

The Effects of a CLIL Programme on Linguistic Progress at Two Different Points in Time

Marta Segura¹, Helena Roquet¹, Carmen Pérez-Vidal²

¹ Universitat Internacional de Catalunya

² Universitat Pompeu Fabra

Correspondence concerning this article should be addressed to Marta Segura Molla, Josep Trueta, s/n, 08195, Sant Cugat del Vallès, Spain. E-mail: mseguram@uic.es

In an attempt to explore the effects of different kinds of English as a Foreign Language (EFL) learning contexts, content and language integrated learning (CLIL) have been at the centre of FL acquisition research over the past decade. Studies have focused on the features and gains this setting brings, whether content is learnt at the same level of success as when taught in the learners' L1, and whether that L1 is negatively affected by CLIL. However, to our knowledge, very little attention has been brought to how the seniority of the programme affects learner progress in the target language. This study aims to fill such a gap in the understanding that the programme will have developed and improved in terms of quality of exposure and interaction, and that learners' EFL performance will be higher. To do that, we measured the efficacy of a long-standing CLIL programme in Barcelona twelve years after it was launched and examined the reading, writing, and lexico-grammatical abilities of CLIL EFL learners aged 8, 11, and 14 compared with results obtained by learners measured at the onset of the programme in 2005. The results showed that the quality of the programme has increased over the last decade, guaranteeing a higher level of EFL student proficiency when raw scores are considered, but not in terms of linguistic gains, in which only improvement in older students' grammar and reading skills can be observed.

Keywords: content and language integrated learning (CLIL), CLIL programme assessment, second language acquisition, English as a foreign language (EFL), language learning

Introduction

Over the last few decades, instructed second language acquisition has experienced a change towards more meaningful activities in classrooms in order to generate a more natural and authentic learning context. In light of such developments, Content and Language Integrated Learning (CLIL) has gained ground in Europe¹, and in this respect, Catalonia, in Spain, is a case in point. It is believed that through CLIL approaches, in which a curricular subject is taught through the medium of a foreign language (FL), learners are immersed in more meaningful communication than in conventional formal instruction (FI). Within CLIL, English as a FL (EFL) learners are exposed to a higher quantity and quality of input, and have more opportunities for interaction. Even if most research shows that CLIL programmes result in benefits for learners (APAC, 2006; Brinton, Wesche, & Snow, 1989; Lasagabaster, 2008; Pérez-Vidal, 2013; Roquet, 2011; Van de Craen & Pérez-Vidal, 2001), more research is needed to assess their real impact in different settings.

Therefore, the goal of the present study is to contribute to this area of research with data obtained from a CLIL programme with English as the medium of instruction in a mainstream educational institution in the city of Barcelona. The study seeks to examine the impact of the school's CLIL programme twelve years after it was launched to analyse its effects as a programme that is no longer at its pilot stage, but that has been running for over a decade. This date is important for the main objective of the study; the evaluation of the programme's efficacy in bringing along higher levels of proficiency in EFL. In this study, we examine CLIL learners' abilities from 8 to 14 years of age, with data collected in 2017, and cross-sectionally contrast it with data collected at the

¹ CLIL programmes have developed more recently in South American countries and Australia but fall outside the scope of the current study.

onset of the programme in 2005 (Roquet, 2011). This should allow us to gauge the expected beneficial impact of the programme and assess its development over the years. The present study thus represents a unique opportunity to examine in detail a CLIL programme that is no longer in its infancy (Pérez-Vidal, Escobar, & Roquet, 2013).

Thus, the present research aims at providing new and meaningful data regarding a specific CLIL programme, namely in terms of the effects of programme lengths, as seen through the students' level and gains in reading, writing, and lexico-grammatical skills. Although CLIL is a branch of EFL acquisition that has been largely researched in recent decades, many studies have stated the infancy of the CLIL programmes analysed as a research limitation. Therefore, the main aim of this paper is to fill in this gap and compare one CLIL programme at two points in time: a couple of years after it was implemented, and twelve years later, when the programme is more established and stable.

Literature Review

The present section is devoted to briefly presenting an overview of the existing CLIL literature to provide a context for the relevance of the current research, starting with a definition of the term CLIL, followed by a description of its main features. We then briefly sketch out the specific context of the study and present a description of the onset of CLIL programmes in Europe. We close this section by reviewing the effects of CLIL on linguistic progress and the possible reasons explaining them, with a special emphasis on how long programmes have been running.

Definition and Features of CLIL

CLIL can be defined as an approach in which 'a language different from the domestic languages is used as the medium of instruction for curricular subjects' (Pérez-Vidal & Roquet, 2015). This way the target language is used in a context of real and meaningful communication (Llinares & Morton, 2010; Lorenzo, Casal, & Moore, 2010). While being the medium of instruction, students are 'learning language through language use' (Brinton, Wesche, & Snow, 1989) by engaging in meaningful and critical interactions (Darvin, Yi Lo, & Lin, 2020), while the curricular subjects taught through CLIL are also object of attention (Pérez-Vidal, 2013). What needs to be highlighted here is that the main goal in CLIL is to teach both the FL and the curricular content in an integrated way (Lasagabaster, 2008). Therefore, a CLIL approach should include both content and language objectives.

As a result, the extant limited amount of hours of L2 exposure in the conventional educational curricula is increased and schools become a multilingual environment, where languages are to be presented as a whole (San Isidro & Lasagabaster, 2019). Therefore, through CLIL approaches, languages are handled in an integrated way in schools. Instead of being taught and used separately and independently from each other, CLIL is intended to raise students' general language awareness (Navarro-Pablo & López Gándara, 2019) and promote translanguaging (García, Aponte, & Le, 2020), that is, the alternation from one language to another (Garcia, 2009) and their integration for different pragmatic purposes, as well as pluriliteracies, as each different language is used for a specific curricular subject in parallel to developing learners' L1s (Nikula & Moore, 2015). In sum, CLIL ultimately promotes plurilingualism with a wide scope (Dalton-Puffer, 2008; Nikula & Moore, 2015).

There is a debate, which is yet not resolved, regarding the extent to which the CLIL subject adopts the L1 cultural approach and pedagogy, considering that usually both students and the teacher share the same L1. Critical to CLIL is the fact that the teachers' and students' L2 proficiency tends to be limited. CLIL programmes also show some variable characteristics that make each CLIL programme slightly different, such as timing, continuation, support for students, resources for teachers, institutional implication, social spreading, and the status of the L2 language and culture (Pérez-Vidal, 2013). Another critical variable feature of CLIL programmes is their seniority; many CLIL research has been conducted with programmes that were still at their infancy, hence, in their pilot phases (Moore, 2009). However, more research is needed to determine whether CLIL programmes that are no longer in their pilot phase can lead to higher benefits for the students in terms of their L2 learning process.

CLIL: from Europe to Spain

Since the onset of the European Union in 1949, the European Commission has shown great concern for Europe's linguistic heritage and cultural diversity, promoting citizens' mobility, linguistic diversity, and multilingualism. With the aim of guaranteeing that citizens in Europe become functionally proficient in their mother tongue and in two other European languages (Cenoz, 2015; Llinares & Morton, 2010; Pérez-Vidal, 2007), it promoted an early starting age of instruction in a second language, an intensive and transdisciplinary approach starting in primary school, the addition of a third language in secondary education taught intensively and trans-disciplinarily, and a university system guaranteeing mobility through ERASMUS (Pérez-Vidal, 2013).

As a consequence, the role and relevance of 'context of learning' began to gain ground in the shape of interdisciplinary teaching approaches (Gené Gil, 2016; Llinares & Morton, 2010; Pérez-Vidal & Juan-Garau, 2012). Thus, CLIL appeared to be the most suitable approach since it provides a real, student-centred, and meaningful context for language learning. Consequently, the CLIL approach clearly gained ground at all education levels in many countries (Lorenzo, Casal, & Moore, 2010; Pérez-Vidal, 2007; Van de Craen & Pérez-Vidal, 2001).

Concerning the specific context of the present research, CLIL programmes have been launched and implemented in the majority of Spanish autonomous territories (Fernández-Sanjurjo, Fernández-Costales, & Arias Blanco, 2017; Pérez-Vidal, Lorenzo, & Trenchs, 2015). This is indeed the case in Catalonia, where the current study takes place: a bilingual community where semi-immersion programmes have gained ground by using English as FL as a medium of instruction of other curricular subjects. Semi-immersion programmes, such as CLIL, allow for more real and meaningful contact with the FL within the classroom while still learning the language in a FL context where there is little or no contact with the language outside of the educational setting. Within this background, in recent years, the number of schools implementing CLIL programmes in Catalonia has grown considerably.

