The impact of age and exposure on EFL achievement in two learning contexts: formal instruction and formal instruction + content and language integrated learning (CLIL)

Gemma Artieda , Helena Roquet & Florentina Nicolás-Conesa

INTERNATIONAL JOURNAL OF BILINGUAL EDUCATION AND BILINGUALISM, 2017

https://doi.org/10.1080/13670050.2017.1373059

The impact of age and exposure on EFL achievement in two learning contexts: formal instruction and formal instruction + content and language integrated learning (CLIL)

Gemma Artieda a, Helena Roquet b and Florentina Nicolás-Conesac

aDepartment of English Studies, Universitat de Barcelona, Barcelona, Spain; bInstitute for Multilingualism, Universitat Internacional de Catalunya, Barcelona, Spain; cDepartment of Languages, Centro Universitario de la Defensa, Murcia, Spain

ARTICLE HISTORY

Received 8 March 2017

Accepted 24 August 2017

KEYWORDS

CLIL; formal instruction; EFL learners; age; exposure

Introduction

Content and Language Integrated Learning (CLIL) is an umbrella term used to describe programmes in which a foreign language is used as the medium of instruction to teach content subjects in mainstream education. CLIL promotes a dual-focused form of instruction where attention is given to both language and content (Coyle, Hood, and Marsh 2010), and in which the non-language subject is not taught in a foreign language but with and through a foreign language (Eurydice 2006, 8). CLIL programmes cover a wide range of educational approaches with differing degree of intensity and exposure to the foreign language; from high-intensity (long-term programmes such as immersion and bilingual education) to medium and low-intensity programmes such as one-CLIL subject courses or language showers (Mehisto, Marsh, and Frigols 2008). These programmes have enjoyed increasing popularity in Europe over the past 20 years, since they were prioritised as a major educational initiative in Europe from 1990 onwards (Eurydice 2006). Their objective is to provide citizens with the ability to be functional in more than one language (European Commission 1995), thus fostering the development of a multilingual society across Europe (Eurydice 2006).

The implementation of CLIL in Europe is heterogeneous due to variables such as compulsory status, intensity, age of onset, starting level or duration (Coyle 2007). The most common CLIL

CONTACT Helena Roquet hroquet@uic.es

model in Europe combines foreign languages with regional and / or minority languages (Pérez-Cañado 2012). In formal instruction (FI) within foreign language (FL) learning contexts, in which the provision of input is scarce, an additional exposure to the target language is one of the main advantages of CLIL for the development of the L2 (Muñoz 2002, 2007). Students in CLIL strands tend to perform better than learners in non-CLIL strands in receptive skills, vocabulary and morphology (Dalton-Puffer 2008), as well as in general fluency (Pérez-Vidal 2011). Practitioners and researchers have formulated an array of additional goals for CLIL. For example, CLIL has been found to increase students' motivation towards learning (Lasagabaster and Sierra 2009), to develop cognitive skills (Novotná and Hofmannová 2000), and to enhance self-confidence, creativity, and risk-taking (Dalton-Puffer 2008), as well as self-regulatory abilities (Pérez-Vidal 2011). In addition, CLIL helps to promote linguistic diversity, language learning, and internationalisation(Van de Craen, Mondt, Allain and Gao 2007), and to boost multilingualism in schools (Grisaleña, Alonso, and Campo 2009; Merino and Lasagabaster 2015).

CLIL research has typically focused on L2 learning outcomes; in general, CLIL learners tend to outstrip non-CLIL students (Pérez-Cañado 2012). Regarding L2 skills that are expected to improve in CLIL settings, some studies report advantages for CLIL students in some skills, whereas others do not find such an advantage for the same skills. For example, there seem to be no clear-cut findings regarding listening, which improves in some studies (Lasagabaster 2008) but not in others (Navés 2011). In studies that look at specific skills, CLIL students have been reported to surpass non-CLIL learners in reading (Admiraal, Westhoff, and de Bot 2006), writing (Järvinen 2010; Lorenzo, Casal, and Moore 2010) and speaking (Admiraal, Westhoff, and de Bot 2006; Lasagabaster 2011). There are also studies in which CLIL students outperform their non-CLIL counterparts in every L2 skill (Lasagabaster 2008; San Isidro 2010). As for other L2 dimensions, advantages have been found for CLIL groups in specific grammar areas, such as morphosyntax (Lázaro Ibarrola 2012), and tense and agreement morphology (Villarreal and García-Mayo 2009). Other studies have found improvements in some syntactic measures and not others (Martínez Adrián and Gutiérrez Mangado 2009).

These findings indicate that there is a need for further research in grammar development and L2 skills in CLIL, with a special urge for researching reading and listening (Merino and Lasagabaster 2015). Additionally, methodological improvements are warranted in CLIL research, as we argue below.

Some studies yielding satisfactory results for CLIL groups have been questioned based on methodological flaws, such as lacking pre-post average scores, not having valid comparison groups, lacking comparable learning contexts (Bruton 2011), or not using moderating variables, not involving longitudinal designs or not using mixed-methods designs (Pérez-Cañado 2016a). Besides, some CLIL programmes in Europe have been accused of being non-

egalitarian in terms of student selection, resulting in the learners with higher ability, motivation and socio-economic status (SES) being placed in CLIL strands (Ruiz de Zarobe 2008; Seikkula-Leino 2007). This is to the disadvantage of non-CLI students, for whom FL learning might become an impoverished alternative (Bruton 2011, 2013). However, as Pérez-Cañado (2016a) states, methodological shortcomings are not proof that CLIL is not successful, but rather a call to researchers to overcome such deficiencies and to improve empirical research, as we are still in much need of 'valid and reliable research' in CLIL (Pérez-Cañado 2016a). For this purpose, Pérez-Cañado (2016b, 20) proposed constructing 'a research agenda which supersedes the methodological shortcomings which have compromised the validity of CLIL investigations […]' This agenda could start by investigating individual differences (ID), as some

*CLIL instruction started in Grade 5 (10/11 years) only for Group A when the school implemented the CLIL programme. Group A and B started FI in Pre-school (5/6 years)

Table 2. Hours of exposure at each data collection time.

Data collection Group A (FI + CLIL) (n = 50) Group B (FI) (n = 50)

Time 1 Grade 7 (12/13 years)

FI: 1120 h + CLIL:210 h = 1330h

Time 2 Grade 8 (13/14 years)

FI: 1260 h + CLIL: 280h = 1540h

Grade 8 (13/14 years) FI: 1260h

Grade 9 (14/15 years) FI: 1400h

researchers argue that part of the inconsistent evidence for the beneficial effects of CLIL in Europe may be due to fundamental differences in age and amount of instruction (Muñoz 2015). Certainly, there is a need for further studies in age by treatment interaction research to shed light on the L2 learning processes that are at work at different ages and under different amounts of exposure (DeKeyser 2012; Robinson 2007). The role of age has been thoroughly investigated in naturalistic and formal second language acquisition (SLA) (Birdsong 2006, 2014; DeKeyser 2000, 2012; Muñoz 2006, 2008), but this has not been the case in CLIL contexts (Muñoz 2015). In the latter, only a few studies have investigated the benefits of CLIL compared to traditional FI at different ages keeping the hours of instruction constant (Schoonjans 2013; Villarreal 2011).

In EFL contexts the provision of L2 input is scarce, and this makes it difficult for learners to attain advanced proficiency levels unless they can enjoy immersion learning experiences (Rifkin 2005). It has been argued that in such input-limited settings young learners do not have the amount of input necessary to benefit from implicit learning mechanisms (DeKeyser 2000; Muñoz 2006), namely, the ability to learn without awareness of what is being learnt (DeKeyser 2013). CLIL has been considered an effective way to increase learners' exposure to the target language in foreign language learning contexts. However, more studies are needed to find out the amount of additional L2 exposure through CLIL that can lead to linguistic gains (Housen 2012; Muñoz 2015) to help CLIL learners benefit from such programmes compared to traditional FI.

