

Does CLIL Enhance Oral Skills? Fluency and Pronunciation Errors by Spanish-Catalan Learners of English

Lucrecia Rallo Fabra and Karen Jacob

1 Introduction

1.1 *Fluency and L2 Learning*

The differences between native and non-native speakers of a language in terms of the speed of delivery and accent are widely acknowledged. Unlike the effortless nature of L1 speech, the production of L2 speech is more cognitively demanding thus affecting the speed of delivery. Kormos (2006) suggests that this difference can be explained by the interaction of different factors, including “poor knowledge of L2 lexis, syntax, morphology and phonology, attentional resources needed for suppressing L1 production procedures, and greater demands on self-monitoring” (p. 154). The lack of automaticity is thus responsible for a slower rate in the speech of L2 learners.

The terms *fluent* and *fluency* have different meanings depending on whether they are used as language testing instruments or as measurable variables in empirical studies of L2 speech learning. In L1 speech, Fillmore (1979) described a “fluent speaker”, quantitatively, as someone who fills time with talk—a non-stop talker—and, qualitatively, as someone whose speech is coherent, complex and dense. Often, when the term fluency is applied to L2 learners, it is used in a broad sense and, as such, it is often conflated with *proficiency* so that we refer to a speaker as being fluent when this speaker has a good overall command of the target language.

Other authors conceptualize fluency in a narrower sense. For instance, Lennon (1990) acknowledges that it is a language testing instrument, but that it is different from other criteria in oral language exams such as accuracy or appropriacy. He defines fluency as a purely performance phenomenon, “an impression on the listener’s part

L. Rallo Fabra (✉) • K. Jacob
Departament de Filologia Espanyola, Moderna i Clàssica,
Universitat de les Illes Balears (UIB), Palma de Mallorca, Spain
e-mail: lucrecia.rallo@uib.es

that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently” (p. 391). Schmidt (1992) claims that fluency is a primarily temporal phenomenon and refers to fluent speech as an “automatic procedural skill that does not require much attention or effort from the speaker, whereas non-fluent speech is effortful and requires a great deal of attention, resulting in non-fluent speakers exhibiting many hesitations and other manifestations of groping for words and attempting to combine them into utterances” (p. 358). Based on Levelt’s (1989) speech production model, Segalowitz (2010) claims that L2 speakers’ fluency has its origins in the formulator, in which lexical access, phonological short-term memory, and control of attention determine the final productions of the articulator.

The aforementioned authors provide accurate conceptualizations of fluency, but they fail to operationalize which aspects of the speech signal are relevant to distinguishing between fluent and non-fluent speakers. More recently, research has focused on determining the factors that contribute to listeners’ perceptions of L2 fluency. For instance, Derwing et al. (2004) measured L2 fluency in low-proficiency Mandarin speakers of English subjectively by means of listeners’ judgments and objectively with five temporal measures, including mean length of runs, self-repetitions, speech rate and a composite measure of “pruned syllables” in which all types of disfluency were removed. Through a series of regression analyses, they found that temporal measures, especially pruned syllables per second, were good predictors of subjective fluency, in that they accounted for listener judgments relatively well.

Similarly, Iwashita et al. (2008) found that speech rate, silent pause rate, and total pause time correlated with proficiency level, speech rate providing the strongest correlation. Cucchiariini et al. (2000) also reported strong correlations between listeners’ judgments of fluency and two objective measures, articulation rate and number of pauses. In a follow-up study (Cucchiariini et al. 2002), they found that speech rate and phonation time ratio were important correlates of fluency for beginners, whereas mean length of runs was an important correlate of fluency in spontaneous speech among intermediate learners.

An important contribution of some of these studies is that they used computer-based techniques that automatically detect silences and syllable nuclei to compute fluency measures without the need to transcribe the speech samples first. These techniques have facilitated the study of L2 speech corpora in formal instruction settings. Along this line, Toivola et al. (2010) investigated the developments in the temporal properties of L2 Finnish spoken by low-proficiency adult learners from different L1 backgrounds—Thai, Chinese, Russian and Vietnamese. They found changes in the articulation rate and the number and duration of pauses of the speech samples obtained at three data collection times over a 1-year period of observation.

Mora and Valls-Ferrer (2012) explored the effects of a study abroad (SA) period on the oral production skills of advanced-level Catalan-Spanish undergraduate learners of English. Speech samples elicited through an interview at three data collection times over a 2-year period were quantitatively assessed for fluency, accuracy and complexity. Fluency was measured through a complete battery of measures including, speech rate, articulation rate, phonation ratio, mean length of runs, dysfluency ratio, pause frequency and pause/time ratio. Overall, participants showed robust fluency gains during SA, moderate

improvement in accuracy, and lack of gains in complexity. No gains were found during the formal instruction period.