Effects of CLIL Programmes

It has been stated that CLIL brings many benefits to learners and many different CLIL programmes have been analysed to see which linguistic skills benefit most from such programmes. (APAC, 2006; Brinton, Wesche, & Snow, 1989; Lasagabaster, 2008; Pérez-Vidal, 2013; Roquet, 2011; Van de Craen & Pérez-Vidal, 2001). There is however one critical aspect in any CLIL programme, the length of time a programme has been running, that has been scarcely considered when interpreting data. Moore (2009) already stressed the fact that much of the research on the impact of CLIL was being conducted with programmes that were barely in their pilot phases and that we needed to give them time to settle in schools and be refined.

In contrast with the pilot programmes that have been analysed in many studies in the last several decades, some of which will be summarized later on, a well- and fully-established CLIL programme would show more stability and complexity, which is a potential key factor for the programme's success (Navés, 2009). Developing and implementing a new CLIL programme is a complex task that requires the teachers to be trained and new materials to be created. After this pilot stage in which the programme is first set and designed, CLIL programmes that have been running for some years will have more experienced teachers who bring consistency to the programme, as well as experience when designing and refining all aspects of its actual implementation (Navés, 2009; Navés & Muñoz, 1999).

Therefore, a CLIL programme that would have been running for many years in the same school is expected to be more robust and have more experienced teachers, which should potentially grant a more stable and complex design and application of the programme. In their study, Navarro-Pablo and López Gándara (2019) suggest that the CLIL group's higher scores may be attributed to the years that the programme has been implemented, the experience of the teachers, and the time students have been part of a CLIL programme. However, as mentioned above, most CLIL research has been conducted with CLIL programmes in their pilot phase. Thus, there is a need for more research to analyse the effects of a well-established CLIL programme as compared to a pilot phase one. This is precisely what the present study seeks to study: the analysis of a long-standing CLIL programme by measuring its efficacy in terms of student's linguistic gains correlated with the programme's maturity and growth.

One study, Roquet (2011), later summarized in Pérez-Vidal and Roquet (2015) and Roquet and Pérez-Vidal (2017), drew attention to this fact when the lack of statistically significant results was attributed to the pilot stage of the newly established programme. Indeed, these authors analysed CLIL learners compared to non-CLIL learners from the same school synchronically. The longitudinal study allowed for an analysis of the degree to which the learners' level of English progressed. All learners were administered five tasks: a written composition, a reading cloze, a dictation², a grammar and vocabulary error correction test, and a grammaticality judgement test. Roquet's (2011) results partially confirmed the hypothesis: the group with CLIL instruction yielded significantly better results in all the tests except the dictation test as compared to their peers of the same age and grade receiving formal instruction (FI) only. It was, therefore, concluded that a FI+CLIL learning context had proved to be more beneficial than a FI learning context only. However, the lack of statistically significant results in some of the tests administered led the authors to state that an academic year may not have been sufficient and, equally importantly, that the programme was still in its infancy when data was collected, at a 'pilot phase' stage, so that it might have still been too early for it to prove to be fully efficient.

To our knowledge, no study thus far has actually tried to prove that a well-established programme should yield better results when learners' proficiency gains are tested than at its onset³. In the wake of such an idea, the present study seeks to explore whether a fully established CLIL programme offered at the school in 2017 yielded significantly higher results in the form of more advanced EFL abilities at two different points in time (Roquet, 2011).

As mentioned above, most research regarding CLIL programmes and their effects on linguistic skills and language development focused on newly-established programmes, comparing the results of students involved in CLIL classes with the students attending only FI EFL lessons. It has been stated that CLIL brings many benefits to learners (APAC, 2006; Brinton, Wesche, & Snow, 1989; Lasagabaster, 2008; Pérez-Vidal, 2013; Roquet, 2011; Van de Craen & Pérez-Vidal, 2001). However, there is still no consensus regarding which areas or language abilities benefit most from it. Lasagabaster (2008) found that CLIL Basque students showed significantly higher oral and written proficiency than their FI peers. Moore (2009) also proved that certain competences are more likely to be improved with CLIL: receptive skills, vocabulary, morphology, fluency, creativity, risk-taking, and emotive-affective factors. Lorenzo, Casal, and Moore (2010) found that CLIL learners outperformed their FI peers in the four main linguistic skills both productively and receptively.

On the contrary, in a longitudinal study conducted with secondary education students, Admiraal, Westhoff, and de Bot (2005) did not find any significant differences between the CLIL and the FI groups in terms of receptive vocabulary. However, the CLIL students in the study did have higher results in speaking and reading, which is contradictory with Pladevall-Ballester and Vallbona (2016), who did not find any significant differences in reading in primary school students. Surprisingly, the same study also showed that the non-CLIL group outperformed the CLIL group in listening, although significant progress was seen in both groups.

Related to oral productive skills, Rallo Fabra and Juan-Garau (2011) found that CLIL students were understood better and showed less accented pronunciation than their FI peers. Such findings are in line with Ruiz de Zarobe's (2008) previous longitudinal study in which better results were seen in the CLIL group not only in terms of pronunciation, but also in vocabulary, grammar, and fluency. Nevertheless, in a later longitudinal study with secondary education students, Rallo Fabra and Jacob (2015) did not find any statistically significant differences between the CLIL and the FI groups in fluency on pronunciation.

In terms of content learning, previous research on CLIL around Europe has reported similar results between groups, or even favourable outcomes for the CLIL cohort in both the learning of the subject content (Hughes & Madrid, 2019) and the languages used as the medium of instruction, leading therefore to state that CLIL programmes do not seem to have a detrimental effect on the learners' L1 or on their achievement when

² The dictation test was not administered in 2017 for technical reasons.

³ Having compared the most relevant data bases (Scopus and JCR), no previous studies were found in which one same programme was compared at two points in time to analyze linguistic progress.

curricular content learning is assessed (Fernández-Sanjurjo, Fernández-Costales, & Arias Blanco, 2017; San Isidro & Lasagabaster, 2019).

However, in their study, Fernández-Sanjurjo, Fernández-Costales, and Arias Blanco (2017) reported opposite results, since non-CLIL students performed slightly better than their CLIL peers when assessing their science knowledge, the subject taught through CLIL. In another study by Hughes and Madrid (2019), primary and secondary CLIL and non-CLIL students' science knowledge was examined, showing slightly higher results in science content in the case of the non-CLIL group in primary education when compared to a CLIL group the same age, a trend which was however changed as the CLIL group reached secondary education. The authors claimed that this might have been the consequence of the younger primary-level learners having a lower FL level as they start following a CLIL programme, an effect which is then mitigated later on (Hughes & Madrid, 2019). Similar results were reported in a longitudinal study by De Dios Martínez Agudo (2019) in which CLIL students showed better results than their peers in all linguistic domains and this tendency kept growing until the end of secondary education.

In sum, it can be claimed that CLIL programmes have been shown to bring about benefits for the students enrolled on them when compared to their FI peers in terms of vocabulary and morphology as well as motivation and creativity (Dalton-Puffer, 2008; Lasagabaster, 2008; Lorenzo, Casal, & Moore, 2010; Pérez-Vidal, 2011). However, there are some other competences for which contradictory results have been found; these include syntax, writing, informal language, pronunciation, fluency, and pragmatics (Dalton-Puffer, 2008; Pérez-Vidal, 2011, Roquet & Pérez-Vidal, 2017). Therefore, more research is needed, namely in the form of longitudinal studies (Lasagabaster & Doiz, 2017; San Isidro & Lasagabaster, 2019) that analyse CLIL effects across all educational levels to see if said CLIL benefits are persistent until the end of secondary education (De Dios Martínez Agudo, 2019; Hughes & Madrid, 2019).

Besides the abovementioned linguistic features, some benefits regarding the students' entire educational process (Lorenzo, Casal, & Moore, 2010) and their knowledge and usage of their mother tongue have also been identified (APAC, 2006), such as greater problem-solving skills, more independence as learners and linguistic spontaneity (DeKeyser, 2000; Derakhshan & Karimi, 2015), and positive attitudes and an interest in learning (Lasagabaster & Sierra, 2009).

Such benefits probably come from the nature of the CLIL environment where students are exposed to high-quality input and opportunities for output practice and meaning negotiation are present (Roquet & Pérez-Vidal, 2017). Generally, the CLIL hours are extra hours of exposure and interaction in the FL (APAC, 2006; Dafouz Milne & Guerrini, 2009). In CLIL classrooms, learners are immersed in transdisciplinary lessons, usually including science, physical education, informatics, or arts and crafts content (Dafouz Milne & Guerrini, 2009; Pérez-Vidal & Roquet, 2015), which allows them to reduce the pressure to learn the language (Auerbach, 2006; Dalton-Puffer, 2008; Lasagabaster & Sierra, 2009; Van de Craen & Pérez-Vidal, 2001).