Muñoz (2015) proposed in a review of empirical studies on CLIL that when learners are matched for hours of exposure, older students in non-CLIL strands seem to do better than younger students in CLIL programmes when the age difference is equal or greater than two years, due to the greater cognitive maturity of older learners (hypothesis one). Conversely, when learners are matched for age, there seems to be an advantage for learners in CLIL strands when the difference in hours of cumulative exposure is greater than 300 h in the CLIL groups (hypothesis two).

The present study aims at empirically testing Muñoz's (2015) two hypotheses to help filling some important gaps in CLIL research. With this aim in mind, the study investigates the effects of two independent variables: students' age in CLIL programmes compared to FI as well as amount of additional L2 exposure through CLIL in two groups of secondary school learners' grammar knowledge and L2 skills (receptive and productive). Our participants were divided into two groups of students that had started learning English at the same early age (5 years) in a foreign language setting. They were tested under two different conditions: (i) when they had received similar hours of instruction1 in two

Table 3. Between-group comparisons.

Comparison between groups

Conditions

Group A (FI + CLIL)

Time 1

Group B (FI)

Time 2

Group A (FI + CLIL)

Time 2

Group B (FI)

Time 1

Condition 1 Different age (2 years) 12/13 years

(Grade 7)

14/15 years (Grade 9)

Similar hours (70 h of difference)

Condition 2 Same age

1330 h 1400 h

13/14 years (Grade 8)

13/14 years (Grade 8)

Different hour (270 h of difference)

1540 h 1260 h

Table 4. Comparison of results in L2 skills (receptive and productive) and grammar knowledge of participants who differed in age (2 years) but had similar hours of exposure to English under different instructional treatments.

Skills/Knowledge	Instruments/Measures	Group A (FI + CLIL) (12/13 years; 1330 h) (n = 50) M(SD)	Mdn	Group B (FI) (14/15 years; 1400 h) (n = 50) M(SD)	Mdn	Mann Whitney U test U	Z	p	Effect size (r)*

*r = z/√N; M = mean; SD = standard deviation; Mdn = median.

Table 5. Comparison of results in L2 skills (receptive and productive) and grammar knowledge of participants who were the same age but differed in hours of exposure to English (280 h) under different instructional treatments.

Skills/Knowledge	Instruments/Measures	Group A (FI + CLIL) (13/14 years; 1540 h) (n = 50) M(SD)	Mdn	Group B (FI) (13/14 years; 1260 h) (n = 50) M(SD)	Mdn	Mann Whitney U test U	Z	p	Effect size (r)*

*r = z/√N; M = mean; SD = standard deviation; Mdn = median.

different instructional treatments (FI vs FI + CLIL), and they differed in the age of testing (two years apart, 12 versus 14 years old respectively); (ii) when the same participants in these two different instructional treatments (FI vs FI + CLIL) were the same age when they were tested (13 to 14 years old), but had different hours of instruction. The proposed research design has high ecological validity as it investigates the most frequent CLIL implementation model in Spain: FI vs FI + CLIL, where the FI

+ CLIL group is a group that receives the same amount of FI than the control group as well as an additional number of CLIL hours. A value-add is that data were collected while CLIL was being implemented in the school, which allowed us to have a pure FI control group to compare our FI + CLIL group with in the same school, which enhanced the comparability of the groups.

We now review the main empirical findings of current research about the two independent variables of our study in CLIL, age and amount of exposure, following Muñoz's (2015) two hypotheses.

Similar hours of exposure in different instructional treatments (FI versus CLIL) and different biological age

In naturalistic settings, children tend to surpass older learners in ultimate attainment in the long term (DeKeyser 2000; Krashen, Long, and Scarcella 1979), albeit there is no evidence that children learn languages faster than adults (DeKeyser 2013).

In contrast, in foreign language (FL) learning contexts, in which exposure to the target language is scarce, older learners have been found to be more efficient learners than younger students, reaching higher levels of achievement and making faster progress (Llanes and Muñoz 2013; Muñoz 2008) due to their superior cognitive development and analytical skills (Muñoz 2006).

Regarding CLIL instruction in foreign language settings, Muñoz (2015) highlighted that only a few studies in Europe have compared learners with different ages at testing and similar hours of instruction, since CLIL hours are usually offered in addition to the FI hours in mainstream education (henceforth FI + CLIL). Villarreal (2011) compared the oral production of verbal morphology in 134 Basque-Spanish CLIL and non-CLIL high-school students at three times. When she compared CLIL and non-CLIL students after matching them for hours of exposure, there was no advantage for CLIL students, since non-CLIL students, who were 2–3 years older than CLIL learners (17–18 years old), only showed a slight advantage. Villarreal (2011) attributed this result to the superior cognitive maturity of non-CLIL students. Along the same lines, Schoonjans (2013) investigated German native speakers who were studying English in three European Schools2 (ES) in Germany, Brussels, and Britain, and compared them to EFL learners in a mainstream school in Germany after 580 h of instruction in both contexts. Results in a general English proficiency test, and in some measures of accuracy and lexical complexity were consistently favourable to the

mainstream school learners in Germany, who were older when they were tested (13 in mainstream vs 10 in ES schools), and had also had a later age of onset (8 in mainstream vs. 6 in ES schools). Only the fluency scores of the mainstream group were lower than the scores obtained by learners in ES schools of Britain and Brussels. Schoonjans (2013) interpreted that the mainstream group benefited from the older age advantage in all L2 proficiency areas except in fluency, a domain in which the hours of exposure at school may not have been enough to compensate for the lack of L2 use outside of school in Germany.

The findings of these studies included in Muñoz's review seem to show that, the older non-CLIL

learners are, the more L2 language benefits can be found in verbal morphology, accuracy and lexical complexity compared to CLIL learners when the age difference is +2 years. This would tally with the higher levels of L2 achievement for older learners proposed by Muñoz's (2015) first hypothesis, although more research is needed to investigate in what other language dimensions (e.g. grammar knowledge) and L2 skills we can expect gains if the amount of L2 exposure is similar in CLIL and non-CLIL programmes. There is not enough research on the influence of an older age on the development of receptive and productive L2 skills when comparing learners in CLIL and non-CLIL programmes in EFL settings, although receptive skills seem to be ahead of productive ones (Muñoz 2015).

On the other hand, when participants are the same age but differ in the number of hours of exposure in distinct instructional programmes, the findings seem to be different, as we explain in the following section.

Same biological age, different number of hours of exposure and treatment (FI versus CLIL)

Following Muñoz's (2015) review, several studies have provided evidence of positive gains for CLIL groups with more than 300 h of CLIL instruction compared to their non-CLIL counterparts. For example, Lázaro Ibarrola (2012), in a study involving Basque-Spanish bilingual learners of English aged 13 and 15, reported that the CLIL group outperformed the non-CLIL group in morphosyntactic measures at time 2, when the CLIL group had accumulated 480 additional hours of exposure. Likewise, Villarreal and García-Mayo (2009) also reported in another study with Basque-Spanish learners aged 15 and 16 that CLIL students, who had between 328 and 363 h of additional exposure, surpassed non-CLIL groups in tense and agreement morphology. On the other hand, Martínez Adrián and Gutiérrez Mangado (2009) conducted a study with learners aged 14, in which the CLIL group, who had 363 h of additional CLIL exposure, performed better in some syntactic measures (e.g. production of embedded clauses) but not in others (e.g. production of null subjects).

According to Muñoz (2015), other studies have reported no statistically significant differences between CLIL and non-CLIL groups when L2 exposure in CLIL does not exceed 300 h. For instance, Ruiz de Zarobe (2007) did not find significant differences between a CLIL and a non-CLIL group (aged 15–16) in pronunciation, vocabulary, grammar, fluency, or content, even though the CLIL group had had 210 additional hours. Similarly, Gallardo, García Lecumberri, and Gómez Lacabex (2009) investigated FL pronunciation in CLIL and non-CLIL learners and found no significant differences between the two groups after 259 additional hours of instruction in the CLIL group.