1.2 *Immersion Programs and Pronunciation*

The positive effects of immersion education on students' linguistic skills are well documented in the literature. For instance, Genesee (1987) reported that English-speaking children in early total French immersion programs in Canada scored as well as the control group on all scales of the oral production evaluation, including comprehension, pronunciation, grammar, vocabulary and communicativeness, in all grade levels. When tested on their L1 linguistic abilities, these students reached parity with their English control peers in listening comprehension, oral production or oral vocabulary skills, showing that immersion education was highly beneficial and it did not affect students' native language development. However, few studies have investigated L2 pronunciation in immersion settings. A study by Hammerly (1991) showed that early French immersion pupils, who entered the program at age five or six, still spoke French with an English accent 12 years later. Late immersion students who started immersion at puberty exhibited even stronger foreign accents.

More recently, Harada (2007) examined the pronunciation of Japanese voiceless stop consonants by English-speaking children aged 6–10 in a Japanese immersion program. He found that the immersion children produced the Japanese /p t k/ with significantly longer VOT¹ values than the monolingual Japanese children and the immersion teachers, but they produced them with significantly shorter VOT values than their English VOT. This suggested that the immersion students were able to distinguish between English and Japanese /p t k/ phonetically, though they implemented the VOT contrast in a non-target-like fashion, producing voiceless stops that were intermediate between English and Japanese.

The studies reviewed investigated total immersion settings in which the target language was introduced in kindergarten (early immersion) or in elementary school (delayed immersion). To our knowledge, only two studies have explored L2 pronunciation in late partial immersion settings, that is, when the onset of immersion is around puberty and the amount of L2 instruction is less than 50 % of the school curriculum. One of these studies was conducted by Gallardo del Puerto et al. (2009) in the Basque Country. Participants were two groups of Basque-Spanish students aged 14–16 with 6–7 years of English exposure through formal language instruction (FI). One group received extra English exposure by means of content and language integrated learning (CLIL). Pronunciation was assessed by inexperienced native English listeners, who rated excerpts from a story-telling task on the basis of foreign accent, foreign accent intelligibility and foreign accent irritation. Results showed

¹VOT (voice onset time) is defined as “the duration of the time interval by which the onset of periodic pulsing either precedes or follows release” (Lisker and Abramson 1964: 387). VOT has been used in many cross-linguistic studies that investigate the acquisition of L2 stops (see Amengual (2012) or Rallo Fabra (1998) for a review).

that CLIL students' pronunciation was rated as more intelligible and less irritating than the non-CLIL students' pronunciation. However, no significant differences in foreign accent ratings were found between the two groups of learners. The authors attributed these outcomes to the fact that the tutors in charge of CLIL instruction were not native English teachers and, as such, CLIL students were not exposed to authentic input as would have been desirable.

Rallo Fabra and Juan-Garau (2011) also investigated the effects of the CLIL program in the Balearic Islands on Spanish-Catalan learners' production. Four experienced native English listeners rated excerpts of read aloud speech for intelligibility and foreign accent. Overall, CLIL learner speech was perceived as more intelligible than non-CLIL speech, but all speech samples were heard by the listeners as equally accented, suggesting that CLIL instruction did not have an ameliorating effect on perceived foreign accent. They also noted that many pronunciation errors could have been the consequence of task effects. Since only read-aloud speech samples were used, many pronunciation errors may have been "spelling-induced" caused by a mismatch between English graphemes and phonemes.

1.3 The Role of Orthography on L2 Pronunciation

From a very early age, infants from different linguistic backgrounds can discriminate the differences between non-native speech sounds, regardless of the language. These language-universal perceptual abilities have been well documented in the literature (see Strange 1995 for a review). As early as the first year of life, speech perception becomes attuned to the speech sounds of the ambient language causing infants' perceptual abilities to decrease gradually. Burnham et al. (2002) claim that the onset of the orthographic period around the 6–8 age span strongly contributes to the decrease of speech perception. Reading instruction and phoneme-to-grapheme conversion rules have been hypothesized to negatively influence children's sensitiveness to non-native speech sounds. In a study with native English children, it was found that reading abilities were correlated with the attenuation of non-native speech perception. These findings were accounted for by contemplating the fact that, orthographically, English is considered an opaque language, in that there is not a one-to-one correspondence between graphemes and phonemes.