It is also important to note that, as pointed out by Bruton (2011) after reviewing several studies on CLIL, it seems that in many cases the students in the CLIL groups are self-selected, that is to say students who choose the CLIL option are more motivated, have a higher socio-economic family status, or have a higher linguistic level. Therefore, those CLIL benefits stated in the previous paragraphs may not only come from the communicative and more meaningful nature of such programmes, but also from the students' sociolinguistic background and initial proficiency level in the L2. However, Hüttner and Smit (2014) claimed that an approach such as CLIL, which increases student engagement in content and language learning simultaneously, attempts to diminish any discriminatory differences in learners' linguistic knowledge.

Finally, one other feature of CLIL pedagogy is that teachers' roles change from the FI conventional ones: the teacher stops being either the FL expert or the subject expert to actually become a combination of both (APAC, 2006), although sometimes the role of the content and the language teacher is still unclear (Darvin, Yi Lo, & Lin, 2020). In this respect, it has been claimed that teacher education in CLIL is a primary requirement. Information about CLIL as a multilingual approach, its challenges and advantages, the existing tools, and well-tested pedagogical practices should be subject to training. CLIL should spur collaboration between the programme's language and content teachers (Darvin, Yi Lo, & Lin, 2020; Jaén Campos, 2016; Papaja, 2013), as they will need to work together.

In sum, the previously mentioned conditions taken together, namely the meaningful context that a fully-established CLIL programme provides, can be said to be optimal for learners' language acquisition and non-linguistic positions vis-à-vis their target language in their language acquisition process (Dalton-Puffer, 2008; Kersten & Rohde, 2013). Nevertheless, many of the abovementioned studies were based on pilot programmes, which are naturally less stable since they are still developing. Therefore, further research is needed that analyses CLIL programmes that have been running for some years and that are more stable and robust. That is precisely what the present research aims at doing, analysing a fully-developed CLIL programme and contrasting how it affects students learning of the FL in comparison with the same programme in its pilot phase twelve years before.

The Present Study

The present study was conducted in a school in Barcelona, Catalonia, which had launched a well-planned CLIL programme in 2004-2005. Through the programme, learners were exposed to one extra subject, science, in English, in addition to the English FI hours. After designing the programme in the 2002-2003 school year, training the teachers, and informing the families, the programme was launched the next year with the third and fifth graders. In the following school years, the programme was extended to secondary education and now it stretches up to the first year of the Baccalauréat, the first of two years of preparation for university entry.

Parallel to the implementation of the programme, the school increased the number of EFL conventional sessions from three to four or five sessions per week, starting at the age of four. As a result, when pupils reached the third grade of primary, they had been exposed to 420 hours of English, which was considered enough to follow a science course in English. The school also made provisions to include the vocabulary needed to understand a science course in English during the English language sessions.

Our study took advantage of the fact that the impact of this same programme on learners' EFL proficiency two years after its implementation had already been analysed in 2005 (see Roquet, 2011), and full access to this data was granted. Thus, the data collected in 2005 were used and compared with the new data collected for an appraisal of the linguistic effects of the CLIL programme more than a decade later.

With that goal in mind, the following research questions were set:

- RQ1.** Did the 2017 CLIL learners display significantly higher linguistic abilities than the 2005 learners from the same academic grades and ages respectively, that is at the third and sixth grades of primary, and at the second grade of secondary?

We hypothesized that the 2017 cohort would yield better results because they would have benefited from an improved, more mature, and better-established CLIL programme.

- RQ2.** What are the differential gains in linguistic abilities (reading, writing, grammar, and vocabulary) between the 2005 cohorts and the 2017 cohort, when comparing cross-sectional gains between the third and sixth grades in primary and between the sixth grade in primary and the second grade in secondary?

We hypothesized that gains between the third and six grades in primary and between the sixth grade in primary and the second grade in secondary would be greater in the 2017 cohort than in the 2005 cohort due to the growth the programme would have experienced.

- RQ3.** What are the differential cross-sectional gains shown in the learners' linguistic abilities (reading, writing, grammar, and vocabulary) between the younger group (third grade in primary) and the eldest group, (second grade in secondary) when comparing the 2005 and the 2017 programmes?

We hypothesized that, as a result of the programme having evolved over the last decade, the 2017 cohort would show higher gains from the third grade of primary school to the second grade of secondary school when compared to the 2005 cohort.

Materials and Methods

In this section, the design is presented, the participants are described, and the materials and procedures for data collection and analyses are given.

Research Design

The current study cross-sectionally measured and contrasted CLIL learners following a CLIL programme consisting of a science subject taught in English, in addition to FI in that language. Students' linguistic abilities were measured in order to identify possible improvements in the programme.

As displayed in Table 1 below, learners were tested in two different years: 2005 and 2017. In both testing times, students were administered the tests in their third trimester, namely between April and May. The 2005 data is actually secondary data in this study as we are reanalysing Roquet's (2011) corpus, which has been contrasted with the new 2017 data for this study.

To that end, learners were tested at three different grades and ages coinciding with the onset, the middle, and the end of the CLIL programme. Data from the third grade in primary education serves as pre-test data, and data from the sixth grade in primary serves both as post-test data for the third graders and as pre-test data for learners in the second grade of secondary education.

Participants

The sample of participants analysed included a total of 90 students from the 2004-2005 cohort, and 90 students from the 2017 cohort. The participants in both groups were Catalan and Spanish bilingual learners from the same school and they were all following the school's CLIL programme, having contact with the English language not only in their FI class, but also in their science CLIL lessons. Thus, students tested in both testing times had had the same exposure to English in the school as their same-age peers from the other data collection cohort.

Participants were distributed in three subgroups according to testing time and educational level: the third year of primary (8-9 year olds), Group 1 in 2005 and Group 2 in 2017; the 6th year of primary (11-12 year olds), Group 3 in 2005 and Group 4 in 2017; and the second grade of secondary (13-14 year olds), Group 5 in 2005 and Group 6 in 2017. Table 1 below shows the distribution of the participants in groups. The team of English teachers and CLIL teachers involved in the programme in 2005 and in 2017 was a stable group. Furthermore, the main parameters of the programme were also kept the same, namely the CLIL programme coordinator was the same and so was the pedagogical approach, as confirmed by the school's management team, guaranteeing the programme's stability and development over the years.

Data were collected from all learners in those three grades at the two testing times, that is from 120 students per grade in primary and 90 students in the second grade of secondary. However, only a randomly selected sample of 30 students from each grade at each testing time was analysed, providing a stratified sample that included students from all the different groups in each school grade. Written consent to participate in the study was requested from all of the students' families through a form in which details regarding the research were presented. All students were tested, as the tests were also included as classroom activities of the school's English subject.

Table 1

Distribution of the participants in groups

Testing time 1 (2005)	Testing time 2 (2017)
Group 1 (N= 30) / 3 rd primary (8 and 9 years old)	Group 2 (N= 30) / 3 rd primary (8 and 9 years old)
Group 3 (N= 30) / 6 th primary (11 and 12 years old)	Group 4 (N= 30) / 6 th primary (11 and 12 years old)
Group 5 (N= 30) / 2 nd secondary (13 and 14 years old)	Group 6 (N= 30) / 2 nd secondary (13 and 14 years old)

Materials

A series of tests were administered to the participants at both testing times: a writing test, a grammar and vocabulary test, and a reading test. Those four main skills were chosen to be the focus of the testing since they provide an overall picture of the students' linguistic competence. However, speaking and listening were not included in the testing due to time and space constraints. The tests were kept the same for both testing times and all three grades tested to guarantee that the scores were comparable across levels and between both testing times. The instructions for all tests were written in Catalan to help understanding.

All tests were created by the researchers in 2005, following guidelines from the schoolteachers to ensure that they corresponded to the educational levels. The tests were piloted in 2005 before the first testing time and were kept the same in 2017. Thus, the tests were criterion-referenced and scored on an absolute scale.

The writing test was based on a picture that showed two policemen taking a statement from a woman and a young boy inside a flat. Students were asked to write three compositions in the space provided: one imagining a dialogue for the picture, another explaining what they thought had triggered the situation, and, finally, another guessing how the situation would end. The three questions that were given to the students were the following (they have been translated from Catalan into English):

Write a dialogue imagining the situation involving the policemen, the boy, and his mother.

Write a short paragraph describing why you think the situation you imagined in the dialogue actually took place.

Write a short paragraph explaining how you think this situation will end?

