Francisco Gallardo del Puerto* and Esther Gómez Lacabex

Oral production outcomes in CLIL: an attempt to manage amount of exposure

Abstract: This study aims at testing the benefits of CLIL (Content and Language Integrated Learning) on oral skills in English by comparing a group of secondary school CLIL learners to two Non-CLIL groups matched for amount of instruction – a two-year-ahead group and a peer group. This sampling design is an attempt to tease the effects of exposure, age and CLIL variables apart, something which has not been addressed in most previous CLIL research. The analyses (holistic evaluation, amount and density of production, and use of compensatory strategies) of participants' story tellings indicate that CLIL learners' oral abilities are superior to those of Non-CLIL groups, especially to those of their exposure-matched peer counterparts. Overall, CLIL learners produce denser and richer oral narrations characterized by better content, vocabulary, grammar and fluency, and a marginal use of the first language. These results could be read as indicative of the beneficial effect of CLIL instruction itself on oral production when intervening variables such as amount of exposure and age are managed. In addition, particular attention is given to the lack of positive effect of CLIL on pronunciation.

Keywords: CLIL, oral skills, English as a Foreign Language, amount of exposure

Resumen: Este estudio intenta comprobar los beneficios del AICLE (Aprendizaje Integrado de Contenidos y Lengua Extranjera) sobre las habilidades orales en inglés al comparar un grupo de estudiantes AICLE en Educación Secundaria con otros dos grupos No-AICLE con una cantidad de instrucción similar: un grupo dos años mayor y un grupo de la misma edad. Este diseño de la muestra constituye un intento por controlar los efectos de las variables exposición, edad y AICLE, algo que no se ha llevado a cabo en la mayoría de las investigaciones previas sobre AICLE. Los análisis (evaluación holística, cantidad y densidad de la producción y uso de estrategias de compensación) de las historias contadas por los partici-

*Corresponding author: Francisco Gallardo del Puerto,** University of Cantabria – Department of Philology, Edificio de Filología. Avenida de los Castros s/n 39005 Santander, Cantabria, Spain, E-mail: francisco.gallardo@unican.es

Esther Gómez Lacabex, University of the Basque Country – English and German Philology and Translation and Interpretation Vitoria-Gasteiz, Araba, Spain, E-mail: esther.glacabex@ehu.eus

pantes indican que las habilidades orales de los estudiantes AICLE son superiores a las de los grupos No-AICLE, especialmente a las de aquellos que tienen su misma edad. En general, los estudiantes AICLE producen unas narraciones orales más densas y ricas caracterizadas por un mejor contenido, vocabulario, gramática y fluidez, así como por un uso marginal de la primera lengua. Estos resultados son indicativos del efecto beneficioso de la instrucción AICLE en si misma sobre la producción oral, cuando variables intervinientes como la cantidad de exposición o la edad están controladas. También se discute la ausencia del efecto positivo del AICLE sobre la pronunciación en particular.

Palabras Clave: AICLE, habilidades orales, Inglés como Lengua Extrajera, cantidad de exposición.

Zusammenfassung: Die vorliegende Studie prüft die Vorteile des CLIL (Content and Language Integrated Learning) in Bezug auf die mündliche Kompetenz im Fach Englisch, indem sie eine CLIL-Lernergruppe der Sekundarstufe mit zwei Gruppen vergleicht, die nicht im Rahmen von CLIL lernen und welche die gleichen Anweisungen erhalten: Eine zwei Jahre ältere und eine gleichaltrige Gruppe. Dieses Testdesign stellt einen Versuch dar, die Effekte der Variablen Alter, CLIL und die Dauer des Fremdsprachenlernens zu kontrollieren – etwas, was in den meisten vorherigen Studien nicht durchgeführt wurde. Die Analyse (holistische Evaluierung, Menge und Dichte der mündlichen Produktion sowie die Anwendung von Kompensierungsstrategien) der Teilnehmeräußerungen weist darauf hin, dass die mündliche Kompetenz der CLIL-Lerner ausgeprägter als die der Vergleichsgruppen ist, insbesondere gegenüber der gleichaltrigen Peergroup. Im Allgemeinen sind die mündlichen Produktionen von CLIL-Lernern umfangreicher an Inhalt, Vokabular und Grammatik. Darüber hinaus sprechen CLIL-Lerner flüssiger als Nicht-CLIL-Lerner. Außerdem benutzen sie ihre Muttersprache nur selten. Die Ergebnisse sind bezeichnend für die positiven Effekte von CLIL auf die mündliche Kompetenz, wenn beteiligte Variablen wie Dauer des Fremdsprachenlernes oder Alter kontrolliert werden. Auch wird besonders das Fehlen positiver Effekte von CLIL auf die Aussprache diskutiert.

Schlüsselwörter: CLIL, mündliche Kompetenz, Englisch als Fremdsprache, Dauer des Fremdsprachenlernens

1 Introduction

The present study aims at exploring the effects of Content and Language Integrated Learning (CLIL) on oral production skills. CLIL can be defined as "an educational approach where curricular content is taught through the medium of a foreign language" (Dalton-Puffer 2011: 183), and where both content and language play an equal role, though it has recently been acknowledged that the integration of these two elements might be far from being optimal in praxis (Dalton-Puffer 2013). As for the language component, the CLIL Compendium (http://clileduca tion.blogspot.com.es/) specifies that developing favourable oral communication skills is one of its purported goals. Furthermore, oral production has been acknowledged to be one of the linguistic aspects which may benefit most from methods which foster the use of the language in meaningful contexts (Block 2003) -CLIL being one of these. However, to date few studies have observed the oral skills in CLIL contexts and some have highlighted the arbitrary results obtained so far regarding the analysis of the speaking skill (Van de Craen, Mondt, Allain and Gao 2007).

Divergent results could be ascribed to the fact that there might be intervening variables such as amount of exposure or age which may have not been addressed in CLIL studies (Bruton 2011). The work presented here is an attempt to manage these variables. We present the results of an investigation where (i) out-of-school exposure is nonexistent and (ii) CLIL learners are compared to both Non-CLIL students two grades ahead and exposure-matched peers. Participants (mean age: 15.9 years) were asked to narrate a story in English and their productions were holistically assessed for pronunciation, vocabulary, grammar, fluency and content. In addition, the amount and density of learners' productions as well as their compensatory strategies were analysed.

The structure of our work is as follows. We start with a literature review of those investigations which have explored the benefits of CLIL in oral production. This section finishes with the presentation of the research questions, which are followed by two blocks dedicated to the methodology of the study – one describing the characteristics of the participant sample and the other one detailing about the materials and procedures employed. Results are subsequently described and discussed, and we finish with a conclusion section which highlights the relevance of the findings and acknowledges the limitations of the study.