The findings of previous studies seem to be in line with the existence of a threshold regarding the amount of L2 exposure that could lead to L2 language benefits in some morphosyntactic measures for CLIL learners. However, further research is needed to empirically test Muñoz's (2015) second hypothesis regarding the influence of additional hours of L2 exposure on grammar knowledge, and on receptive and productive L2 skills when CLIL and non-CLIL learners are compared at the same biological age.

Research questions

The present study focuses on the influence of age and hours of exposure on the L2 competence achieved by young Spanish-Catalan learners in grammar knowledge as well as in receptive and productive L2 skills comparing a FI context to a FI + CLIL context. The CLIL instruction that our participants received is partial given that it consisted of only one subject (Science) being taught in English as the medium of instruction, the most common CLIL implementation model in primary and secondary schools in Spain nowadays. CLIL students had two additional hours of exposure to English per week compared to the learners in the FI only group.

The study addresses the following research questions:

At different ages and after similar hours of L2 exposure in different instructional treatments, who will show greater linguistic benefits in L2 skills (receptive and productive), and grammar knowledge, younger students receiving FI + CLIL instruction or older learners receiving only FI instruction? Based on the literature review, we hypothesise that older learners will show greater L2 benefits than young learners as they are two years older than their younger counterparts.

At the same age and after different number of hours of L2 exposure in different instructional treatments, who will show greater linguistic benefits in L2 skills (receptive and productive) and grammar knowledge, students receiving FI + CLIL instruction or learners receiving only FI? According to the literature reviewed, we hypothesise that students with more hours of CLIL

instruction will show greater linguistic benefits than their non-CLIL counterparts because of 280 additional hours of exposure being very close to the 300-hour threshold.

Materials and methods

This is a follow-up study of a larger piece of research (Roquet 2011) that explored the impact of two individual differences, gender and age, on the L2 skills and grammar knowledge of two groups of participants in two different instructional treatments (FI versus FI + CLIL), as well as the context effects on the participants' L2 gains longitudinally. The larger study and the present study use the same participant sample and the same instruments of data collection.

Participants and context

Participants in this study were 100 bilingual speakers of Catalan and Spanish, who were learning English in a semi-private secondary school in Barcelona, Spain. At the first time of data collection learners were studying either Grade 7 (12/13 years, Group A) or Grade 8 (13/14 years, Group B). There were 50% of males and 50% of females in each group.

Group A (n = 50) included learners who were studying English as a foreign language as part of the mainstream curriculum, and who in addition to the regular FI hours also received instruction in CLIL (FI + CLIL). Students in Group B (n = 50) only had the standard EFL hours established in the official curriculum (FI). This division of participants into two equal groups (n = 50) was possible by testing all learners who were finishing Grade 7 and Grade 8 once at Time 1 (T1), and then again when the same students were finishing Grade 8 and Grade 9 respectively at Time 2 (T2). There were two classes per grade, and all students were tested, so four intact classes participated in the study. There was no selection of participants. All students who completed all tests at T1 and T2 were included. Both groups shared the same age of onset and intensity of exposure to English as a third language (L3) from pre-school: five years old, four hours of FI per week.

The school started a CLIL programme in 2002–2003. Given that at the first time of data collection only the participants who were in Grade 7 had been forced to enter the compulsory CLIL programme at school back when they were in Grade 5, it was possible to divide learners into two groups: students who received FI + CLIL (Grade 7 at T1) and students who only received FI at school without CLIL (Grade 8 at T1). In other words, we collected data from the last students who were studying only through FI, and from the first cohort of learners who joined the newly implemented FI + CLIL programme. Attending the CLIL programme was compulsory for all students from 2002 to 2003 onwards, and there

were no entry examinations. This inclusive model contributed to the homogeneity of the sample in several ways, as recommended by Pérez-Cañado (2016a): it avoided that the most academically motivated students opted for the CLIL strand if they had been given the choice (no student self-selection), which would have confined students with lower levels of ability and motivation to non-CLIL strands (Bruton 2011, 2013). Regarding SES, only one school was involved, which also guaranteed a similar, medium-high level of SES among participants.

Only two teachers taught English and Science to our learner sample and were involved in the research: an English FI teacher and a CLIL teacher. In fact, it had been the FI teacher, who was also the FL coordinator at the school, who proposed the implementation of a CLIL programme to the school board. CLIL experts from two universities3 provided one-year implementation assistance and CLIL training for all teachers upon request. Both staff and administrators were highly motivated and involved in the project.

Research design

Participants in both groups were tested at the end of two consecutive academic years. As shown in Table 1, data on Group A (FI+ CLIL) were collected at the end of Grade 7 (T1) and Grade 8 (T2),

when learners were 13 and 14 years old respectively. At T1, participants in Group A had received CLIL classes for three years (from Grade 5 onwards), as well as FI for 8 years (from pre-school to the end of Grade 7). On the other hand, data on Group B (FI only) were gathered at the end of Grade 8 (T1) and Grade 9 (T2), when students were 14 and 15 years old respectively. At T1 participants in Group B had followed the traditional FI only for 9 years (from pre-school to the end of Grade 8).

Regarding the number of hours of L2 exposure in each group, students received approximately 140 h of FI every year since pre-school (age 5). Furthermore, learners enrolled in the CLIL programme received 70 additional hours of instruction in CLIL from Grade 5 (age 10) onwards. Table 2 shows that Group A (FI + CLIL) had accumulated 1330 h of instruction at T1 and 1540 h at T2. On the other hand, Group B (FI only) had had 1260 h of instruction at T1 and 1400 h at T2.

To answer the two research questions about the influence of age and exposure on learners' L2 skills and grammar knowledge, two between-group comparisons were conducted (see Table 3).

As shown in Table 3, in the first condition, participants in the two groups differed in age (2 years of difference) but had a similar amount of L2 exposure. The difference in hours between Group A and B was small (70 h) considering the threshold of 300 h which could

affect L2 results, as proposed by Muñoz (2015). Therefore, the independent variable in this comparison was participants' age.

In the second condition, participants in the two groups were the same age (14 years) but differed in the number of hours of exposure in favour of the CLIL group (280 h of difference). Accordingly, the independent variable in the second condition was number of hours of instruction in English, which was very close to the threshold of 300 h proposed by Muñoz (2015).

Instruments and data collection procedures

Data on receptive skills (i.e. reading and listening) were gathered using (i) a cloze test for reading comprehension and (ii) a dictation for listening comprehension. Data on productive skills (writing) were collected using a composition based on pictures. Participants' grammar knowledge was measured using a multiple choice test and a Grammaticality Judgement Test (GJT). All the data were collected by school teachers in two hours. Students were given some free time (5–10 min) after each test.

The order that was followed for the collection of data using the battery of instruments was the following: writing, reading, listening and grammar knowledge (see Appendix A). Data were collected using the same instruments and following the same order of administration at T1 and T2. Although the instruments used were the same at both times of data collection, this was not considered to affect the results due to the large time span (12 months) between the first and the second time of data collection. Participants were unlikely to remember the answers given in such a long period of time. In addition, the answers were not provided to the students after having collected the data at T1, and participants were not informed about the existence of a second time of data collection one year later. In what follows, we describe the instruments used to gather data.

Cloze test

Reading comprehension was measured using a cloze test because these tests are considered valid to analyse students' ability to process and understand lengthy passages and to decode interrupted messages (Weir 1998). In our study, participants had to read a short text with blanks that they had to complete choosing the correct options. The text was about science (e.g tsunamis) and all the participants were acquainted with the topic since they had already studied that content in science classes. Students completed the test in 15 min.