The grammar and vocabulary test consisted of two separate tasks. The first one was an error correction task. It included 30 statements grouped into three sections of 10 sentences each, with an increasing degree of difficulty, which needed to be corrected in a multiple-choice answer format by choosing either the correct option to fill in the blank in the sentence or selecting the correct sentence out of the three options given. Some examples of the questions included in the error correction test are given below:

_____? No, I'm sorry, I haven't got any change.

- A. Do you can change this 5 € note?
- B. Have you change for this 5 € note?
- C. Have you got change for this 5 € note?

I visited London last year.

- A. How long have you stayed there?
- B. How long did you stay there?
- C. How long you stay there?

How _____ people are there in class?

- A. much
- B. more
- C. many

Why are you putting on your coat?

- A. Because I took the dog for a walk.
- B. Because I'll take the dog for a walk.
- C. Because I'm going to take the dog for a walk.

The second grammar and vocabulary test, a grammaticality judgement test, consisted of 20 sentences of increasing difficulty for which the students had to decide whether they were correct or incorrect. Some examples of the questions included in the grammaticality judgement test are presented below:

She can't to sky.

- Correct
- Incorrect

I am taller than my brother.

- Correct
- Incorrect

The cat is sleep under the chair.

- Correct
- Incorrect

I am going to write a novel some day.

- Correct
- Incorrect

The reading test consisted of a cloze activity with a multiple-choice answer format. The learners were administered a text dealing with a topic studied in the science course with 20 missing words and four options for each of the blanks. The second paragraph (of three in total) of the reading test, with its corresponding possible answers, is presented below:

Tsunami is a Japanese word that (6) ‘harbour wave’. But why do tsunamis (7)? Tsunamis are usually caused by earthquakes at the bottom of the sea. At first, the (8) in the sea is quite small, but it moves very (9). When the wave gets close to the coast, the ocean floor makes it grow enormously. By the time it reaches the (10) it has become huge. Some tsunamis can be 30 meters (11). These giant waves can hit Japan, Indonesia, Central (12) and South America.

- | | | | | |
|------|---------|-----------|----------|---------|
| (6) | mean | signifies | means | tells |
| (7) | happens | happen | exists | go |
| (8) | wave | water | animals | one |
| (9) | quick | speedy | rapid | quickly |
| (10) | sea | cost | costal | coast |
| (11) | high | height | tall | of high |
| (12) | america | America | American | USA |

Procedure

Data collection procedures were kept basically the same in 2017 as in 2005. In 2005, the participants were tested in their classrooms over one two-hour session. During the first hour, three tests were administered: the error correction test (15 minutes), the grammaticality judgement test (10 minutes), and the writing test (20 minutes). In the second hour, the reading test (15 minutes)⁴ was completed. All tests were pen and paper, and teachers were given previous instructions such as to strictly control the time and to guarantee an exam-like environment.

In the second testing time, in 2017, the tests were administered over two separate sessions. The first one took place in the classrooms and included the writing test (20 minutes), which was a pen and paper test, as in 2005. The second session took place in the computer room and included the rest of the tests: the error correction test (15 minutes), the grammaticality judgement test (10 minutes), and the reading test (15 minutes), which were digitalized and administered through three different Google Forms questionnaires. Teachers were also given precise instructions for test administration. The change introduced in this data collection was a reflection of the well-established use of computers both in the school and in the learners’ individual work outside school hours, which made ‘pen-and-paper only’ tests obsolete to a certain extent, and perhaps less valid as a measure of proficiency.

In addition to the abovementioned tests, students were administered a socio-linguistic background questionnaire, tapping onto the participants’ linguistic background such as their L1 and their knowledge of other FLs. The results obtained from the sociolinguistic background questionnaire allowed for controlling external factors such as the students’ mother tongue and their participation in English extracurricular lessons. Thus, students with English as their mother tongue or enrolled in English extracurricular activities were

⁴ The reading test was not administered to the third grade of primary school group in 2005.

excluded from the sample of participants. However, even though those two main factors were controlled, it is important to underline that in 2017, as compared to 2005, there was more extramural contact in English, namely students had more access to resources in English in their daily lives.

Analyses

The following procedure was followed for the analysis. To correct the grammar and vocabulary tests, and the reading test, correct answers added points and incorrect answers did not. As for the writing test, learners' compositions were evaluated on the basis of Friedl and Auer's (2007) wholistic rating model, which includes four main areas (task fulfilment, organization, grammar, and vocabulary) rated along a scale where 0 is the lowest grade and 5 the highest. Interrater reliability was calculated on the basis of 10% of the compositions, which were corrected by two different ratters.

With the raw results obtained from the grading procedure explained above, statistical analyses were conducted, namely t-tests. All data was statistically analysed with Version 16 of IBM's SPSS Statistics package and the p value was set at ($p=0.05$).

Results

The first research question enquired into the comparative results obtained from the participants in 2005 and 2017 at the three educational levels analysed. It was hypothesized that the 2017 cohort would obtain significantly higher results than the 2005 cohort. When comparing the results in all tests, as shown in Table 2 and Figure 1 below, participants in the third grade of primary obtained higher grades in 2017 than in 2005. Regarding the writing test, the mean score obtained in 2005 was 1.05 out of 10 points; while in 2017 it was 1.92, the difference being statistically significant ($p=0.000$). In the error correction test, the mean in 2005 was 3.28; while in 2017, it was 4.60, with a statistically significant difference ($p=0.001$). In the grammaticality judgement test, the mean in 2005 was 3.90; while in 2017 the mean was 5.76, such difference being statistically significant ($p=0.000$). Concerning the reading test, no comparison was possible as this test was not administered in 2005 to this grade.

Table 2

Group statistics for third graders in 2005 (Group 1) and 2017 (Group 2)

Group		N	Mean	Std. Deviation	Std. Error	P-value
Writing test	1	30	1.050	.6345	.1158	.000
	2	30	1.917	1.0178	.1858	.000
Error correction test	1	30	3.283	1.3110	.2394	.001
	2	30	4.600	1.6785	.3064	.001
Grammaticality judgement test	1	30	3.900	1.9538	.3567	.000
	2	30	5.767	1.3309	.2430	.000
Reading test	1	30	.000	.0000	.0000	.000
	2	30	3.850	1.2536	.2289	.000

Regarding the results obtained by the sixth primary grade learners, as presented in Table 3 and Figure 2 below, higher grades were reported in 2017 for the writing and on the grammaticality judgement test. As for the writing test, the mean score in 2005 was 4.22; while in 2017, it was 4.63. In the grammaticality judgement test, the mean in 2005 was 5.34; while it was 6.04 in 2017. Such a difference was not statistically significant in the writing test ($p=0.128$) whereas it was significant for the grammaticality judgement test ($p=0.026$).

On the contrary, lower results were reported in 2017 than in 2005 for error correction and the reading tests. Regarding the error correction test, the mean score in 2005 was 6.82; while in 2017, it was 6.70, but this was not statistically significant ($p=0.729$). On the reading test, the mean in 2005 was 6.62; while in 2017, it was 5.82, the difference not being statistically significant either ($p=0.106$).