2 Literature review

It has been recently advocated that the alleged language benefits of CLIL instruction in language aspects such as vocabulary (Lo and Murphy 2010; Ruiz de Zarobe 2010) or different aspects of oral production (Admiraal, Westhoff and de Bot 2006; Huttner and Rieder-Bunemann 2010; Lasagabaster 2008; Maillat 2010; Moore 2009) could be compromised given that variables such as amount of exposure have not been sufficiently controlled for in research (Bruton 2011). Indeed, much conducted research has focused on CLIL vs. non-CLIL peer comparisons where CLIL students have received more exposure than non-CLIL ones precisely due to their participation in the CLIL programmes. Nevertheless, trials to control for amount of exposure have been made in some experiments. These have adopted different approaches. Some studies have compared CLIL students to older traditional students in higher grades in an attempt to equal the number of instruction hours (Lasagabaster 2008), others have monitored effects longitudinally by including a pre-test (Navés and Victori 2010; Rallo Fabra and Juan Garau 2010) and authors such as Villarreal Olaizola (2011) have incorporated correlation analyses between amount of exposure (out-of-school lessons) and language outcomes in their data analysis procedure so as to explore whether this variable interacts with the language aspect under analysis. In general, these works have also evinced some advantage in the aspects analysed on the part of the learners undergoing the CLIL programmes, this being perhaps more modest as sometimes significance was not reached or not all aspects under study were boosted similarly.

We shall now review those studies which have analysed the effects of CLIL on oral skills and in which amount of exposure was an important factor. In the Basque Country (Spain) we find several investigations with this design undertaken by the LASLAB (Language and Speech Laboratory) Research Group. Lasagabaster (2008) controlled amount of exposure by comparing general competence of three groups – a CLIL group in Secondary Grade 3 (aged 14–15), a CLIL group in Secondary Grade 4 (aged 15–16) and a traditional English as foreign language (EFL) group in Secondary Grade 4. The study analysed speaking, writing, listening, and grammar variables for the three groups. The speaking skill was holistically assessed by using the following 5 criteria: pronunciation, vocabulary, grammar, fluency and content. The study revealed that both CLIL groups, the one matched for age (but not exposure) and the one matched for exposure (but not for age), scored higher than the Non-CLIL group in the speaking skill. Although the advantage did not reach statistical significance for the speaking skill, it did so for overall general competence, a result which the author described as positive, interpreting it as younger CLIL learners being able to catch up with one year older Non-CLIL peers.

Ruiz de Zarobe (2008) analysed the effects of two different CLIL exposure levels on spoken production by collecting and comparing data from three different groups which differed in amount of exposure (a traditional EFL group, a 1-subject-CLIL group, and a 2-subject CLIL group) at three different levels: Secondary Grade 3, Secondary Grade 4 and Baccalaureate. The design also analysed oral production holistically by using the same five scales in Lasagabaster's study (2008): pronunciation, vocabulary, grammar, fluency, and content. The study showed that, when analyzing the whole sample, the more exposure, the better the oral outcomes, as the traditional group scored lower than the 1-subject CLIL group, which in turn, scored lower than the 2-subject CLIL group. When oral skills were analysed by school year, the CLIL groups tended to outperform the traditional group in the three grades analysed. Nevertheless, the author did not statistically account for differences between the two CLIL groups, which differed in amount of exposure.

Finally, Gallardo del Puerto and Gómez Lacabex (2013) also controlled for out-of-school exposure focusing on oral skills development in CLIL with similar gauging variables in addition to age range (3rd and 4th graders) as the two studies previously reported. Their analyses indicated that Secondary Grade 3 and 4 CLIL learners' productions were holistically perceived to exhibit better fluency, lexis and grammar than those of Non-CLIL peers while no differences were found as regards content and pronunciation. Besides, although Non-CLIL learners' productions were greater in quantity and longer in time, CLIL learners produced denser and more fluent narrations. Additionally, CLIL learners resorted to their first language to a lesser extent and demanded fewer vocabulary clarifications.

The GRAL (Research Group On Language Acquisition) Group in Catalonia (Spain) has also looked into CLIL overall proficiency achievement (see Navés and Victori 2010 for a review article) in a design which compared groups of four different age ranges (n=837 students) and which included oral perceptive skills, listening and dictation in a proficiency test. Interestingly, these studies revealed that the CLIL approach seems to benefit from cognitive maturation and that, the higher the grades, the more it may be ready to boost learner's language proficiency. They found that, first, the CLIL group tended to outperform the traditional group in each testing age (11, 13 14 and 15 years old) and that, second, as age increased, the CLIL group's proficiency skills caught up with those of the one-year-ahead group and were able to eventually outperform them.

The previous studies have quantitatively analysed oral production skills from a holistic and analytic perspective. To date, very few studies have focused on more qualitative views or more specific conditions such as pronunciation, an aspect on which CLIL has been reported to have little impact (Dalton Puffer 2008; Ruiz de Zarobe 2011, 2015). Gallardo del Puerto, Gómez Lacabex and García

Lecumberri (2009) compared the degree of foreign accent of students that learned English through traditional classroom instruction with those learning in CLIL environments, in a sample in which out-of-school exposure was also eliminated. Additionally, they tested the communicative effects of foreign accent, specifically the intelligibility and irritation produced by learners' non-native speech. They concluded that despite the fact that no differences in terms of degree of foreign accent were discovered, the CLIL students' oral narrations were judged to be more intelligible and less irritating than those of the students receiving a traditional EFL approach.

Along the same lines, Rallo Fabra and Juan-Garau (2010) conducted a study in which intelligibility and accentedness differences between CLIL and traditional foreign language (FL) students were also explored, this time longitudinally. This study analysed differences between the two groups over a year and also added a comparison to a group of English monolingual speakers of the same age. Results in a reading aloud task also showed that CLIL students were more intelligible than the FL ones and that differences in accentedness were slight. Interestingly, no differences between the two testing times (1 year apart) were found in the CLIL group, indicating that one year of CLIL instruction may not be sufficient to improve aspects such as intelligibility or accentedness. The authors also suggest that, in fact, these aspects may not improve unless they are given specific attention.

Finally, a recent analytic account of CLIL effects on pronunciation has been provided in Rallo Fabra and Jacob (2015). The authors examined fluency and number of vowel errors in a CLIL group and a traditional EFL group (14–15 years old) at two times: when the CLIL programme started and after the CLIL programme had run for two years. Quantitative data including several speech rate measures or silent pauses per minute for fluency were analysed while an experienced native speaker coded the quality of the vowels produced in a read-aloud task and accurate matches were sent for analysis. The authors did not find significant differences in the fluency of the story-telling task or in the rate of vowel errors in the read-aloud task between the CLIL group and the traditional EFL group after two years of CLIL instruction for the former and traditional language instruction for the latter. More precisely, both groups improved in their fluency rates to a similar extent, while they did not experience amelioration in the quality of their vowels between testing times. In line with what had previously been estimated, these novel results, which await replication and further research extension to other pronunciation aspects and populations, seem to indicate that the pronunciation aspect is unlikely to improve in language learning contexts in which intensity of exposure and opportunities for production are favoured but where quality of exposure is still limited. Also, pronunciation may not benefit

from contexts in which instructors are not native speakers of the target language or from learning/teaching contexts in which the written form still eclipses the oral form, as the authors suggest.