Dictation

A 15-minute dictation was used to collect data on listening comprehension. Dictations are considered suitable instruments to test learners' comprehension as they involve both lower and higher-order

abilities to decode and write what is heard (Cohen 1980; Johansson 1973; Oller 1979). Furthermore, dictations draw on a broad range of integrative skills, such as the ability to discriminate phonological units and to define word boundaries so as to reconstruct the message (Oller 1979). The topic of the dictation used in this study was familiar to all learners since it dealt with content already studied in Science classes.

Composition

Learners were given a picture. Then they were asked to write a dialogue about it individually and to answer three questions: 'What happened?', 'Why did that situation happen?' and 'How do you think the situation will end?' (see Appendix A). The pictures and questions were used because they were considered inspiring for secondary school students (Tavakoli and Foster 2008). For this task, students were given 20 min.

Multiple choice test and grammaticality judgement test (GJT)

These tests were used to measure students' grammar knowledge. The tests were created based on what secondary school students were expected to know according to the official curriculum for the teaching of foreign languages (e.g. present tense; third person singular agreement morpheme; comparative and superlative adjectives; quantifiers). The tests complemented each other; the multiple choice test gathered information on learners' grammatical knowledge by having students recognise a correct answer out of three possible answers, whereas the GJT measured students' grammatical knowledge by having them assess the correctness of isolated sentences. The sentences of the GJT were related to the questions formulated in the multiple choice test.

The mutiple choice test consisted of 30 questions of increasing difficulty that had to be answered in 15 min. As for the GJT, it consisted of a list of 20 sentences that learners had to read and decide whether they were grammatically correct or not in 10 min.

Data analysis

Different procedures were followed to analyse the data. The cloze test, the multiple choice test and the GJT were corrected on the basis of the number of correct answers. One point was granted for each correct answer. The total number of correct answers obtained in the

multiple choice test and the GJT was added up to obtain a single measure of grammar knowledge.

To score the dictation, we followed Oller (1979, 1993). This involved counting the number of correct words appearing in the dictations written by the participants. A high number of words written that corresponded to the original words of the dictation was indicative of high listening skills. Words did not have to be spelled perfectly to be considered correct, as the objective of the test was to measure listening comprehension and not spelling.

Compositions were analysed both quantitatively and qualitatively by two experienced raters. For the quantitative analysis, the compositions were transcribed using the Computerized Language Analysis programme (CLAN). Then, we measured the fluency, accuracy, and lexical and syntactic complexity of the compositions following the measures recommended by Wolfe-Quintero, Inagaki, and Kim (1998) in their review of analytic measures in writing development.

Fluency was computed on the basis of the total number of words written. Since the compositions were written in 20 min, the higher the number of words produced, the more fluent the participant was. Accuracy was calculated by counting the total number of errors per word. Finally, complexity was measured with a double dimension regarding syntax and lexis. At the level of syntax, complexity was measured using the coordination index following Bardovi-Harlig (1992), who proposed that, at initial levels of L2 competence, coordination may be a more valid measure of complexity than subordination. We calculated the coordination index of sentences by dividing the independent coordination clauses by the total number of coordinate and subordinate clauses. As for complexity at the level of lexis, we explored the lexical richness of texts using Guiraud's index, since the increase of

lexical richness in texts could be an indication of the development of complexification when using more advanced language (Norris and Ortega 2009). Guiraud's index was used to compensate for the effect of text length, which impacts other measures of lexical richness such as type token ratio. Guiraud's index was calculated by dividing the total number of types by the square root of the total number of tokens.

When all the quantitative measures of writing were computed, frequencies were obtained using CLAN, and then they were introduced into a Stats Graphic matrix to calculate ratios.

As for the qualitative analyses of compositions, two experienced raters carried out the assessment, as recommended by McNamara (2009), to avoid subjectivity in ratings. After mixing up the compositions of both groups of participants, they blindly evaluated compositions together until agreement was achieved. Then, each reviewer proceeded with the individual assessment of compositions; again, blindly. The two raters assessed participants' linguistic competence (i.e. grammar and vocabulary) and communicative competence (i.e. task fulfilment and organisation) in writing, after Friedl and Auer's 6-point-rating scale (2007) (see Appendix A). The maximum score that learners could obtain in all

qualitative measures of writing was 20.

Linguistic competence was measured on account of grammar (i.e. errors, structure variety, use of complex structures) and vocabulary features (i.e. range and choice of words, spelling, comprehensibility, accuracy and word usage). On the other hand, participants' communicative competence in writing was based on task fulfilment and organisation. To measure task fulfilment the raters analysed the appropriateness of the compositions' text format, length (i.e. not too short, not too long), register (i.e. frequency of gaps, redundancy of information) and content (i.e. relevance). Finally, the raters evaluated the organisation of compositions assessing structure, paragraphing, cohesion, coherence, editing, and punctuation.

When all the data obtained from the different instruments (cloze test, dictation, composition, multiple choice test, GJT) were computed, statistical analyses were conducted. Mann Whitney U-tests were used to compare the results of the two groups (FI + CLIL vs FI) given the small sample size of the study and the violation of some assumptions of ANOVA.

Results

Regarding the first research question, and as shown in Table 4, when the participants differed in age (Group A, FI + CLIL: 12/13 years vs Group B, FI: 14/15 years) but had similar number of hours of exposure to English (only 70 h of difference), the FI participants, who were two years older, performed significantly better in the receptive skill of listening and in some dimensions of the productive skill of writing measured quantitatively: (i) use of coordination and (ii) accuracy4 compared to the FI + CLIL learners. No significant differences between the two groups were found in grammar knowledge.

Concerning the second research question, Table 5 indicates that when participants in both groups were the same age (13/14 years) but differed in the number of hours of exposure to English (280 h of difference in favour of Group A, FI + CLIL), the FI + CLIL participants, who received more hours of instruction, outperformed the FI learners in (i) the receptive skill of reading; (ii) in one quantitative dimension of writing, the Guiraud index; and (iii) in all the qualitative dimensions of writing related to linguistic and communicative competence. Regarding linguistic competence, the FI + CLIL group outstripped the FI group in grammar, and vocabulary. As for the communicative competence of writing, the FI + CLIL outperformed the FI group in task fulfilment and organisation. No significant differences between the two groups were found in grammar knowledge.

Discussion

This study set out to investigate the influence of age and exposure to English in two different instructional treatments (FI + CLIL vs FI) on EFL learners' L2 skills (receptive and productive) and grammar knowledge.

As for our first research question, our findings show that when learners were matched for hours of exposure in two different instructional programmes but differed in age, older students without CLIL instruction were better than their CLIL counterparts in one receptive skill (i.e. listening) and in two quantitative dimensions of writing (accuracy and coordination index). Therefore, in this study an older age seems to counterbalance the positive impact of a CLIL programme in some language dimensions. These results confirm Muñoz's (2015) first hypothesis: when learners have the same hours of L2 exposure, older students in non-CLIL programmes seem to have a faster rate of learning than younger learners in CLIL strands in some language dimensions due to their greater cognitive maturity.

Our findings are consistent with previous CLIL studies. Villarreal (2011) found that when matching CLIL and non-CLIL groups for hours of instruction, non-CLIL learners, who were 2 years older, showed slight advantages over CLIL students in the oral production of verbal morphology. Likewise, Schoonjans (2013) also reported that mainstream learners, who were also two years older, were better than ES students, after similar hours of exposure in both instructional contexts (580 h), in general English proficiency, as well as in some measures of accuracy and lexical complexity, as we also found in our study.

Our results are also in line with studies from instructed learning contexts, which have consistently shown that older students are faster and more efficient learners than younger ones due to their higher capacity for abstraction and logical thinking (García-Mayo and García Lecumberri 2003; Muñoz 2006, 2008).

From a cognitive perspective, it seems that our older FI participants, who received only formal instruction, outperformed younger students in FI + CLIL in listening due to the characteristics of the test, which favoured learners with greater cognitive maturity (Cummins 1980). Older students could handle highly cognitively-demanding, context-reduced tasks (Cummins 1982) more successfully. The listening test involved using lower and higher-order thinking skills to recognise words as part of a dictation, as well as making sense of the words to retrieve the content previously studied in Science classes.