Figure 1
Test results for the third year of primary at both testing times

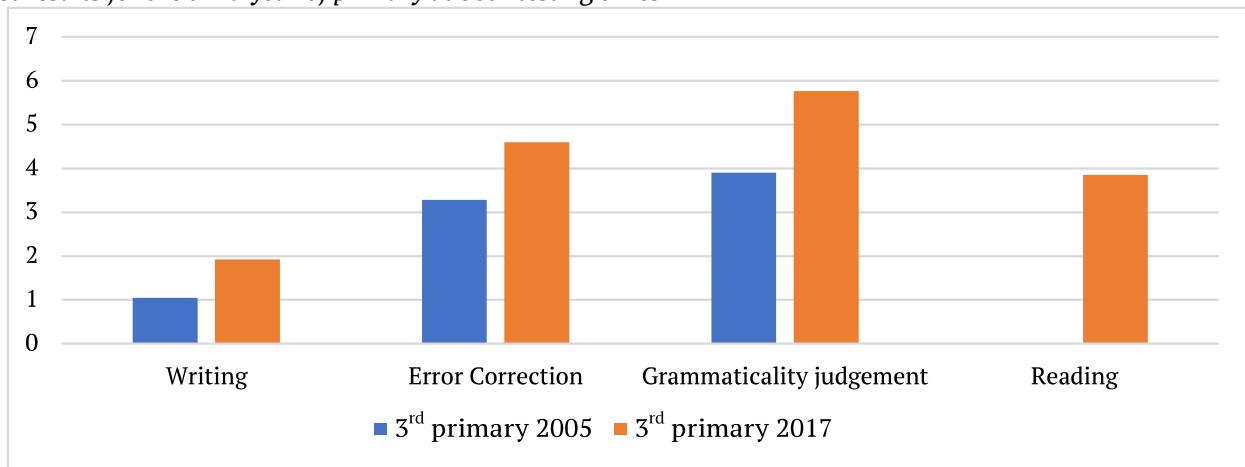
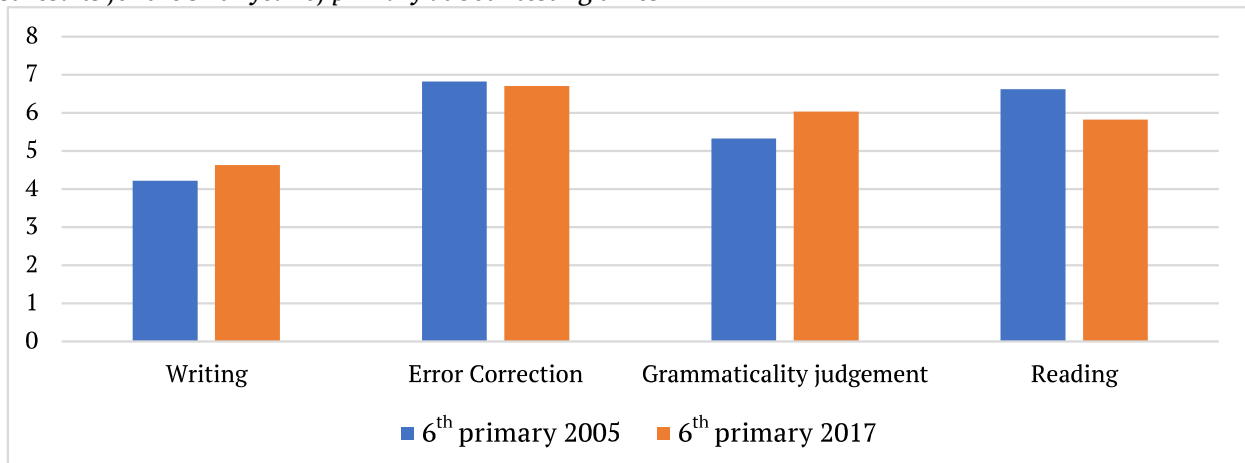


Table 3
Group statistics for the sixth graders in 2005 (Group 3) and 2017 (Group 4)

Group	N	Mean	Std. Deviation	Std. Error	P-value
Writing test	3	4.217	1.0722	.1958	.128
	4	4.633	1.0165	.1856	.128
Error correction test	3	6.823	1.4689	.2682	.729
	4	6.700	1.2630	.2306	.729
Grammaticality judgement test	3	5.333	1.0450	.1908	.026
	4	6.033	1.3126	.2397	.026
Reading test	3	6.617	1.7003	.3104	.106
	4	5.817	2.0615	.3764	.107

Figure 2
Test results for the sixth year of primary at both testing times



When analysing the results of participants in the second grade of secondary, as displayed in Table 4 and Figure 3 below, higher grades were reported in all four tests in 2017, although in the case of the error correction test, the difference was very small: the mean in 2005 was 8.17, while in 2017 it was of 8.19, thus being slightly higher in 2017, however not statistically significant ($p=0.909$). On the writing test, the mean score in 2005 was 5.62; while in 2017 it was 6.50, a difference that was statistically significant ($p=0.032$). Concerning the grammaticality judgement test, the mean reported in 2005 was 7.48; while in 2017 it was 7.75, the difference not being

statistically significant ($p=0.353$). Finally, for the reading test, the mean reported in 2005 was 7.72; while in 2017 it was 8.53, with a statistically significant difference ($p=0.013$).

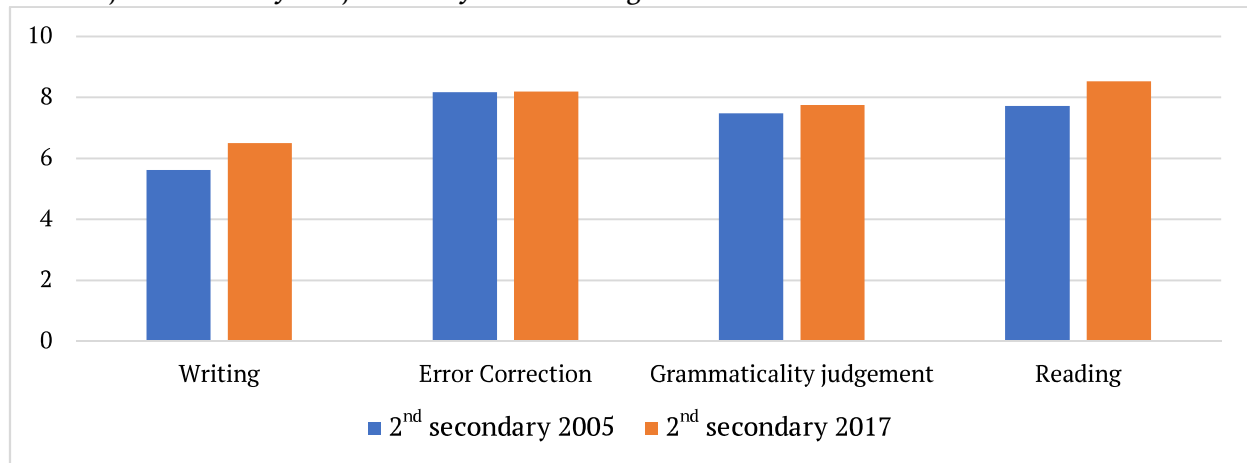
Table 4

Group statistics for the second year of secondary in 2005 (Group 5) and 2017 (Group 6)

Group	N	Mean	Std. Deviation	Std. Error	P-value
Writing test	5	5.617	1.5125	.2761	.032
	6	6.500	1.5920	.2907	.032
Error correction test	5	8.170	.7479	.1366	.909
	6	8.193	.8166	.1491	.909
Grammaticality judgement test	5	7.483	1.2490	.2280	.353
	6	7.750	.9354	.1708	.353
Reading test	5	7.717	1.5011	.2741	.013
	6	8.533	.8996	.1642	.014

Figure 3

Test results for the second year of secondary at both testing times



In sum, the previous results prove that the first hypothesis can only be partially confirmed. The 2017 cohort showed higher results than the 2005 cohort on most tests. Two exceptions must be noted: the sixth graders in primary school whose error correction test and reading test in 2005 yielded higher results; and the second graders in secondary whose error correction test scores were almost the same at both data collection times.

The second research question enquired into the gains shown cross-sectionally between the third and sixth graders in primary; and those same sixth graders and second graders in secondary. It was hypothesized that gains in the 2017 cohort would be higher than those reported for the 2005 cohort.

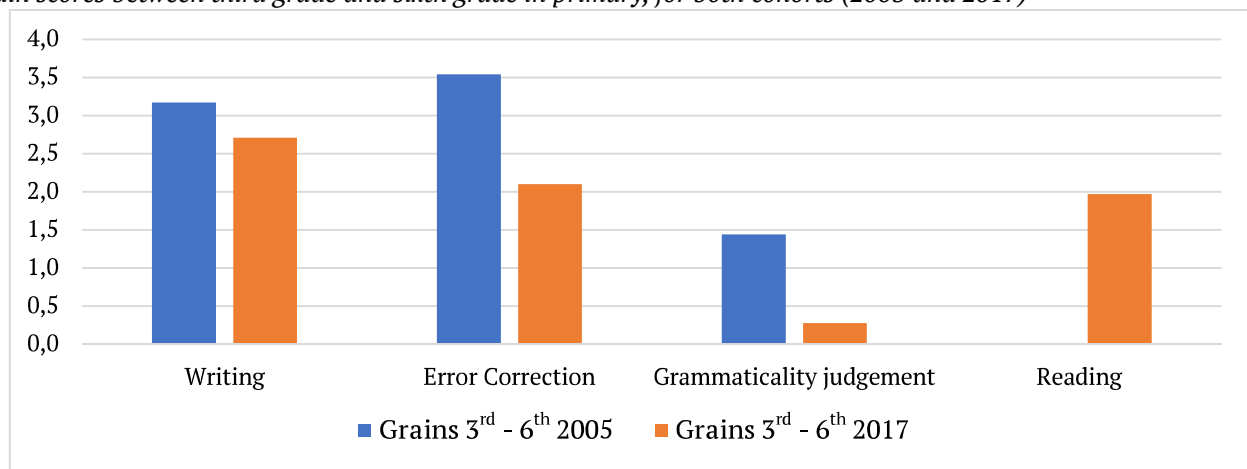
When comparing third and sixth graders in primary, significantly better results were obtained by the 2005 cohort in the case of the error correction test and the grammaticality judgement test, and a statistical tendency showed in writing⁵. Indeed, the results for the writing test, as displayed in Table 5 and Figure 4 below, reveal that the gains obtained in 2005 were 3.17 out of 10; while in 2017, they were 2.71, but the difference was not statistically significant ($p=0.208$). On the error correction test, the gains in 2005 were 3.54 points; while they were 2.10 in 2017, such a difference being statistically significant ($p=0.014$). On the grammaticality judgement test, the gains in 2005 were 1.44 points; while in 2017, they were 0.28, with a statistically significant difference ($p=0.027$).