Research on the impact of CLIL on communicative oral skills such as conversational interaction and negotiation skills should also be mentioned given that the CLIL methodology is likely to promote classroom talk and more communicative-oriented tasks than traditional language lessons (Llinares, Morton and Whittaker 2012). Much research has been conducted on the teacher-learner interaction mode. Some of this research points to the fact that teachers do not always provide negative feedback (Mariotti 2007). Milla Melero and García Mayo (2014) also found that practitioners do not facilitate corrective feedback in the CLIL classroom more often than in the EFL classroom, which, as pointed out by the authors, may result in insufficient opportunities for producing comprehensible output or for feeding learners' uptake (students' responses to corrections, as in Lyster and Ranta 1997), both regarded as relevant L2 learning processes in classroom settings.

More recent research has investigated the learner-learner interaction mode. Mesquida and Juan-Garau (2013) designed a comparative study in which a traditional EFL group was compared to an age-matched CLIL group with additional exposure. This study controlled for out-of school exposure by ruling out private tuition, visits to English speaking countries or English speaking environments. Negotiation strategies such as self-repairs, clarification requests or confirmation checks were analysed. The CLIL group activated more frequent and more varied negotiation strategies than the Non-CLIL group, which, despite not reaching statistical significance, the authors interpreted as evidence of CLIL being able to "complement formal language teaching" (2013:47). García Mayo and Lázaro Ibarrola (2015) have also recently explored negotiation for meaning in another EFL-CLIL comparative study with younger learners, in which the CLIL learners were able to negotiate more and to access their L1 less frequently than the EFL learners.

3 Research Questions

Given that oral production in CLIL is still in need of further examination as inconclusive results have been reported (Rallo Fabra and Jacob 2015; Van de Craen et al. 2007) and that the time/exposure advantage has been less monitored in previous studies (Bruton 2011), the present study intended to look into the effect of CLIL on oral production skills when amount of exposure was monitored. We present the results of an investigation with secondary English learners where out-of-school exposure is non-existent and CLIL learners are compared to non-CLIL students with a similar amount of exposure. We matched them with i) a

group of two-grade-ahead students (FL+2) who started learning English at the same age and ii) a group of same-age students (FLpeer) who started learning English at an earlier age (see Martínez Adrián and Gutiérrez Mangado 2015 for a study on grammar with a similar sampling design). We entertained the following research questions, which aim at addressing both a holistic and an analytic account of results:

3.1 *Do CLIL and Non-CLIL learners' oral skills differ when holistically analyzed?*

3.2 *Are there any differences between CLIL and Non-CLIL learners as far as the amount and density of oral production are concerned?*

3.3 *Do CLIL and Non-CLIL learners use compensatory strategies in oral production differently?*

4 The study

4.1 Participants

The participants in this study were 48 Basque-Spanish bilingual school children who had been exposed to English exclusively at school. They came from three different schools located in middle-class neighbourhoods, and following similar educational programmes. They were divided into three different groups, a CLIL group and two non-CLIL ones: FLpeer, a group matched for age and exposure, and FL+2, a two-year older group also matched for exposure. Those learners who had received extra-curricular lessons in English (private academies, summer camps, etc.) were excluded from our sample. We were able to monitor participants' attitudes towards the English language and motivation towards learning English. They were administered two different Likert-scale questionnaires which revealed that the older group (FL+2) showed slightly higher rates than the two younger learner groups (FLpeer and CLIL). We did so given that recent research has discovered a deleterious effect of CLIL on primary school learners' motivation (Fernández Fontecha and Canga Alonso 2014). However, in our sample, differences did not turn out to be statistically significant when t-tests were computed for CLIL vs. FLpeer ($p > .05$) and CLIL vs. FL+2 ($p > .05$) comparisons, revealing that the groups exhibited similar attitudinal and motivational profiles towards the English language and English language learning. Table 1 displays the characteristics of each of these groups, namely testing age, exposure, onset age, and attitudinal and motivational profile towards the English language.

Table 1: Characteristics of the Sample.

Group	N	Age	Exposure (hours)	Onsetage	Attitudes% (SD)	Motivation% (SD)
CLIL	17	15;4	1155	8	69.1(14.2)	68.7(9.2)
FLpeer	19	15;4	1148	4	62.8(18.3)	73.1(15.9)
FL+2	12	17;5	990	8	74.7(17.7)	81.9(14.4)

Subjects were being instructed in Basque (the minority language of the community). Spanish (the majority language in the Basque Country) was just a school subject to which 4 hours per week were devoted. English (a language with a foreign status in Spain) was just a school subject for the FLpeer and FL+2 groups. These groups devoted 3 hours per week to it. In addition to these FL hours, the CLIL group had been receiving 3 hours of content-based English instruction per week for 4 years from the beginning of secondary school (age 12) in subjects such as science, biology and geography/history, which made up about 400 hours of additional CLIL exposure overall. The CLIL programme they participated in was not elective but administered to all secondary education students. CLIL teachers, like Non-CLIL ones, were non-native speakers of English. The minimum requirement to become a CLIL teacher in these programmes was a B2 level of English according to the Common European Framework of Reference for Languages.

Regarding the onset age of English learning, the CLIL and the FL+2 groups had started learning English when they were 8 years old. However, these two groups differed in terms of their testing age, FL+2 students being two years older at testing time than the CLIL participants. The former had received an inferior amount of English exposure, namely 14% less exposure, than the latter, but they were the oldest group available for comparison, being in their last year of non-compulsory secondary education. Participants older than those had already left the school. The FLpeer students shared their testing age with the CLIL group (15;4) but they had started learning English 4 years earlier (age 4). Hence, in this comparative analysis we were able to control for cognitive development at the time of testing.

4.2 Materials

The instrument employed to analyse participants' spoken productions in English was a story-telling activity in which students were individually presented a

series of wordless black and white vignettes narrating the story of a frog (*Frog, where are you?* by Mayer 1969). Students had to look at the pictures and tell the interviewer the story in English. Stories were analysed holistically and also analytically.

As regards the holistic analysis, two-minute excerpts were extracted from participants' productions, randomised and presented to listeners so that they could judge learners' productions with regard to *Pronunciation* (how accented/ intelligible the speech was), *Vocabulary* (richness and right choice of lexicon), *Grammar* (degree of syntax accuracy), *Fluency* (how smooth speech was, number of pauses and hesitations) and *Content* development (how cogent and elaborated the narration was). Listeners were two trained judges, one male and one female, aged 30–35 who were proficient in English, experienced teachers of English, and had received previous linguistic training as former students of Linguistics and English Studies. They were postgraduates in the English Language Department at the University of the Basque Country in Spain, and were specialised in the acquisition of English as a foreign language. Furthermore, they had considerable experience of the holistic assessment of the story telling activity used in our study. They lived in the Basque Country and were Spanish monolinguals with a good knowledge of Basque.