Along the same lines, our FI students' higher performance in two quantitative dimensions of writing (i.e. accuracy and coordination) could also be related to their greater cognitive ability, which may have enabled them to deal with linguistic aspects (Krashen, Long, and Scarcella 1979). As for accuracy, it could be argued that older FI learners had an advantage over younger FI + CLIL students on rule acquisition (Snow and Hoefnagel-Höle 1978), which is essential to make fewer errors related to morphology, syntax and lexis when writing. Likewise, the significant difference in coordination index when writing in favour of

the FI group (older students) could be indicative of complexification at initial levels of development (Bardovi-Harlig 1992). In the case of grammatical knowledge, the FI group also tended to score higher than the FI + CLIL group, although the difference was not significant.

Regarding the second research question, we found that when participants in two different instructional treatments (FI + CLIL vs FI) were the same age but had a different number of hours of instruction in favour of the FI + CLIL group (280 h of difference), participants in the FI + CLIL treatment outperformed the FI students in one receptive skill (i.e. reading), in one quantitative dimension of writing (i.e. lexical richness), and in all qualitative measures of writing (i.e. grammar, vocabulary, task fulfilment and organisation). These findings are in line with Muñoz's (2015) second hypothesis, according to which 300 h of additional L2 exposure seem to make a difference between CLIL and non-CLIL groups in favour of CLIL strands. In our study, there were 280 h of difference in favour of the FI + CLIL group, which is fairly close to the 300-hour threshold. Maybe for this reason students begin to show improvements in certain receptive (i.e. reading) and productive L2 skills (i.e. writing) but not in all L2 skills (i.e. listening) or in grammar knowledge.

It is interesting to note that our FI + CLIL participants were significantly better in the range of productive use of vocabulary measured both quantitatively (i.e. lexical richness) and qualitatively in writing. These findings are consistent with previous studies that indicate that the vocabulary of

CLIL learners tends to be lexically richer (Jiménez-Catalán, Ruiz de Zarobe, and Cenoz 2006; Moreno 2009). This lexical richness could also be related to differences in reading comprehension favourable to the FI + CLIL group. To choose the correct option in the cloze test, our participants needed not only to read and understand the text but also to apply L2 knowledge, more specifically, lexico-gramma-tical knowledge. As indicated by previous research, there is a positive relationship between learners' knowledge of vocabulary and the quality of reading comprehension (Coady et al. 1993; Laufer 1992). These findings coincide with those reported by Jiménez-Catalán, Ruiz de Zarobe, and Cenoz (2006), who found that primary school students in CLIL programmes outperformed their FI conterparts in an EFL context in both reading comprehension and a writing composition which measured the productive use of vocabulary.

Given that our participants in both instructional treatments (FI + CLIL vs FI) were the same age when they were tested, but they differed in the amount of L2 exposure in favour of the FI + CLIL, we suggest that our results are indicative of the benefits derived from a larger exposure to the L2 language. However, we cannot deny the possible effects of the quality of exposure and instruction on the specific language skills that were specifically practised in the CLIL classroom (Lasagabaster 2008; Navés and Victori 2010). A noteworthy portion of the additional hours in the FI + CLIL group were CLIL hours, in which practice was meaning-oriented basically through reading L2 texts about Science in each CLIL lesson. In contrast,

the FI group was also acquainted with the content of the reading comprehension text used for the data collection, but this FI group did not practise reading texts about Science in the L2. A strong focus on reading in the Science class may have boosted the L2 literacy abilities of students in the FI + CLIL group, not only by reading and understanding texts about Science but also by composing using the texts (in terms of L2 language use, structure and content) that they read in each CLIL lesson. In this respect, it should be noted that FI + CLIL group did not often practise L2 writing in the classroom. Therefore, extensive exposure to reading in the FI + CLIL group could account for the significant differences of these CLIL students in reading, as well as for the significant differences in all the qualitative dimensions of writing regarding linguistic competence (accuracy in grammar and vocabulary, variety in lexis and grammar) and communicative competence (task fulfilment and organisation). Along these lines, research has consistently shown the intimate relationship between reading and writing (e.g. Tierney and Pearson 1983), finding correlations between reading achievement and writing ability (e.g. Abbott and Berninger 1993; Berninger et al. 2002; Koons 2008).

Conclusion

This study intended to overcome some of the pinpointed research design flaws of previous studies (Pérez-Cañado 2016a) by controlling for variables such as ability and motivation to study the L2 language in CLIL strands, as well as SES, and finally by matching students first for age and later for exposure as part of the research design. The study contributes to research on age by treatment (DeKeyser 2012) and on the influence of exposure in different instructional treatments, FI and FI + CLIL (Muñoz 2015) with two main relevant findings.

Firstly, we can conclude that biological age has an impact on L2 learning when comparing FI students to FI + CLIL learners. Older students (at least 2 years) in the FI group outperformed younger learners in the FI + CLIL group (Muñoz 2015) in some L2 language dimensions when learners in both groups had similar hours of instruction in both treatments. We argue that this difference is due to the higher abstract and logical thinking abilities of the older students.

Secondly, when learners in the FI and the FI + CLIL groups were matched for age, our results highlight the L2 learning potential of a partial, low-intensity CLIL programme at secondary school level (introduced at Grade 10) in some receptive (e.g. reading) and productive L2 skills (e.g. writing) after only 280 additional hours of instruction distributed across two academic years. This finding underscores the efficacy of partial or low-intensity CLIL programmes for secondary school students.

Our results can guide much needed further research on biological age in CLIL programmes, as well as inform institutional stakeholders on the benefits of middle or late CLIL

implementations following partial or low intensity programmes in foreign language learning contexts.

Future research should investigate the role of both biological age, age of onset, and exposure in CLIL with a greater variety of age ranges and programme intensities (high, medium or low), to fully understand the influence of age and exposure on L2 language benefits (skills and lexico-grammatical knowledge) in secondary education levels. Future investigations should continue aiming at improving research designs. Finally, the instruments used to collect data about L2 skills and knowledge should be illustrative of classroom practises in both FI and CLIL treatments, to control the possible influence of transfer of knowledge and skills from classroom practise.

Notes

In this paper the terms 'hours of instruction', 'exposure' and 'hours of exposure' are used interchangeably to refer to the instructional hours (FI or CLIL depending on the condition) that students received at school.

European Schools (ES) offer studying the L2 from Grade 1, first as a subject, and then in grades 3–5 the L2 becomes the language of instruction during one to three 45-minute periods. A maximum of 30% of the curriculum can be taught through the second language, including FL classes. In secondary education, this percentage can increase up to 60% of the total school time table.

Universitat Pompeu Fabra and Universitat Autònoma de Barcelona.

It should be reminded that accuracy in writing was measured in terms of the number of errors per word. Therefore, the higher the score obtained in the measure of accuracy, the less accurate the writing was.

Instructions were provided to students in their L1. We have translated the instructions of all the instruments into English for the reader.

Acknowledgements

We are very grateful to Dr. Carmen Pérez-Vidal, who implemented CLIL at the school where data where gathered and supervised the larger study (Roquet 2011). Similarly, we would like to acknowledge the insightful comments of Dr. Carmen Muñoz and an anonymous reviewer on earlier versions of this article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Dr. Gemma Artieda is Adjunct Lecturer at the English Studies Department of the University of Barcelona (UB), where she teaches at the Master in Applied Linguistics and Language Acquisition in Multilingual Contexts. Her research interests are multilingualism, individual differences, adult foreign language acquisition, language aptitude, and CLIL (Content and Language Integrated Learning). She has published her research in peer-reviewed international journals such as Learning and Individual Differences and System.