⁵ As mentioned above, no reading test was administered to the third graders in 2005, for technical reasons, so no comparison is possible in this respect.

Table 5
Mean gains in all tests between the third graders and the sixth graders in primary

Variables	Group	Gains	P-value
Writing test	2005	+3.17	.208
	2017	+2.71	
Error correction test	2005	+3.54	.014
	2017	+2.10	
Grammaticality judgement test	2005	+1.44	.027
	2017	+0.28	
Reading test	2005		
	2017	+1.97	

Figure 4
Gain scores between third grade and sixth grade in primary, for both cohorts (2005 and 2017)



We find the reverse pattern when contrasting those same sixth graders and second graders in secondary. Table 6 and Figure 5 below show that between sixth grade in primary and second grade in secondary, a tendency towards greater gains for the 2017 testing time accrued, reaching significance only in the case of the reading test. On the writing test, the gains were 1.40 points in 2005; while they were 1.87 in 2017, a difference that was not statistically significant ($p=0.301$). On the error correction test, the gains in 2005 were 1.35; while they were 1.49 in 2017, such a difference not being statistically significant ($p=0.712$). On the grammaticality judgement test, the gains were 2.14 in 2005; while in 2017 they were 1.71, without statistical significance ($p=0.277$). Finally, on the reading test, the gains in 2005 were 1.09; while they were of 2.71 in 2017, a difference that was statistically significant ($p=0.005$).

Table 6
Mean gains in all tests between the sixth graders in primary and the second graders in secondary

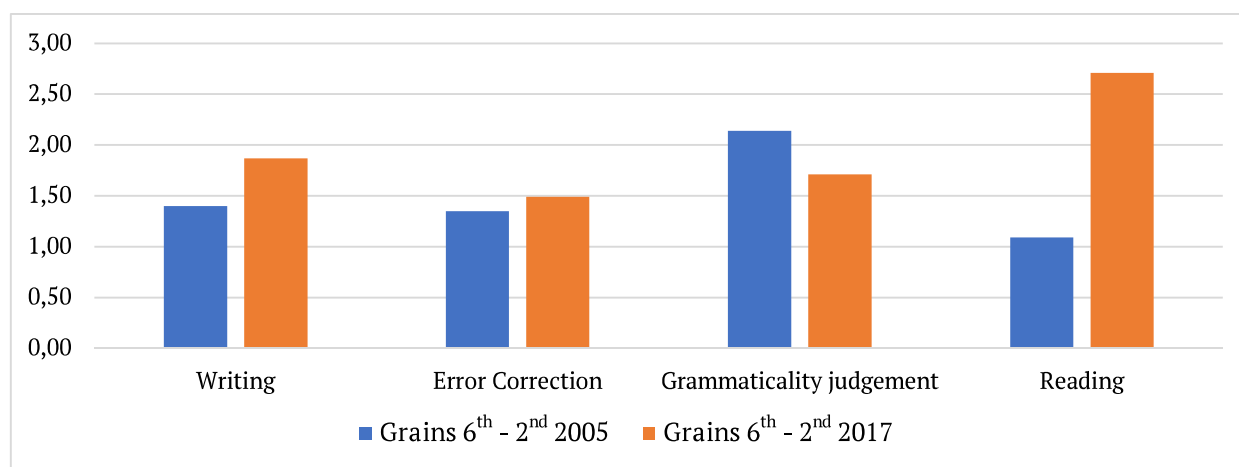
Variables	Group	Gains	P-value
Writing test	2005	+1.40	.301
	2017	+1.87	
Error correction test	2005	+1.35	.712
	2017	+1.49	
Grammaticality judgement test	2005	+2.14	.277
	2017	+1.71	
Reading test	2005	+1.09	.005
	2017	+2.71	

By examining the previous results, it can be stated that the second hypothesis cannot be confirmed. Significantly larger gains were obtained from the third to sixth grade of primary by the 2005 cohort on the

error correction and the grammaticality judgement tests, and a statistical tendency towards significance was shown in writing (reading was not measured). In contrast, the pattern seemed to change when we compared the sixth graders in primary and the second graders in secondary: significantly larger gains were found for the 2017 cohort in the case of reading and a tendency in the case of the writing, the error correction. Nevertheless, it has to be noted that in 2017, the sixth graders in primary showed lower results on the error correction and reading tests than those reported in 2005 (see Table 3). This can explain why, in the case of reading, they showed higher gains in this comparison, they had more room for improvement given their lower initial level.

Figure 5

Gain scores between the sixth graders in primary and the second graders in secondary for both cohorts (2005 and 2017)



The third and last research question enquired into the possible gains between the initial testing time (third grade in primary) and the last testing time (second grade in secondary). It was hypothesized that the 2017 cohort would show larger gains than the 2005 cohort.

Table 7 and Figure 6 below display the results, which do not confirm our hypothesis. They were similar at both testing times for writing: in 2005 gains reached 4.57; while in 2017, they were 4.58, and the difference was not statistically significant ($p=0.972$). In the case of the error correction test, higher gains were obtained in 2005, 4.89 points; than in 2017, when they were 3.59, such difference being statistically significant ($p=0.003$). The same was true in the case of the grammaticality judgement test, for which the gains obtained in 2005 reached 3.58, also a superior figure than in 2017, 1.99, a difference that was also statistically significant ($p=0.006$). As for the reading test, in 2005, third graders did not take this test.

Table 7

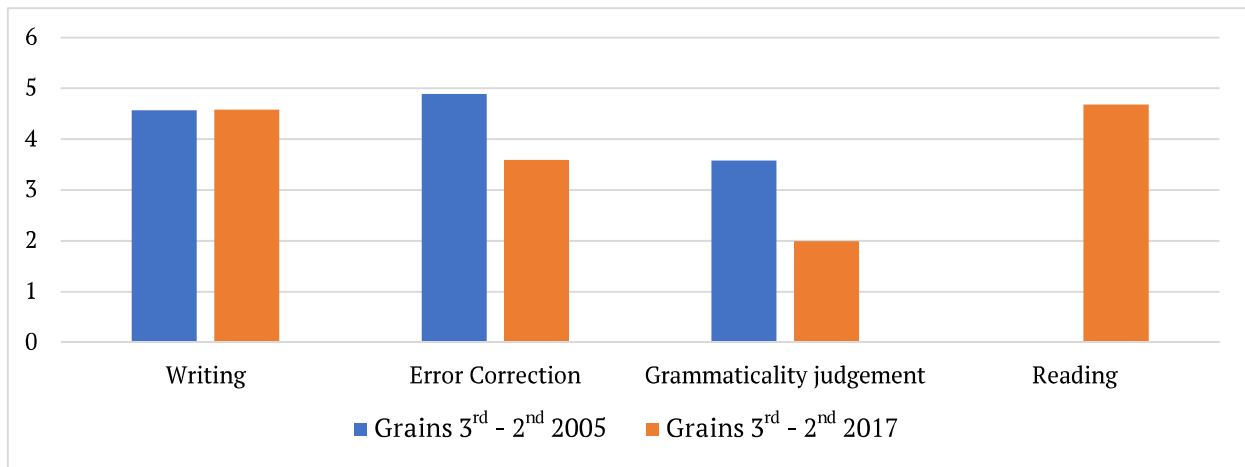
Mean gains in all tests between the third graders in primary and the second graders in secondary

Variables	Group	Gains	P-value
Writing test	2005	+4.57	.972
	2017	+4.58	
Error correction test	2005	+4.89	.003
	2017	+3.59	
Grammaticality judgement test	2005	+3.58	.006
	2017	+1.99	
Reading test	2005		
	2017	+4.68	

In conclusion, the third hypothesis cannot be confirmed either, since for both the error correction test and the grammaticality judgment test, significantly larger gains were obtained in 2005 than in 2017, with a statistical tendency shown in the case of the writing.

Figure 6

Gain scores between initial testing time and the last testing time for both cohorts (2005 and 2017)



Discussion

The current study has aimed to contrast the benefits of a FI+CLIL EFL programme at two different points in time. It sought to examine whether learners' linguistic proficiency proved to be higher at the time when the programme was no longer at its onset, that is twelve years after it was launched. It was assumed that more than a decade should have resulted in a well-established learning context no longer at the pilot stage, and, as a result, was potentially more efficient.