As far as the analytic assessment was concerned, a series of variables were computed in CLAN (http://childes.psy.cmu.edu/clan/) after the productions were transcribed and coded. These variables were assorted in three groups according to (i) how much output learners produced (amount of production), (ii) how rich and compact/ed the output was (density of production) and iii) how much students had to resort to their first language/s or the interviewer when telling the story in English (compensatory strategies). Regarding amount of production, four counts were provided: the mean of the total number of words which the learners in each group used in the narrations (*Total no. of words),* the total number of words excluding borrowings or Spanish/Basque words (*Total no. L2 words),* and a mean count for the different words used by each learner *(Total no. different words).* We also provided counts for the number of utterances (*Total no. utterances*), coded as speech sequences (one or more words) which are preceded and followed by silence and/or prosodically signaled, and the number of turns (*Total no. turns),* which indicate change of speaker or a new utterance. Density of production was rendered by means of calculating the number of different words over the number of words with two lexical richness indices: *No. different words/ No. words* and *D-index*. The latter was included for controlling variations in the size of the narrations (see McCarthy and Scott 2007, 2010). We also included the number of utterances over number of turns *(No. utterances/turn)* and the number of words over number of turns *(No. words/turn)*. Finally, compensatory strategies

were operationalised by counting the number of Spanish/Basque words in students' productions (*L1/s transfer*) and their appeals for assistance, that is, the times the interviewer was required to answer learners' questions on lexicon (*Interviewer turns*).

4.3 Results

The results of our analyses will be structured as follows. Firstly, the results of the comparison between CLIL learners and non-CLIL learners two years ahead (FL+2) will be presented in section 4.3.1 for holistic assessment (Table 2), amount of production (Table 3), density of production (Table 4) and compensatory strategies (Table 5). In the next section (4.3.2), we will display the results of the comparison between CLIL learners and their non-CLIL peer counterparts (FLpeer) who had started learning English 4 years earlier in the same order as above: the holistic assessment in Table 6, amount of production in Table 7, density of production in Table 8 and compensatory strategies in Table 9.

Samples were normally distributed and hence t-tests were used for the paired comparisons. The confidence interval was set at 95%. A first analysis explored the reliability of the holistic assessment by correlating the data provided by the judges for the three groups. Moderate correlational indices were found for most variables assessed – *Vocabulary*: $r(48) = .60$, $p < .0001$; *Fluency*: $r(48) = .62$, $p < .0001$; *Content*: $r(48) = .49$, $p < .0001$-, except for *Grammar*: $r(48) = .37$, $p < .05$, with a slightly lower value, and *Pronunciation* $r(48) = -.05$, $p > .05$.

4.3.1 CLIL vs FL+2 Comparison

A t-test analysis was computed so as to establish comparisons between the means (range: 1–10) of the 5 variables assessed by the judges in the holistic evaluation of oral productions by the CLIL group and the two-grade-ahead matched-for-exposure group (FL+2). Judgements were moderately correlated at $r(36) = .59$, $p < .0001$ for the mean of all five variables. As we can observe in Table 2, no statistical significant differences were found, which indicates that both groups were perceived to perform similarly in all the oral speech dimensions analysed.

Table 2: Listeners' holistic assessment for CLIL vs. FL+2 comparison.

	CLIL 1		FL+2		*Significance*
	Mean	*(SD)*	**Mean**	*(SD)*	
Pronunciation	6.1	*0.7*	6.6	*0.8*	>.05
Vocabulary	7.2	*0.9*	6.6	*0.9*	>.05
Grammar	6.5	*0.8*	6.9	*1.0*	>.05
Fluency	6.9	*0.8*	6.6	*1.1*	>.05
Content	7	*0.8*	7.3	*0.9*	>.05
Total	6.7	*0.6*	6.8	*0.9*	>.05

The t-test computed in the frequency evaluation revealed more abundant productions on the part of the FL+2 group in all the variables analysed (see Table 3). Some of these comparisons yielded highly significant differences, as in the case of *Total no. utterances* (t(26) = -4.48, *p* < .0001, 95% CI = -29.1,-10.9) and *Total no. turns* (*t*(26) = -3.16, *p* < .005, 95% CI = -10.4,-2.2). The same trend was observed for the *Total no. different words* variable, which nearly reached statistical significance (*t*(26) = -1.95, *p* = .06, 95% CI = -43.3, 0.6).

Table 3: Amount of production for CLIL vs. FL+2 comparison.

	CLIL 1		FL+2		*Significance*
	Mean	*(SD)*	**Mean**	*(SD)*	
Total no. words	343.1	*169.1*	402.4	*135.6*	>.05
Total no. L2 words	334.4	*163.7*	378.7	*140.2*	>.05
Total no. different words	95.4	*30.6*	116.8	*26.6*	>.05
Total no. utterances	30.4	*12.5*	50.4	*10.8*	<.0001
Total no. turns	5.2	*4.2*	11.5	*6.6*	<.005

Those variables which measured density of production (Table 4) revealed that the CLIL group produced longer and richer narrations as shown by the variables *No. different words/No. words, D-index No. utterances/turn* and *No. words/turn*. This difference became highly significant in the last case (*t*(26) = 3.43, *p* < .005, 95% CI = 139.9, 557.9) and was marginally significant in *No. utterances/turn* measure (*t*(26) = -1.81, *p* = .08, 95% CI = -32.7, 519.2).

Table 4: Density of production for CLIL vs. FL+2 comparison.

	CLIL		FL+2		*Significance*
	Mean	*(SD)*	**Mean**	*(SD)*	
No. different words/No. words	0.3	*0.1*	0.03	*0.0*	>.05
D-index	25.5	*7.1*	28.9	*4.9*	>.05
No. utterances/ turn	8.0	*4.0*	5.5	*2.7*	>.05
No. words/turn	11.4	*3.3*	7.9	*1.5*	< .005

As for compensatory strategies (Table 5), it was observed that the CLIL group had recourse to the L1/s significantly less than the FL+2 group *(t*(26) = 7.75, *p* < .05, 95% CI = -27.4,-2.6). It was also observed that interviewers interacted significantly less with the CLIL group *(t*(25) = -2.97, *p* < .05, 95% CI = -10.3,-1.87), an indicator that these subjects demanded lexical clarifications less often than the FL+2 group.

Table 5: Compensatory strategies for CLIL vs. FL+2 comparison.

	CLIL		FL+2		*Significance*
	Mean	*(SD)*	**Mean**	*(SD)*	
L1/s transfer	8.7	*10.1*	23.8	*21.9*	<. 05
Interviewer turns	4.4	*4.2*	10.5	*6.6*	<. 05

4.3.2 CLIL vs FLpeer Comparison

To come to the listeners' evaluation, a t-test was computed with the mean scores of the two judges for the five variables analysed in the holistic assessment of oral narrations by CLIL learners and their Non-CLIL exposure-matched peer counterparts. Judgements were moderately correlated at r(29) = .41, p < .05 for the mean of all five variables. As can be seen in Table 6, the CLIL group outscored the FLpeer group in *Vocabulary (t*(33) = 5.02, *p* < .0001, 95% CI = 1.03, 2.43), *Grammar (t*(33) = 3.39, *p* < .005, 95% CI = .42, 1.68), *Fluency (t*(33) = 4.53, *p* < .0001, 95% CI = .777, 2.04) and *Content (t*(33) = 3.21, *p* < .005, 95% CI = .33, 1.50). However, *Pronunciation* ratings did not distinguish both groups significantly.

