatrix/pdf/wsrepD3E2005_6.pdf). Accessed 20 October 2017.