Dr. Helena Roquet is the Director of the Institute for Multilingualism at Universitat Internacional de Catalunya (UIC Barcelona), where she is a Tenure-track Lecturer of Foreign Language Acquisition. Currently she is also Adjunct Lecturer at the Department of Translation and Language Sciences at Universitat Pompeu Fabra (UPF). Helena is a research team member of the consolidated research group ALLENCAM (Language Acquisition from Multilingual Catalonia) at UPF. Her main research interests lie within the field of second and third language acquisition, bilingualism, CLIL (Content and Language Integrated Learning) and ICLHE (Integrating Content and Language in Higher Education). Her articles have been published in leading peer-reviewed international journals like International Journal of Bilingual Education and Bilingualism and Applied Linguistics.

Florentina Nicolás-Conesa is Assistant Professor at the University Center of Defence in the Spanish Air Force Academy in Murcia, Spain. Her research interests include the analysis of writing processes, written products, students' cognition, and L2 learning and teaching. Her doctoral dissertation on "Development of Mental Models of Writing in a Foreign Language Context: Dynamics of Goals and Beliefs" received the best doctoral thesis award by the Spanish Society of Applied

Linguistics. She has published her work in the Journal of Second Language Writing, in the TESOL Encyclopedia of English Language Teaching, in the Handbook of Second and Foreign Language Writing and in Spanish journals.

Gemma Artieda http://orcid.org/0000-0001-8807-8571

Helena Roquet http://orcid.org/0000-0001-6694-3288

References

Abbott, R. D., and V. W. Berninger. 1993. “Structural Equation Modeling of Relationships among Developmental Skills and Writing Skills in Primary and Intermediate Grade Writers.” Journal of Educational Psychology 85 (3): 478–508.

Admiraal, W., G. Westhoff, and K. de Bot. 2006. “Evaluation of Bilingual Secondary Education in the Netherlands: Students’ Language Proficiency in English.” Educational Research and Evaluation 12 (1): 75–93.

Bardovi-Harlig, K. 1992. “A Second Look at T-unit Analysis: Reconsidering the Sentence.” TESOL Quarterly 26: 390–395. Berninger, V., R. D. Abbott, S. P. Abbott, S. Graham, and T. Richards. 2002. “Writing and Reading Connections between

Language by Hand and Language by Eye.” Journal of Learning Disabilities 35 (1): 39–56.

Birdsong, D. 2006. “Age and Second Language Acquisition and Processing: A Selective Overview.” Language Learning 56

(1): 9–49.

Birdsong, D. 2014. “The Critical Period Hypothesis for Second Language Acquisition: Tailoring the Coat of Many Colors.” In Essential Topics in Applied Linguistics and Multilingualism, Second Language Learning and Teaching, edited by M. Pawlak and L. Aronin. Switzerland: Springer International. doi:10.1007/978-3-319-01414-2_3.

Bruton, A. 2011. “Is CLIL so Beneficial, or Just Selective? Re-evaluating Some of the Research.” System 39: 523–532. Bruton, A. 2013. “CLIL: Some of the Reasons Why … and Why Not.” System 41: 587–597.

Coady, J., J. Magoto, P. Hubbard, J. Graney, and K. Mokhtari. 1993. “High Frequency Vocabulary and Reading Proficiency in ESL Readers.” In Matsuoka and Hirsh: Second Language Reading and Vocabulary Learning, edited by T. Huckin, M. Haynes and J. Coady, 217–228. Norwood, NJ: Ablex.

Cohen, L. 1980. Research Methods in Education. London: Croom Helm.

Coyle, D. 2007. “Content and Language Integrated Learning: Towards a Connected Research Agenda for CLIL Pedagogies.” The International Journal of Bilingual Education and Bilingualism 10 (5): 543–562.

Coyle, D., P. Hood, and D. Marsh. 2010. CLIL. Content and Language Integrated Learning. Cambridge: Cambridge University Press.

Cummins, J. 1980. “The Entry and Exit Fallacy in Bilingual Education.” In An Introductory Reader to the Writings of Jim Cummins, edited by C. Baker and N. Hornberger, 110–138. Clevedon: Multilingual Matters.

Cummins, J. 1982. “Tests, Achievement, and Bilingual Students.” National Clearinghouse for Bilingual Education 9: 2–9.

Dalton-Puffer, C. 2008. “Outcomes and Processes in Content and Language Integrated Learning (CLIL): Current Research from Europe.” International Journal of Bilingual Education and Bilingualism 10: 543–562.

DeKeyser, R. 2000. “The Robustness of Critical Period Effects in Second Language Acquisition.” Studies in Second Language Acquisition 22 (4): 499–533.

DeKeyser, R. 2012. “Interactions between Individual Differences, Treatments, and Structures in Second Language Acquisition.” Language Learning 62 (2): 189–200.

DeKeyser, R. 2013. “Age Effects in Second Language Learning: Stepping Stones Toward Better Understanding.” Language Learning 63 (1): 52–67.

European Commission. 1995. White Paper on Education and Learning. Brussels: European Commission.

Eurydice. European Commission. 2006. Content and Language Integrated Learning (CLIL) at School in Europe. Brussels: Eurydice.

Friedl, G., and M. Auer. 2007. “Rating Scale Used for Assessment of the Writing Task.” Erläuterungen zur Novellierung del Reifeprufungsverordnung fur AHS, lebende Fremdsprachen. BIFIE. Wien – St. Pölten. http://www.bifie.at/ publikationen.

Gallardo, F., M. L. García Lecumberri, and E. Gómez Lacabex. 2009. “Testing the Effectiveness of Content and Language Integrated Learning in Foreign Language Contexts: Assessment of English Pronunciation.” In Content and Language Integrated Learning: Evidence from Research in Europe, edited by Y. Ruiz de Zarobe and R. M. Jiménez Catalán, 63–80. Bristol: Multilingual Matters.

García-Mayo, M. P., and M. L. García Lecumberri. 2003. Age and the Acquisition of English as a Foreign Language. Clevedon: Multilingual Matters.

Grisaleña, J., E. Alonso, and A. Campo. 2009. “Enseñanza Plurilingüe en Centros de Educación Secundaria: Análisis de Resultados.” Revista Iberoamericana de Educación 49 (1): 1–12.

Housen, A. 2012. “Time and Amount of L2 Contact Inside and Outside the School. Insights from the European Schools.” In

Intensive Exposure Experiences in Second Language Learning, edited by C. Muñoz, 111–138. Bristol: Multilingual Matters.

Järvinen, H. 2010. “Language as a Meaning Making Resource in Learning and Teaching Content.” In Language use and Language Learning in CLIL Classrooms, edited by C. Dalton Puffer, T. Nikula and U. Smit, 145–168. Amsterdam: John Benjamins.

Jiménez-Catalán, R. M., Y. Ruiz de Zarobe, and J. Cenoz. 2006. “Vocabulary Profiles of English Foreign Language Learners in English as a Subject and as a Vehicular Language.” Views 15 (3): 23–26.

Johansson, S. 1973. “Partial Dictation as a Test of Foreign Language Proficiency.” Swedish-English Contrastive Studies

Report No. 3.

Koons, H. 2008. “The Reading-writing Connection: An Investigation of the Relationship Between Reading Ability and Writing Quality Across Multiple Grades and Three Writing Discourse Modes.” PhD Thesis., University of North Carolina. Krashen, S., M. H. Long, and R. C. Scarcella. 1979. “Age, Rate, and Eventual Attainment in Second Language Acquisition.”

TESOL Quarterly 13: 573–582.

Lasagabaster, D. 2008. “Foreign Language Competence in Content and Language Integrated Courses.” The Open Applied Linguistics Journal 1 (1): 30–41.

Lasagabaster, D. 2011. “English Achievement and Student Motivation in CLIL and EFL Settings.” Innovation in Language Learning and Teaching 5: 3–18.

Lasagabaster, D., and J. M. Sierra. 2009. “Language Attitudes in CLIL and Traditional EFL Classes.” International CLIL Research Journal 1 (2): 4–17.