Table 6: Listeners' holistic assessment for CLIL vs. FLpeer comparison.

	CLIL 1		FLpeer		Significance
	Mean	*(SD)*	**Mean**	*(SD)*	
Pronunciation	6.1	*0.7*	5.7	*0.7*	>.05
Vocabulary	7.2	*0.9*	5.5	*1.1*	<.0001
Grammar	6.5	*0.8*	5.5	*1.0*	<.005
Fluency	6.9	*0.8*	5.5	*1.1*	<.0001
Content	7.0	*0.8*	6.1	*0.9*	<.005
Total	6.7	*0.6*	5.6	*0.9*	<.005

The t-test computed in the amount-of-production evaluation did not reveal many statistical differences between the age-and-exposure matched groups, as shown in Table 7. Only one significant difference was found (*Total no. turns*) in favour of the FLpeer group (*t*(33) = 7.75, *p* < .05, 95% CI = -27.4,-2.6), indicating that the FLpeers initiated more turns than the CLIL group. This contrasts with the nearly significant difference in favour of the CLIL group for the *Total no. L2 words*- (*t*(33) = 2.0, *p* = .05, 95% CI = -1.06, 178.4), which indicated that, despite initiating fewer turns, the CLIL learners tended to produce more English words than the FLpeers.

Table 7: Amount of production for CLIL vs. FLpeer comparison.

	CLIL 1		FLpeer		Significance
	Mean	*(SD)*	**Mean**	*(SD)*	
Total no. words	343.1	*169.1*	273.4	*110*	>.05
Total no. L2 words	334.4	*163.7*	245.7	*96.9*	>.05
Total no. different words	95.4	*30.6*	94.2	*27.0*	>.05
Total no. utterances	30.4	*12.5*	36.3	*15.5*	>.05
Total no. turns	5.18	*4.2*	12.3	*7.2*	< .005

This tendency was confirmed in the density of production analysis (Table 8), in which t-tests revealed that the CLIL learners were able to produce turns with a significantly higher number of words (*t*(33) = 4.73, *p* < .0001, 95% CI = 217.5, 544.7) and utterances (*t*(33) = 4.88, *p* < .0001, 95% CI = 254.5, -677.4) than the FLpeer group.

Table 8: Density of production for CLIL vs. FLpeer comparison.

	CLIL		FLpeer		Significance
	Mean	(SD)	Mean	(SD)	
No. different words/No. words	0.3	0.1	0.4	0.1	>.05
D-index	25.5	7.1	28.7	7.4	>.05
No. utterances/turn	8	4.0	3.3	0.9	<.0001
No. words/turn	11.4	3.3	7.6	1.2	<.0001

The analyses of the compensatory strategies (Table 9) used by both groups revealed that CLIL learners resorted to the L1/s significantly less frequently (*L1/s transfer: (t*(33) = 2.64, *p* < .005, 95% CI = -30.6, -7.4) and demanded fewer lexical clarifications as measured by the number of *Interviewer turns (t*(32) = -3.36, *p* < .005, 95% CI = -10.9, -2.7).

Table 9: Compensatory strategies for CLIL vs. FLpeer comparison.

	CLIL		FLpeer		Significance
	Mean	(SD)	Mean	(SD)	
L1/s transfer	8.7	10.1	27.7	21.5	<.005
Interviewer turns	4.4	4.2	11.3	7.2	<.005

5 Discussion

The present study intended to explore the effect of CLIL as opposed to traditional classroom teaching on the oral production abilities of English learners in a study design in which all participants were matched for amount of exposure. The study compared a secondary CLIL group with, first, a group of two-year-ahead students who started learning English at the same age (FL+2) and, second, a group of same-age students who started learning English at an earlier age (FL peers).

As regards the first research question (*Do CLIL and Non-CLIL learners' oral skills differ when holistically analyzed?*), CLIL learners significantly outscored their age-and-exposure matched FLpeers in all variables (vocabulary, grammar, fluency, content development) with the exception of pronunciation. However, their performance was statistically similar to that of the FL two-year-ahead learners (FL +2 group). It can be inferred that, first, when amount of exposure is monitored, CLIL students are still judged to outperform non-CLIL peers, in accordance with

those previous comparative studies in which CLIL groups had received more exposure than mainstream FL groups (Gallardo del Puerto and Gómez Lacabex 2013; Jiménez Catalán and Ruiz de Zarobe 2009; Ruiz de Zarobe 2007). Second, the CLIL group exhibits a similar oral competence to that of a two-grade-ahead FL group. It could be said that CLIL boosts language outcomes similar to those that a traditional EFL group develops after two years of cognitive advance, a variable which has been found to exert a considerable influence in samples of exposure-matched school students of different ages where cognitively more mature, older learners consistently outperform less cognitively developed, younger learners (García Mayo and García Lecumberri 2003; Muñoz 2006). Despite the fact that it must be acknowledged that our FL+2 group had received 14% less exposure than the CLIL group, these tentative results go in line with studies such as Lasagabaster (2008) and the projects conducted by the GRAL research group in Catalonia (Navés and Victori 2010), which have provided evidence that CLIL learners are able to catch up with older Non-CLIL peers with regard to their oral skills.

It is interesting to note the point to a lack of effect of CLIL on the pronunciation variable, which was never boosted. CLIL does not seem to exert pronunciation benefits either when age and exposure are controlled or when exposure is controlled but not age (and thus cognitive maturation). This outcome is not surprising given that, first, instructors in CLIL programmes are mostly non-native speakers (Gallardo del Puerto, Gómez Lacabex and García Lecumberri 2009; Rallo Fabra & Jacob 2015) and hence intelligibility levels are high in teacher-learner and learner-learner classroom interactions. The so called "L1 match intelligibility benefit" (Bent and Bradlow 2003), by which both the group of learners and the instructor share the L1, is likely to discourage the instructor from addressing pronunciation issues which the learner may show when interacting with native speakers or with speakers of other languages. Second, given the orientation of CLIL lessons towards meaning and communication, a tendency to prioritize fluency over accuracy has been observed (Alonso, Grisaleña and Campo 2008) and in such teaching conditions aspects such as pronunciation may be under-appreciated. Thirdly, it has been shown that CLIL instructors do not activate corrective feedback as much as in traditional language sessions (Milla Melero and García Mayo 2014), a common correction resource in pronunciation teaching. Fourthly, language teachers are not always provided with sufficient training/ background on pronunciation teaching and phonetic knowledge (Derwing and Munro 2015), an issue which may be aggravated in the case of the content teacher, who may have not been provided with appropriate language training and for whom intelligibility, despite not being a learning goal in his/her classroom, should be a learning tool. Finally, we want to point out that in our study the pronunciation scale showed the weakest correlation amongst judges, despite

their common linguistic background. Indeed the perception of foreign accent is typically a rather varying parameter which it has been suggested may also be affected by non-phonetic factors such as fluency or grammar (Gallardo del Puerto, García Lecumberri and Gómez Lacabex 2015; Derwing 2013). Hence, the results and interpretations of the present paper still need to be supported by further pronunciation analysis by either including more judges or performing further testing which can better isolate and measure the pronunciation factor such as acoustic analyses, quantitative analyses of specific pronunciation elements or intelligibility and/or comprehensibility tests.