Most studies analysing the benefits of CLIL have been conducted with programmes that are in their pilot stages (Moore, 2009). It is therefore of utmost interest to analyse whether the infancy of the programme may negatively impact its outcomes by comparing the same programme years after its launching. A well-established and robust programme may be expected to provide better learning conditions by guaranteeing greater stability and complexity, leading to higher results in the students' linguistic proficiency (Moore, 2009; Navarro-Pablo & López Gándara, 2019).

Therefore, the first research question in the present study enquired into the EFL linguistic benefits of the programme by comparing learner proficiency at three different learner ages with a cross-sectional population. It was hypothesized that learners enrolled in the 2017 programme would show higher proficiency than their peers in the 2005 programme on the four tests administered, tapping into writing, grammar and vocabulary, and reading. This first hypothesis has been only partially confirmed, since participants in the 2017 cohort showed significantly higher grades on 8 out of 11 tests: all tests in the third grade in primary; writing and the grammaticality judgement test in the sixth grade of primary; and writing, the grammaticality judgement test, and reading in the second grade in secondary. On the contrary, on three occasions higher results were found in 2005 than in 2017: primary sixth graders on the error correction and reading tests, and secondary second graders on the error correction test.

Previous CLIL research has also reported contradictory results regarding the linguistic benefits of the programmes scrutinized with respect to syntax and writing (Dalton-Puffer, 2008; Pérez-Vidal, 2011). In the present study, the 2017 cohort yielded higher results in writing than the 2005 for the three age groups tested with a statistically significant difference for two of them. Such results may be due to the longevity and robustness of the programme, which may allow students to benefit from a more complex programme and more experienced teachers. As for syntax, out of the six grammar tests that were administered in total (error correction and grammaticality judgement in all three grades), higher results were reported in the 2017 cohort in four, with the exception of the error correction tests for sixth graders in primary and second graders in secondary. Finally, vocabulary has been found to be a skill benefitting from CLIL (Dalton-Puffer, 2008; Pérez-Vidal & Roquet, 2015). As for reading, higher results were found in the present study in older students (2nd secondary), while younger learners (6th primary) had higher results in the 2005 programme.

Against the previous backdrop, it seems evident that although most of the previous research in the field has focused on comparing CLIL and FI learners' language development, most of the skills that were previously found to benefit from CLIL have also showed higher results in the well-established programme analysed in the current study. Therefore, such results may mean that a more robust and stable CLIL programme could potentially lead to greater benefits in said skills.

Thus, these findings allow us to state that when groups from the early and the well-established programmes are compared at the ages of 8-9, 11-12, 13-14 and the same conditions are kept in terms of number of hours of instruction and pedagogical approach, learners' progress was higher in 2017 than in 2005 in most of the tests administered (8 out of the 11 tests). Indeed, the programme seems to have had a higher positive impact on learners' linguistic progress once fully established, twelve years after it was launched for most tests when raw scores are analysed, with the exception of the error correction and reading tests mentioned above. These results are in line with a previous study (Navarro-Pablo & López Gándara, 2019), in which higher results in the CLIL group were attributed to the seniority of the programme and teachers' expertise.

The second and third research questions approached the comparison of the impact of the CLIL programme when it first was launched and twelve years later, by measuring the gains in linguistic abilities between the three different grades tested. When comparing learners' gains between the third and sixth grades in primary at both testing times (2005 and 2017), much higher gains were seen in the 2005 cohort as well as significant gains on the error correction and grammaticality judgement tests. Hence the programme's length of time in operation does not seem to have yielded extra benefits in terms of linguistic gains. The opposite results were obtained when analysing the gains between the sixth grade in primary and the second grade in secondary, namely higher gains in 2017 for all tests except for the grammaticality judgement test, with a significant difference in reading and a statistical tendency towards significance in writing. Considering these findings, it cannot be claimed that learners will reap higher benefits from a well-established CLIL programme in terms of linguistic gains between the beginning, middle, and end points of the programme. However, there is a tendency towards greater improvement in the older students, who showed higher gains in 2017 than in the areas of writing, grammar, and reading.

Finally, when comparing the gains between the younger and the older groups, almost equal gains were reported on the writing test and higher gains were found in 2005 on the error correction and grammaticality judgement tests. As a result, the second and third hypotheses cannot be confirmed.

To our knowledge, no previous studies on the influence of programme length have analysed linguistic gains; however, one study (Navarro-Pablo & López Gándara, 2019) focused on analysing the effects of programme seniority using raw scores in English proficiency. Therefore, further research is needed in this domain to further examine the effects of programme length.

Conclusion

Overall, it can be stated that the quality of the programme has indeed improved since learners in the 2017 cohort had better results on most of the tests at all three educational levels when raw results are compared. Such improvement in the programme can be attributed to the programme's seniority, since the programme has been kept within the same parameters in terms of the number of hours, the CLIL coordinator, and the pedagogical approach followed. After 15 years of running, the programme in 2017 was more robust and stable, the teachers were more experienced, and its design and application more complex. This allowed students to benefit from a much more developed programme and attain higher levels in English as a FL.

In contrast, the analysis of gains does not seem to prove an effect of the programme's seniority, neither when contrasting the development between the three educational levels, nor between the initial and final testing times. Thus, it cannot be confirmed that a fully-established CLIL programme will lead to higher results in terms of the linguistic gains in the students enrolled in the programme, as compared to the programme in its pilot phase. Additionally, the absence of higher gains in 2017 may, however, be attributed to a ceiling effect: initial scores in 2017 were slightly higher on most tests both in the third and sixth grades of primary school, which were used as pre-test scores to analyse gains; therefore, starting levels of English were higher

in 2017. Thus, higher initial language levels may have left less room for improvement and, as a result, explain the lower gains.

However, this study has some limitations that may partially explain its results. The first and possibly the largest limitation is that the data used was not longitudinal, but cross-sectional. Although students' exposure and contact with English inside and outside the school was controlled through the sociolinguistic background questionnaire at both testing times, the cross-sectional nature of the study poses a limitation, since the raw scores and gains compared were from different samples of participants. Second, individual differences are considered to have an impact on FL learning; however, we did not take them into account due to time and space constraints. Only the mother tongue of the subjects and their contact with the English language outside of school was considered to make sure that none of the learners had English as their native language or had more hours of contact with the language beyond the school sessions. In addition, it must be noted that the amount of extramural exposure to English in 2017 was much higher than in 2005 as access to English resources accessible from home has increased in the last decade. Therefore, we cannot be completely sure that some of the higher results reported in 2017 were not also enhanced due to the fact that children nowadays have more access to the English language.

This study has sought to contribute with new, meaningful, and updated data to our understanding of the effects of a CLIL programme in mainstream education. In a nutshell, our results reveal that the programme's quality has indeed changed over the years, guaranteeing a higher level of EFL proficiency when raw scores are contemplated, but not when gains are analysed. This calls for further studies being conducted that might add longitudinal data to the cross-sectional data used in this study.

Acknowledgments

This research was supported by grants PGC2018-098815-B-I00 and 2017 SGR 1028. Marta Segura is a Universitat Internacional de Catalunya PhD grant student at the Institute for Multilingualism. We would like to thank the editors for their insightful comments.

Conflict of interests

The authors declare that they have no conflict of interest.

References

- Admiraal, W., Westhoff, G., & de Bot, K. (2005). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English 1. *Educational Research and Evaluation* 12(1), 75-93. <https://doi.org/10.1080/13803610500392160>
- APAC. (Ed.). (2006). *CLIL in Catalonia: From theory to practice*. Associació de Professors d'Anglès de Catalunya.
- Auerbach, E. R. (2006). Reexamining English only in the ESL Classroom. *TESOL Quarterly*, 27(1), 9-32. <https://doi.org/10.2307/3586949>
- Brinton, D., Wesche, M., & Snow, A. (Eds.). (1989). *Content-based second language instruction*. Nebury House Publishers.
- Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, 39, 523 – 532. <https://doi.org/10.1016/j.system.2011.08.002>
- Cenoz, J. (2015). Content-based instruction and content and language integrated learning: The same or different? *Language, Culture and Curriculum*, 28(1), 8–24. <https://doi.org/10.1080/07908318.2014.1000922>
- Dafouz Milne, E., & Guerrini, M. C. (Eds.). (2009). *CLIL across educational levels: Experiences from primary, secondary and tertiary contexts*. Richmond Publishing – Santillana Educación S.L.
- Dalton-Puffer, C. (2008). Outcomes and processes in Content and Language Integrated Learning (CLIL): Current research from Europe. In W. Delanoy, & L.Volkman (Eds.), *Future perspectives for English language teaching* (pp. 139-157). Carl Winter.
- Darvin, R., Yi Lon, Y., & Lin, A. M. Y. (2020). Examining CLIL through a critical lens. *English Teaching and Learning*, 44(2), 103-108. <https://doi.org/10.1007/s42321-020-00062-2>

- DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(4), 499-533. <https://doi.org/10.1017/S0272263100004022>
- Derakhshan, A., & Karimi, E. (2015). The Interference of first language and second language acquisition. *Theory and Practice in Language Studies*, 5(10), 2112-2117 <http://dx.doi.org/10.17507/tpls.0510.19>
- De Dios Martínex Agudo, J. (2019). The impact of CLIL on English language competence in monolingual context: A longitudinal perspective. *The Language Learning Journal*, 48(1), 36-47. <https://doi.org/10.1080/09571736.2019.1610030>
- Fernández-Sanjurjo, J., Fernández-Costales, A., & Arias Blanco, J. M. (2017). Analysing students' content learning in science in CLIL vs. non-CLIL programmes: Empirical evidence from Spain. *International Journal of Bilingual Education and Bilingualism*, 22, 661-674. <https://doi.org/10.1080/13670050.2017.1294142>
- García, O. 2009. *Bilingual education in the 21st century: A global perspective*. Wiley-Blackwell.
- García, O., Aponte, G.Y., & Le, K., (2020). Primary bilingual classrooms: Translations and translanguaging. In S. Laviosa & M. González Davies, (Eds.). *The Routledge handbook of translation and education* (pp. 81-94). Taylor and Francis. <https://doi.org/10.4324/9780367854850>
- Gené Gil, M. (2016). *Efecto del AICLE sobre la producción escrita en estudiantes de secundaria de inglés como lengua extranjera* [CLIL effects on written production in secondary school students of English as a foreign language]. Universitat de les Illes Balears
- Hughes, S. P., & Madrid, D. (2019). The effects of CLIL on content knowledge in monolingual contexts. *The Language Learning Journal*, 48(1), 48-59. <https://doi.org/10.1080/09571736.2019.1671483>
- Hüttner, J., & Smit, U. (2014). CLIL (Content and Language Integrated Learning): The bigger picture. A response to: A. Burton. (2013) CLIL: Some of the reasons why... and why not. *System*, 41, 587-597. *System*, 44, 160-167. <https://doi.org/10.1016/j.system.2014.03.001>
- Jaén Campos, M. (2016). *Redefining the roles of the EFL teacher within CLIL secondary settings*. Universidad Complutense de Madrid.
- Kersten, K., & Rohde, A. (2013). Teaching English to young learners. In A. Flyman Mattsson & C. Norrby (Eds.), *Language acquisition and use in multilingual contexts* (pp. 107-118). Lund University.
- Lasagabaster, D. (2008). Foreign Language Competence in Content and Language Integrated Courses. *The Open Applied Linguistics Journal*, 1, 31-42. <https://doi.org/10.2174/1874913500801010030>
- Lasagabaster, D., & Doiz, A. (2017). A longitudinal study on the impact of CLIL on affective factors. *Applied Linguistics*, 38, 688-712. <https://doi.org/10.1093/applin/amv059>
- Lasagabaster, D., & Sierra, J. M. (2009). Language attitudes in CLIL and international EFL classes. *International CLIL Research Journal*, 2(2), 4-17.
- Llinares, A., & Morton, T. (2010). Historical explanations as situated practice in content and language integrated learning. *Classroom Discourse*, 1(1), 46-65. <https://doi.org/10.1080/19463011003750681>
- Lorenzo, F., Casal, S., & Moore, P. (2010). The effects of content and language integrated learning in European education: Key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics*, 31(3), 418-442. <https://doi.org/10.1093/applin/amp041>
- Moore, P. 2009. *On the emergence of L2 oracy in bilingual education: A comparative analysis of CLIL and mainstream learner talk* [Unpublished doctoral dissertation]. Universidad de Sevilla. <https://doi.org/10.1093/applin/amp041>
- Navarro-Pablo, M., & López Gándara, Y. (2019). The effects of CLIL on L1 competence development in monolingual contexts. *The Language Learning Journal*, 48(1), 18-35. <https://doi.org/10.1080/09571736.2019.1656764>
- Navés, T. (2009). Effective Content and Language Integrated (CLIL) programmes. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and Language Integrated Learning. Evidence from research in Europe* (pp. 22-40). Multilingual Matters.
- Navés, T., & Muñoz, C. (1999). CLIL experiences in Spain. In D. Marsh & G. Langé (Eds.), *Implementing Content and Language Integrated Learning*. Jyväskylä.
- Nikula, T., & Moore, P. (2015). Exploring translanguaging in CLIL. *International Journal of Bilingual Education and Bilingualism*, 22, 237-249. <https://doi.org/10.1080/13670050.2016.1254151>
- Papaja, K. (2013). The role of a teacher in a CLIL classroom. *Glottodidactica An International Journal of Applied Linguistics*, 40(1) 147-154. <https://doi.org/10.14746/gl.2013.40.1.11>
- Pérez-Vidal, C. (2007). The need for focus on form (FoF) in content and language integrated approaches: An exploratory study. *Revista Española de Lingüística Aplicada*, 1, 39-54.
- Pérez-Vidal, C. (2011). Language acquisition in three different contexts of learning: Formal instruction, study abroad and semi-immersion (CLIL). In Y. Ruiz de Zarobe, J.M. Sierra, & F. Gallardo del Puerto (Eds.), *Content*

- and foreign language integrated learning: Contributions to multilingualism in European contexts* (pp. 25-35). Peter Lang. <https://doi.org/10.3726/978-3-0351-0171-3>
- Pérez-Vidal, C. (2013). Perspectives and lessons from the challenge of CLIL experiences. In C. Abello-Contesse, P. M. Chandler, M. D. López-Jiménez, & R. Chacón-Beltrán (Eds.), *Bilingual and multilingual education in the 21st century building on experience* (pp. 59-82). Multilingual Matters. <https://doi.org/10.21832/9781783090716>
- Pérez-Vidal, C., Escobar, C., & Roquet, H. (2013). Perspectives sobre l'aprenentatge de llengües i multilingüisme [Perspectives on language learning and multilingualism]. *Butlletí Informatiu del Col·legi de Doctors i Llicenciats en Filosofia i Lletres i en Ciències de Catalunya*, 14, 4-5.
- Pérez-Vidal, C., & Juan-Garau, M. (2012). To CLIL or not to CLIL? From bilingualism to multilingualism in Catalan/Spanish communities in Spain. In Y. Ruiz de Zarobe, & D. Lasagabaster (Eds.), *CLIL in Spain: Implementation, results and teacher training* (pp. 115–138). Cambridge Scholars Publishing.
- Pérez-Vidal, C., Lorenzo, N., & Trenchs, M. (2015). Les llengües en l'educació: El plurilingüisme i la internacionalització [Languages in education: plurilingualism and internationalization]. In J. M. Vilalta (Ed.), *Reptes de l'educació a Catalunya. Anuari 2015* [Education challenges in Catalonia. Annual report 2015] (pp. 139-197). Fundació Jaume Bofill.
- Pérez-Vidal, C., & Roquet, H. (2015). The linguistic impact of a CLIL Science programme: An analysis measuring relative gains. *System*, 54, 80-89. <https://doi.org/10.1016/j.system.2015.05.004>
- Pladevall-Ballester, E., & Vallbona, A. (2016). CLIL in minimal input contexts: A longitudinal study of primary school learners' receptive skills. *System*, 58, 37-48. <https://doi.org/10.1016/j.system.2016.02.009>
- 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 (Edition: Educational Linguistics)* (pp. 163–177). Springer. https://doi.org/10.1007/978-3-319-11496-5_10
- Rallo Fabra, L., & Juan-Garau, M. (2011). Assessing FL pronunciation in a semi-immersion setting: The effects of CLIL instruction on Spanish-Catalan learners' perceived comprehensibility and accentedness. *Poznan Studies in Contemporary Linguistics*, 49(1), 96–108. <https://doi.org/10.2478/psicl-2011-0008>
- Roquet, H. (2011). *A study of the acquisition of English as a foreign language integrating content and language in mainstream education in Barcelona*. Universitat Pompeu Fabra.
- Roquet, H., & Pérez-Vidal, C. (2017). Do productive skills improve in content and language integrated learning contexts? The case of writing. *Applied Linguistics*, 15(3), 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), 60-73
- San Isidro, X., & Lasagabaster, D. (2019). The impact of CLIL on pluriliteracy development and content learning in a rural multilingual setting: A longitudinal study. *Language Teaching Research*, 23(5), 584-602. <https://doi.org/10.1177/1362168817754103>
- Van de Craen, P., & Pérez-Vidal, C. (2001). *The multilingual challenge, Le Défi Multilingue* [The multilingual challenge]. Printulibro Intergrup, S.A.