Laufer, B. 1992. “Reading in a Foreign Language: How Does L2 Lexical Knowledge Interact with the Reader’s General Academic Ability?” Journal of Research in Reading 15: 95–103.

Lázaro Ibarrola, A. 2012. “Faster and Further Morphosyntactic Development of CLIL vs. EFL Basque-Spanish Bilinguals Learning English in High-school.” International Journal of English Studies 12 (1): 79–96.

Llanes, A., and C. Muñoz. 2013. “Age Effects in a Study Abroad Context: Children and Adults Studying Abroad and at Home.” Language Learning 63 (1): 63–90.

Lorenzo, F., S. Casal, and P. Moore. 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. Martínez Adrián, M., and M. J. Gutiérrez Mangado. 2009. “The Acquisition of English Syntax by CLIL Learners in the

Basque Country." In Content and Language Integrated Learning: Evidence from Research in Europe, edited by Y. Ruiz de Zarobe

and R. M. Jiménez Catalán, 176–196. Bristol: Multilingual Matters.

McNamara, T. F. 2009. Personal Communication at the University of Vienna. Vienna: University of Vienna.

Mehisto, P., D. Marsh, and M. J. Frigols. 2008. Uncovering CLIL. Content and Language Integrated Learning in Bilingual and Multilingual Education. London: Macmillan.

Merino, J. A., and D. Lasagabaster. 2015. "CLIL as a Way to Multilingualism." International Journal of Bilingual Education and Bilingualism. doi:10.1080/13670050.2015.1128386.

Moreno, S. 2009. "Young Learners' L2 Word Association Responses in Two Different Learning Contexts." In Content and Language Integrated Learning: Evidence from Research in Europe, edited by Y. Ruiz de Zarobe and R. M. Jiménez Catalán, 93–111. Bristol: Multilingual Matters.

Muñoz, C. 2002. "CLIL-AICLE (Aprendizaje Integrado de Contenidos y Lengua Extranjera). Relevance and Potential of CLIL." In CLIL/EMILE. The European Dimension. Actions, Trends and Foresight Potential, edited by D. Marsh, 33–36. Jyväskyla: University of Jyväskyla.

Muñoz, C. 2006. "The Effects of Age on Foreign Language Learning: The BAF Project." In Age and the Rate of Foreign Language Learning, edited by C. Muñoz, 1–40. Clevedon: Multilingual Matters.

Muñoz, C. 2007. "CLIL: Some Thoughts on its Psycholinguistic Principles." In Models and Practice in CLIL, edited by F. Lorenzo, S. Casal, V. de Alba, and P. Moore, 17–26. Logroño: Monográfico de la Revista Española de Lingüística Aplicada (Resla).

Muñoz, C. 2008. "Age-related Differences in Foreign Language Learning. Revisiting the Empirical Evidence." IRAL -International Review of Applied Linguistics in Language Teaching 46: 195–220.

Muñoz, C. 2015. "Time and Timing in CLIL: A Comparative Approach to Language Gains." In Content-based Learning in Multilingual Educational Environments, edited by M. Juan-Garau and J. Salazar-Noguera, 87–105. Berlin: Springer.

Navés, T. 2011. "How Promising are the Results of Integrating Content and Language for EFL Writing and Overall EFL Proficiency?" In Content and Foreign Language Integrated Learning: Contributions to Multilingualism in European Contexts, edited by Y. Ruiz de Zarobe, J. M. Sierra and F. Gallardo del Puerto, 155–186. Bern: Peter Lang.

Navés, T., and M. Victori. 2010. "CLIL in Catalonia: An Overview of Research Studies." In CLIL in Spain: Implementation, Results and Teacher Training, edited by Y. Ruiz de Zarobe, and D. Lasagabaster, 30–54. Newcastle upon Tyne: Cambridge Scholars Publishing.

Norris, J. M., and L. Ortega. 2009. “Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity.” Applied Linguistics 30 (4): 555–578.

Novotná, J., and M. Hofmannová. 2000. “CLIL and Mathematics Education.” Proceedings of the international conference on

‘mathemathics for living: The mathematics education into the 21st century project, CLIL, 226–230.

Oller, J. W. 1979. Language Testing at School: A Pragmatic Approach. London: Longman.
Oller, J. W. 1993. Methods that Work. Boston: Heinle & Heinle.

Pérez-Cañado, M. L. 2012. “CLIL Research in Europe: Past, Present and Future.” International Journal of Bilingual Education and Bilingualism 15 (3): 315–341.

Pérez-Cañado, M. L. 2016a. “Stopping the ‘Pendulum Effect’ in CLIL Research: Finding the Balance between Pollyanna and Scrooge.” Applied Linguistics Review. doi:10.5565/rev/jtl3.667.

Pérez-Cañado, M. L. 2016b. ““From the CLIL Craze to the CLIL Conundrum: Addressing the Current CLIL Controversy.”

Bellaterra Journal of Teaching & Learning Language & Literature 9 (1): 9–31.

Pérez-Vidal, C. 2011. “Language Acquisition in Three Different Contexts of Learning: Formal Instruction, Stay Abroad, and Semi-immersion (CLIL).” In Content and Foreign Language Integrated Learning: Contributions to Multilingualism in European Contexts, edited by Y. Ruiz de Zarobe, J. M. Sierra and F. Gallardo, 103–127. Bern: Peter Lang.

Rifkin, B. 2005. “A Ceiling Effect in Traditional Classroom Foreign Language Instruction: Data from Russian.” The Modern Language Journal 89 (1): 3–18.

Robinson, P. 2007. “Aptitudes, Abilities, Contexts and Practice.” In Practice in a Second Language. Perspectives from Applied Linguistics and Cognitive Psychology, edited by R. DeKeyser, 256–286. Cambridge: Cambridge University Press.

Roquet, H. 2011. Integrating Content and Language in Mainstream Education in Barcelona. A Study of the Acquisition of English as a Foreign Language. Saarbrücken: Lambert Academic Publishing.

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 (Special Issue: Current Research on CLIL 2) 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 (1): 60–73.

San Isidro, X. 2010. “An Insight into Galician CLIL: Provision and Results.” In CLIL in Spain: Implementation, Results and Teacher Training, edited by D. Lasagabaster and Y. Ruiz de Zarobe, 55–78. Newcastle: Cambridge Scholars Publishing. Schoonjans, E. 2013. “Second/Foreign Language Learning in Context. Investigating Linguistic Proficiency and Socio-

affective Dispositions in Four Different Learning Contexts.” Doctoral Diss., Universiteit Brussel, Vrije.

Seikkula-Leino, J. 2007. “CLIL Learning: Achievement Levels and Affective Factors.” Language and Education 21 (4): 328–341.

Snow, C. E., and M. Hoefnagel-Höle. 1978. “The Critical Period for Language Acquisition: Evidence from Second Language Learning.” Child Development 49: 1114–1128.

Tavakoli, P., and P. Foster. 2008. “Task Design and Second Language Performance: The Effect of Narrative Type on Learner Output.” Language Learning 58 (2): 439–473.

Tierney, R. J., and P. D. Pearson. 1983. “Toward a Composing Model of Reading.” Language Arts 60 (5): 568–580.

Van de Craen, P., K. Mondt, L. Allain, and Y. Gao. 2007. “Why and How CLIL Works. An Outline for a CLIL Theory.” Vienna English Working Papers 16 (3): 70–78. CLIL Special Issue.

Villarreal, I. 2011. “Tense and Agreement in the Non-native English of Basque-Spanish Bilinguals: Content and Language Integrated Learners vs. English as a School Subject Learners.” Doctoral Diss., Universidad del País Vasco, Spain.

Villarreal, I., and M. P. García-Mayo. 2009. “Tense and Agreement Morphology in the Interlanguage of Basque/Spanish Bilinguals: CLIL Versus Non-CLIL.” In Content and Language Integrated Learning: Evidence from Research in Europe, edited by Y. Ruiz de Zarobe and R. M. Jiménez Catalán, 157–175. Bristol: Multilingual Matters.