With regard to the second research question (*Are there any differences between CLIL and Non-CLIL learners as far as the amount and density of oral production are concerned?*), in the case of amount of production, we must say that this was the least informative variable given that few significant differences were found. Still, some considerations can be made. The variable revealed that despite the fact that the CLIL group produced i) more L2 words and more different words than the FLpeer group, and ii) fewer words, L2 words and different words than the FL+2 group, differences were not significant in either comparison. However, both Non-CLIL groups were found to produce significantly more turns (significant in both comparisons) and more utterances (significant for the CLIL vs. FL+2 comparison only) than the CLIL group. This indicated that CLIL turns may be expected to be more lexically abundant given that the same amount of lexicon was presented in a significantly smaller number of turns. This was confirmed in the density of production measures, which revealed that the CLIL students were able to produce richer turns than students in the mainstream FL groups. CLIL learners' turns contained both a larger number of words and a larger number of utterances than the turns produced by the students in the two Non-CLIL groups. This is indicative of CLIL learners' better capacity to produce denser, more compact and synthetic oral narrations when amount of exposure and/or testing age are better controlled, which confirms the tendency found in previous studies that have not monitored these variables (e.g., Gallardo del Puerto and Gómez Lacabex 2013; Lasagabaster 2008, Ruiz de Zarobe 2008). Nevertheless, no inter-group differences were found when the number of different words over the number of words was computed, in the two indices used (*No. different words*/*No. words* and *D-index)* which exhibit lexical affluence. The fact that the production task was not extemporaneous, which would have allowed the learners to resort to their English lexicon in a broader sense, may have contributed to this lack of differences. Also, it shall be considered that the two variables in this computation include L1 transfer items, which occurred significantly more often in the Non-CLIL groups.

Regarding the third research question (*Do CLIL and Non-CLIL learners use compensatory strategies in oral production differently?*) both Non-CLIL groups

were found to transfer a significantly higher number of L1 words and to demand more lexical clarifications from the interviewer than CLIL learners. In other words, traditional FL learners need to compensate for the lack of lexical knowledge in English to a significantly larger extent than CLIL learners when narrating a story orally. These findings mirror previous results from research where CLIL learners, in comparison to Non-CLIL ones, had received additional exposure (Gallardo del Puerto and Gómez Lacabex 2013). Hence, it can be conceded that it is CLIL, and not exposure or age (cognitive development) that leads to a lesser reliance on L1-based knowledge and on the interviewer, that is, to a decrease in the use of negotiation and repair strategies characterising foreigner talk (Gass and Varonis 1991). In other words, CLIL makes more independent storytellers.

6 Conclusions and limitations

The main aim of the present study was to better tackle the limitations shown by most previous research as regards the ways in which the variables amount of exposure and testing age have been handled. When testing age was held constant there happened to be a typical mismatch in the number of hours of exposure received by CLIL vs. Non-CLIL participants, CLIL learners having received more exposure as a consequence of their participation in CLIL programmes, a fact which may have masked the effects of CLIL in its own right in these studies. On the contrary, when amount of exposure was more alike in research comparing CLIL and traditional FL groups, the latter happened to be older and thus at a higher cognitive developmental stage, a fact which might have played against CLIL participants in these investigations (Lasagabaster 2008; Ruiz de Zarobe 2008). So as to better disentangle the confounding effects of both exposure and age variables in CLIL research, our study presented data from a CLIL group when compared to both an exposure-matched two-year-ahead Non-CLIL group and an exposure-matched peer Non-CLIL group. This sampling design permitted us to better tease the effects of exposure, age and CLIL apart.

Our study has provided evidence of the CLIL group's overall more efficient output than exposure-matched traditional FL peers with the exception of pronunciation. CLIL learners produced larger and denser oral narrations with better content, vocabulary, grammar and fluency, in addition to minimal use of the L1. Hence, the benefits of CLIL reported in the literature where CLIL participants have received more hours of instruction than Non-CLIL learners persist, as shown by our study, when CLIL and Non-CLIL peers receive the same amount of instruction. This finding can be read as indicative of the positive influence of the nature of the CLIL approach itself. Nevertheless observations of CLIL lessons would be desir-

able in order to find stronger supportive evidence for this suggestion. The data agree with the positive findings of research on the implementation of CLIL in Europe, where the target language, mainly English, is not natively spoken in the community and learnt in exclusively formal instructional settings (Bürgi 2007; Dalton-Puffer and Nikula, 2006; Huibregtse 2001; Jiménez Catalán, Ruiz de Za-robe and Cenoz 2006; Laitinen 2001; Nikula 2005; Sylvén 2004; Villarreal Olaizola and García Mayo 2009).

The comparison of CLIL learners to the two-grade-ahead FL group led to divergent outcomes as CLIL students were found to produce denser and less L1-reliant speech but their oral narrations were shorter and did not exhibit better content, fluency, vocabulary, grammar or pronunciation than those of older Non-CLIL learners. We have to acknowledge that, in this group comparison, the inferior amount of exposure received by the two-grade-ahead Non-CLIL group might also be a factor accounting for CLIL learners' better performance (denser and less L1-reliant speech), so our conclusions in this regard cannot be as robust as the ones from the peer-group comparison. As for the lack of differences in the holistic analysis, this finding can be read as regular FL learners' higher testing age (and thus higher maturational and cognitive stage) masking some of the purported beneficial effects of CLIL, a finding which would support a call for further research in which the effect of testing age is ruled out and at the same time the amount of exposure is held constant.

There are some limitations in the present study which should be highlighted. We have not controlled onset age, as CLIL learners had started learning English 4 years earlier than their Non-CLIL peers. In not doing so, our peer groups have come to differ in terms of rate of learning and intensity of exposure. First, it has been acknowledged that younger students have a slower learning rate. This is well documented in natural acquisition contexts (Snow and Hoefnagel/Höehle 1978) as well as in formal learning environments (García Mayo and García Lecumberri 2003; Muñoz 2006). Consequently, it can be conceded that the FLpeer group may be experiencing a slower learning rate than the CLIL group. Second, in this same group comparison, there may also be an intensity of exposure effect operating, as the CLIL group exhibited an intensity advantage over the FL peer group, having started learning English four years later than the FLpeer group. We are aware that immersion studies (Spada and Lightbown 1989; White and Turner 2005) have revealed that advantages may be due to intensity of exposure, some-thing which has been proved to be true in formal instructional settings too (Serrano 2010). We acknowledge that this factor might be operating in the present study.

As for the neutral effect of CLIL on the acquisition of phonology, we make a call for more research exploring the effects of explicit pronunciation training in

CLIL settings (see Gómez Lacabex and Gallardo del Puerto 2014), given the poor results of this language skill in formal (García Lecumberri and Gallardo del Puerto 2003) and CLIL (Gallardo del Puerto et al. 2009; Rallo Fabra and Juan-Garau 2010; Rallo Fabra and Jacob 2015) instruction settings and the recent research routes of explicit pronunciation training that are currently being explored (Saito 2012; Thomson and Derwing 2015).

All in all, our paper attempted to contribute further to the understanding of oral production outcomes in CLIL by carrying out intergroup comparisons so as to minimize the shortcomings present in previous CLIL vs. Non-CLIL comparative studies where CLIL learners had received a considerably higher number of hours of exposure, precisely as a consequence of their participation in CLIL programs. More specifically, the oral production of three matched groups of students was compared: a CLIL group, a Non-CLIL group with students of the same age, and a second non-CLIL group with students who are two years older. The results indicated advantages for CLIL students, whose productions were judged to be richer and denser in some of the holistic and analytic variables computed, particularly when they were compared to their Non-CLIL peers and testing age was controlled. As for their comparison with the two-year- ahead Non-CLIL group, where fewer significant differences were detected, it may be conceded that Non-CLIL learners' older age, and thus higher cognitive development and test taking abilities, contributes to mask some of the positive effects of CLIL. We must also acknowledge that our study could not fully disentangle intensity of exposure and the nature of the CLIL program. Yet, it seems to point to the fact that CLIL learners were able to produce more compact and synthetic narrations while they did not exceed their Non-CLIL peers in aspects such as pronunciation.

References

Admiraal, Wilfried, Gerard Westhoff & C. Kees de Bot. 2006. Evaluation of bilingual secondary education in the Netherlands: Students language proficieny in English. *Educational research and Evaluation* 12(1). 75–93.

Alonso, Esmeralda, Jesus Grisaleña & Alejandro Campo. 2008. Plurilingual education in secondary school: Analysis of results. *International CLIL Reserch Journal* 1(1). 36–49.

Bent, Tessa & Ann R. Bradlow. 2003. The interlanguage speech intelligibility benefit. *Journal of the acosutical Society of America*, 114(3). 1600–1610.

Block, David. 2003. *The social turn in second language acquisition*. Edinburgh: Edinburgh University Press.

Bruton, Anthony. 2011. Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System* 39. 523–532.

Bürgi, Heidi. 2007. *Im Sprachbad. Besseres Englisch durch Immersion. Eine Evaluation zweisprachiger Ausbildungsgänge an drei kantonalen Gymnasien in der Schweiz*. Bern: H.E.P.

Dalton-Puffer, Christiane. 2008. Outcomes and processes in content and language integrated learning (CLIL): current research from Europe. In Werner Delanoy and Laurenz Volkmann, (eds.), *Future perspectives for English language teaching*, 139–157. Heidelberg: Carl Winter.

Dalton-Puffer, Christiane. 2011. Content and language integrated learning: from practice to principles. *Annual Review of Applied Linguistic*s, 31. 182–204.

Dalton-Puffer, Christiane. 2013. A construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics* 1(2). 1–38.

Dalton-Puffer, Christiane & Tarja Nikula (eds). 2006. Current research on CLIL. Special issue of *VIEWS Vienna English Working Papers* 15(3).

Dalton-Puffer, Christiane, Tarja Nikula & Ute Smit (eds). 2010. *Language use and language learning in CLIL classrooms*. Amsterdam: John Benjamins.

Derwing, Tracey, M. 2013. Intelligibility in SLA. In Peter Robinson (ed.), *The Routledge encyclopedia of second language acquisition*. New York: Routledge.

Derwing, Tracey, M., & Murray J. Munro. (2015). *Pronunciation Fundamentals: Evidence-based perspectives for L2 teaching and research* (Vol. 42). John Benjamins Publishing Company.

Fernández Fontecha, Almudena & Andrés Canga Alonso. 2014. A preliminary study on motivation and gender in CLIL and non-CLIL types of instruction. *International Journal of English Studies* 14(1). 21–36.

Gallardo del Puerto, Francisco, Esther Gómez Lacabex & María L. García Lecumberri. 2009. Testing the effectiveness of content and language integrated learning in foreign language contexts: the assessment of English pronunciation. In Yolanda Ruiz de Zarobe & Rosa M. Jiménez Catalán (eds.), *Content and language integrated learning: Evidence from research in Europe*, 63–80. Bristol: Multilingual Matters.

Gallardo del Puerto, Francisco & Esther Gómez Lacabex. 2013. The Impact of Additional CLIL Exposure on Oral English. *Journal of English Studies* 11. 113–131.

Gallardo del Puerto, Francisco, María Luisa García Lecumberri & Esther Gómez Lacabex. 2015. The assessment of foreign accent and its communicative effects by naïve native judges vs. experienced non-native judges. *International Journal of Applied Linguistics*. 25(2). 202–224.

García Lecumberri, María Luisa & Francisco Gallardo del Puerto. 2003. English FL sounds in school learners of different ages. In María Pilar García Mayo & María Luisa García Lecumberri (eds), *Age and the acquisition of English as a foreign language*, 115–135. Clevedon: Multilingual Matters.

García Mayo, María Pilar & María Luisa García Lecumberri (eds). 2003. *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.

García Mayo, María Pilar & Amparo Lázaro Ibarrola. (2015). Do children negotiate for meaning in task-based interaction? Evidence from CLIL and EFL settings. *System* 54. 40–54.

Gass, Susan. M. & Evangeline M. Varonis. 1991. Miscommunication in nonnative speaker discourse. In Nikolas Coupland, Howard Giles & John M. Wiemann. (eds.), *Miscommunication and problematic talk*, 121–145. Newbury Park: Sage Publications.

Gómez Lacabex, Esther & Francisco Gallardo del Puerto. 2014. Two phonetic training procedures for young learners: Investigating instructional effects on perceptual awareness. *Canadian Modern Language Review*. 70(4). 500-531.

Huigbregtse, Ineke. 2001. *Effecten en Didaktiek van Tweetalig Voortgezet onderwijs in Nederland*. PhD Thesis, University of Utrecht.

Hüttner, Julia & Angelika Rieder-Bünemann. 2010. A cross-sectional analysis of oral narratives by children in CLIL and Non-CLIL instruction. In Christiane Dalton-Puffer, Tarja Nikula and Ute Smit (eds.), *Language use and language learning in CLIL*, 61–80. Amsterdam, the Netherlands: John Benjamins.

Jiménez Catalán, Rosa María, Yolanda Ruiz de Zarobe & Jasone Cenoz. 2006. Vocabulary profiles of English foreign language learners in English as a subject and as a vehicular language. *VIEWS: Vienna English Working Papers (Special Issue: Current Research on CLIL 1)* 15(3). 23–27.

Jiménez Catalán, Rosa María & Yolanda Ruiz de Zarobe. 2009. The Receptive Vocabulary of EFL learners in Two Instructional Contexts: CLIL vs. Non-CLIL Instruction. In Yolanda Ruiz de Zarobe and Rosa M. Jiménez Catalán (eds.), *Content and Language Integrated Learning: Evidence from Research in Europe*, 81–92. Bristol: Multilingual Matters.