Weir, C. J. 1998. Communicative Language Testing with Special Reference to English as a Foreign Language. Vol. 11. Exeter: Exeter Linguistic Studies, University of Exeter.

Wolfe-Quintero, K., S. Inagaki, and H. Kim. 1998. Second Language Development in Writing: Measures of Fluency, Accuracy and Complexity. Hawai’i: University of Hawai’i at Manoa.

Appendices

Appendix A. Instruments

Cloze test to measure reading comprehension

In this activity5 you need to choose the correct word for each gap. Read the whole text before before choosing the correct options so that you get a general idea of what the text is about. Check your choice for each answer. You have 15 min. Good luck!

Natural disasters such as volcanic eruptions, fires, floods and avalanches happen every year, somewhere in the world. But

(1) isanother, and perhaps more dangerous, (2) disaster, which we do not hear about very (3). A tsunami is (4) huge wave that can cause terrible damage (5) destruction.

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 metres (11). These giant waves can hit Japan, Indonesia, Central (12) and South America.

In 1960, there was an earthquake that measured 8.6 on the Richter (13) at the (14) of the Pacific Ocean. It started a giant wave that (15) the coast of Chile and Peru (16). Result was that between 1500 and 2000 people were killed, and whole villages were (17). The wave then travelled (18) 15 h to Hawaii, to the town of Hilo, and another 200–300 people lost their

(19). It was only one wave, but it caused $500 million of damage.

Tsunamis do not happen very often, but Hawaii now has a 'tsunami watch' station that (20) for the next one to come. The station was opened to warn people and give them time to protect themselves against the killer waves.

Dictation to measure listening comprehension

This activity is a dictation. Pay attention and write what the teacher says. The teacher will stop for a few seconds after each group of words.

Don’t worry if you don’t understand everything. Just write the words you understand. If you miss any words you can revise your text later and note down the word that you think is missing.

Go ahead!

Antarctica

Composition based on pictures to measure writing skills

Look at this picture carefully and write short texts following the instructions given. You have 20 min to write as much as you can.

Multiple choice test to measure learners’ grammar knowledge

In this exercise you need to choose the option that you think is correct. You have three options, but only one is right. You have 15 min to finish the activity. You cannot ask any questions. Do not answer if you don’t know the answer to the question. This is going to be an easy activity for you!

Do you can change this 5 € note?

Have you change for this 5 € note?

Have you got change for this 5 € note? No, I’m sorry, I haven’t got any change.

Sally often to do her homework.

forgetting

is forgetting

forgets

How old is your sister?

She has got ten years old.

She is ten years old.

She is ten years.

That green cup on the table is .

my

mine

you

Monday, I must go to school.

At

In

On

Sarah plays tennis the weekend.

in

on

at

What are you wearing?

I wearing green socks

I'm wearing green socks

Wearing green socks

There are children in the park.

a lot of

much

so more

My car is than yours.

the biggest

bigger

very big

Mary don’t usually speak with people.

Mary do no usually speak with people.

Mary doesn’t usually speak with people.

Now try the following questions:

Did you have lunch at home yesterday?

No, I haven’t.

No, I didn’t.

No, I hadn’t.

I visited London last year.

How long have you stayed there?

How long did you stay there?

How long you stay there?

What were you see in London last year?

What have you seen in London last year?

What did you see in London last year?

What is your father doing now?

He has just have lunch.

He is having lunch.

He has lunch.

Where do you live?

At, 45 Green Street, London.

In 45, Green Street, London.

From 45, Green Street, London.

My mother usually cooks dinner for us, but she's in hospital, so

my father does it this week.

has done it this week.

is doing it this week.

I really love her.

She is the intelligentest person I know.

She is the most intelligent person I know.

She is the more intelligent person I know.

I think this is film I have ever seen.

the worse

the baddest

the worst

My mother is than my father.

happy

more happy

happier

How people are there in class?

much

more

many

The following questions are a bit more dificult, but surely you can answer some of them too.

Oh, dear! We don't have money.

many

much

too many

There wasn't in the building when the fire started.

nobody

somebody

anybody

You mustn't smoke in class!

You don't smoke in class!

You can not smoke in class!

Why are you putting on your coat?

Because I took the dog for a walk.

Because I'll take the dog for a walk.

Because I'm going to take the dog for a walk.

Please, do not disturb me.

I am trying to study.

I try to study.

I will try study.

I to London.

never went

have never been

didn't ever go

My friend John has shown me his car the day before yesterday.

My friend John showed me his car the day before yesterday.

My friend John shown me his car the day before yesterday.

I think it is going to rain tomorrow.

I think it is raining tomorrow.

I think it rained tomorrow.

She learns everything very

easy.

easily.

easier.

They invited us dinner with them.

having

to have

have

Grammaticality judgement test to measure grammar knowledge

In this activity you have to decide whether the sentences in English presented below are correct or not. Tick the correct box accordingly. You have 10 min to do the activity. You cannot ask any questions.

Appendix B. Rating scale used for the assessment of the writing task (Friedl and Auer 2007)

Task fulfilment:content and relevance; text format, length and register

5 Task fully achieved, content entirely relevant; appropriate format, length and register.

4 Task almost fully achieved, content mostly relevant; mostly appropriate format, length and register.

3 Task adequately achieved, some gaps or redundant information, acceptable format, length and register.

2 Task achieved only in a limited sense, frequent gaps or redundant information, often inadequate format, length and register.

1 Task poorly achieved, major gaps or pointless repetition; inadequate format, length and register.

0 Not enough to evaluate.

Organisation: Structure, paragraphing, cohesion and coherence, editing and punctuation

5 Clear overall structure, meaningful paragraphing; very good use of connectives, no editing mistakes, conventions of punctuation observed.

4 Overall structure mostly clear, good paragraphing, good use of connectives, hardly any editing mistakes, conventions of punctuation mostly observed.

3 Adequately structured, paragraphing misleading at times, adequate use of connectives; some editing and punctuation errors.

2 Limited overall structuring, frequent mistakes in paragraphing, limited use of connectives; frequent editing and punctuation errors.

1 Poor overall structuring, no meaningful paragraphing, poor use of connectives; numerous editing and punctuation errors.

0 Not enough to evaluate.

Grammar: Accuracy/errors, variety of structures, readiness to use complex structures

5 Accurate use of grammar and structures, hardly any errors of agreement, tense, word order, articles, pronouns, etc.; meaning clear, great variety of structures, frequent use of complex structures.

4 Mostly accurate use of grammar and structures, few errors of agreement etc.; meaning mostly clear; good variety of structures, readiness to use complex structures.

3 Adequate use of grammar and structures; some errors of agreement etc.; meaning sometimes not clear; adequate variety of structures; some readiness to use complex structures.

2 Limited use of grammar and structures; frequent errors of agreement etc.; meaning often not clear; limited variety of structures; limited readiness to use complex structures.

1 Poor use of grammar and structures; numerous errors of agreement etc.; meaning very often not clear; poor variety of structures.

0 Not enough to evaluate.

Vocabulary: Range and choice of words, accuracy, spelling, comprehensibility

5 Wide range of vocabulary; very good choice of words; accurate form and usage; hardly any spelling mistakes; meaning clear.

4 Good range of vocabulary; good choice of words; mostly accurate form and usage, few spelling mistakes; meaning mostly clear.

3 Adequate range of vocabulary and choice of words; some repetitions; some errors of form and usage; some spelling mistakes; meaning sometimes not clear; some translation from mother tongue.

2 Limited range of vocabulary and choice of words; frequent repetitions; frequent errors of form and usage; frequent spelling mistakes; meaning often not clear; frequent translation from mother tongue.

1 Poor range of vocabulary and choice of words; highly repetitive; numerous errors of form and usage; numerous spelling mistakes; meaning very often not clear; mainly translation from mother tongue.

0 Not enough to evaluate.