Laitinen, Jaana. 2001. *English immersion in Finland – dreams or reality?* PhD Thesis, University of Jyväskyla.

Lasagabaster, David. 2008. Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal* 1. 31–42.

Lo, Yuen-Yi. & Victoria A. Murphy. 2010. Vocabulary knowledge and growth in immersion and regular language-learning programmes in Hong Kong. *Language and Education,* 24. 215–238.

Llinares, Ana, Tom Morton & Rachel Whittaker. 2012. *The Roles of language in CLIL*. Cambridge: Cambridge University Press.

Lyster, Roy & Leila Ranta. 1997. Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition* 19. 37–66.

MacWhinney, Brian. 2012[2000]. *The CHILDES project: Tools for analyzing talk*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Maillat, Didier. 2010. The pragmatics of L2 in CLIL. In Christiane Dalton-Puffer, Tarja Nikula and Ute Smit (eds.), *Language use and language learning in CLIL,* 39–58. Amsterdam, the Netherlands: John Benjamins.

Mariotti, Cristina. 2006. Negotiated Interactions and Repair Patterns in CLIL Settings. In Christiane Dalton-Puffer & Tarja Nikula (eds), *Vienna English Working Papers (Views) Special Issue: Current Research on CLIL* 15(3). 33–41.

Martínez Adrián, María & Junkal Gutierrez Mangado. 2015. Is CLIL instruction beneficial in terms of general proficiency and specific areas of grammar? *Journal of Immersion and Content-Based Language Education* 3. 51–76.

Mayer, Mercer. 1969. *Frog, where are you?*. New York: Penguin.

McCarthy, Philip & Scott Jarvis. 2007. Vocd: A theoretical and empirical evaluation. *Language Testing* 24. 459–488.

McCarthy, Philip & Scott Jarvis. 2010. MTLD, Vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*. 381–392.

Mesquida, Francesca & María Juan Garau. 2013. CLIL Instruction and its effects on the development of negotiation strategies. *Miscelánea: A Journal of English and American Studies,* 47. 125–144.

Milla Melero, Ruth & María Pilar García Mayo 2014. Corrective feedback episodes in oral interaction: A comparison of a CLIL and an EFL classroom. *International Journal of English Studies* 14(1). 1–20

Moore, Pat. 2009. *On the emergence of L2 oracy in bilingual education: A comparative analysis of CLIL and mainstream learner talk* (Unpublished doctoral thesis). Universidad Pablo de Olavide, Sevilla, Spain.

Muñoz, Carmen (ed). 2006. *Age and rate of foreign language learning*. Clevedon: Multilingual Matters.

Navés, Teresa & Mia Victori 2010. CLIL in Catalonia: An overview of research studies. In David Lasagabaster & Yolanda Ruiz de Zarobe (eds.), *CLIL in Spain: Implementation, results and teacher training,* 30–54. NewCastle Upon Tyne: Cambridge Scholars Publishing.

Nikula, Tarja. 2005. English as an object and tool of study: interactional effects and pragmatic implications. *Linguistics and Education*, 16(1). 27–58.

Rallo Fabra, Lucrecia, & Karen Jacob. 2015. Does CLIL enhance oral skills? Fluency and pronunciation errors by Spanish-Catalan learners of English. In María Juan Garau & Joana Salazar-Noguera (eds.), *Content-based language learning in multilingual educational environments,* 163–177. Berlin: Springer.

Rallo Fabra, Lucrecia & María Juan Garau. 2010. Intelligibility and foreign accentedness in a context-and–language–integrated-learning (CLIL) setting. *New Sounds 2010: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech*. Poznan (Poland): Federal Adam Mickiewicz University.

Ruiz de Zarobe, Yolanda. 2007. CLIL in a bilingual community: Similarities and differences with the learning of English as a foreign language, 20–22. Paper presented at *CLIL Symposium*, Vienna.

Ruiz de Zarobe, Yolanda. 2008. CLIL and foreign language learning: A longitudinal study in the Basque Country. *International CLIL Research Journal 1*(1). 60–73.

Ruiz de Zarobe, Yolanda. 2010. Written production and CLIL: An empirical study. In Christiane Dalton-Puffer, Tarja Nikula, & Ute Smit (eds.), *Language use and language learning in CLIL classrooms*, 191–209. Amsterdam/Philadelphia: John Benjamins.

Ruiz de Zarobe, Yolanda. 2011. Which language competencies benefit from CLIL? An insight into Applied Linguistics Research. In Yolanda Ruiz de Zarobe, Juan Manuel Sierra & Francisco Gallardo del Puerto (eds), *Content and foreign language integrated learning: Contributions to multilingualism in European Contexts*, 129–153. Bern: Peter Lang.

Ruiz de Zarobe, Yolanda. 2015. The effects of implementing CLIL in Education. In María Juan Garau & Joana Salazar-Noguera (eds.), *Content-based language learning in multilingual educational environments,* 51–68. Berlin: Springer.

Saito, Kazuya. 2012. Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quaterly*, 46(4), 842–854.

Serrano, Raquel. 2010. Development of English language skills in oral production by adult students in intensive and regular EFL courses. *Vigo International Applied Linguistics (VIAL)* 10. 99–115.

Snow, Catherine E, & Marian Hoefnagel-Höhle. 1977. Age differences in the pronunciation of foreign sounds. *Language and Speech,* 20. 357–365.

Spada, Nina & Patsy M. Lightbown. 1989. Intensive ESL programmes in Quebec primary schools. *TESL Canada Journal*, 7(1). 11–32.

Sylvén, Liss Kerstin. 2004. *Teaching English or English teaching? On the effect of content and language integrated learning on Swedish leaerners' incidental vocabulary acquisition*. PhD thesis. Göteborg: Göteborg University.

Thomson, Ron I. & Tracey M. Derwing. (2015). The effectiveness of L2 pronunciation instruction. A narrative review. *Applied Linguistics*, 36(3). 326–344.

van de Craen, Piet, Katrien Mondt, Laure Allain and Ying Gao. 2007. Why and how CLIL works: An outline for a CLIL theory. *VIEWZ Vienna English Working Papers (Special Issue: Current Research on CLIL 2)* 16(3). 70–78.

Villarreal Olaizola, Izaskun. 2011. *Tense and agreement on the non-native English of Basque-Spanish bilinguals: Content and language integrated learning vs. English as a school subject learners*. Published PhD Thesis. Leioa: University of the Basque Country/Euskal Herriko Unibertsitatea.

Villarreal Olaizola, Izaskun & María Pilar García Mayo. 2009. Tense and agreement morphology in the interlanguage of Basque/Spanish bilinguals: CLIL vs. non-CLIL. *In* Yolanda Ruiz de Zarobe & Rosa María Jiménez Catalán (eds.), *Content and language integrated learning: Evidence from research in Europe*, 157–175. Bristol: Multilingual Matters.

White, Joanna L. & Carolyn E. Turner. 2005. Comparing children's oral ability in two ESL programs. *Canadian Modern Language Review* 61(4). 491–517.