More recently, Erdener and Burnham (2005) investigated the effect of orthography on non-native speech production and writing by two groups of monolinguals, Australian-English (opaque orthography) and Turkish (transparent orthography). Participants were recorded performing various pronunciation tasks in different orthographic and audio-visual conditions, combining auditory information, visual information and orthography. The target non-words were taken from two languages varying in orthographic depth, namely, Spanish (transparent orthography) and Irish (opaque orthography). They found that orthography had a facilitating effect in pronunciation if the target language had a transparent orthography, such as Spanish. If the target language had an opaque orthography, as is the case of Irish or English,

orthography increased substantially the number of phonemic errors produced by the speakers. More importantly, the authors suggested that, at the first stages of exposure to a foreign language, orthographic input may not be beneficial to learners whose L1 has a transparent orthography, as in the case of Spanish or Catalan.

The role of orthography in L2 pronunciation by learners from different linguistic backgrounds has also been reviewed by Basetti (2009). In line with Erdener et al., she acknowledges that languages differ in terms of phonological transparency and that, in some cases, orthographic input may facilitate L2 production. However, orthographic input can also trigger non-target-like pronunciations, which might have been avoided if learners had only been exposed to auditory input. This was illustrated with the case of inexperienced Chinese speakers, who would pronounce the English spelling *ui* as /uei/, whereas Italian or Spanish speakers would say /ui/. These non-target-like pronunciations are the consequence of a mismatch between the L1 and L2 grapheme-phoneme conversion rules and question the facilitative role of orthography in classroom-based L2 learning.

2 The Present Study

This study examined oral language performance in two groups of EFL learners; a group of students who received CLIL instruction in English, plus the compulsory traditional English language classes, and another group who only received FI. Our primary aim was to investigate the impact of 2 years of CLIL instruction on students' oral performance in English. Two research questions were addressed:

1. Does CLIL instruction ameliorate students' fluency in English?

We addressed this question by measuring various temporal features of speech from a picture story task at two points in time, the onset of the participants' inclusion in the CLIL program (time 1) and an additional point about 2 years after the onset of the CLIL program (time 3).² The differences between CLIL and FI contexts as well as the development of fluency over time were analyzed statistically by means of two-way ANOVAs.

2. Does CLIL instruction reduce the number of vowel errors in English?

To answer this question, a native English speaker with phonemic transcription expertise transcribed a selection of vowel sounds taken from the reading aloud task at times 1 and 3. Differences between CLIL and FI students as well as time effects were examined statistically by means of two-way ANOVAs. Additional analyses were run to explore any possible effects of orthography on the pronunciation of English vowels by the two groups of learners.

²Participants were also recorded at the end of the academic year in which the CLIL program started (time 2). The language samples obtained at time 2 were analyzed for intelligibility and foreign accent in a previous study (Rallo Fabra and Juan-Garau 2011).

In the light of the findings of previous studies on L2 speech production in immersion and CLIL settings (Genesee 1987; Hammerly 1991; Harada 2007; Gallardo del Puerto et al. 2009) and other FI settings (Fullana 2006; Mora and Valls-Ferrer 2012), we predicted that CLIL students might exhibit modest gains in some fluency measurements, but they were unlikely to show gains in pronunciation measured as a percentage of target-like production of vowel sounds.

2.1 Method

2.1.1 Participants

Participants ($N=43$) were selected from two state-run secondary schools situated on the bilingual Spanish-Catalan island of Mallorca, in the Balearic Islands. All the students included in the analysis were from a Spanish-Catalan-speaking background. Data were collected as part of the COLE project, a state-funded project based in Catalonia and the Balearic Islands, Spain (see Juan-Garau and Salazar-Noguera 2015). For this chapter, data collected at T1 and T3 have been analyzed. The average age of the participants in both groups at T1 and T3 was 14 and 15 years respectively. All participants were included in the fluency analyses, however, a smaller subgroup was selected for the vowel error analysis (CLIL, $N=14$ and FI, $N=16$).

Group 1—a secondary school in Calvià, a touristic area on the south-west coast of the island—comprised 21 students who followed a combination of FI and CLIL instruction in English Language Teaching (ELT). They had 3 hours of FI per week along with the study of English in their social science class following a CLIL methodology. T1 corresponded to history and T3 corresponded to geography. Group 2 (FI), a secondary school situated inland in a village not too far from the capital of the island, Palma, comprised 22 students who followed a FI approach to ELT. This involved 3 h of instruction per week.

2.1.2 Speech Materials

The speech samples were obtained on the school premises. Each student was requested to perform two oral tasks, a guided task consisting in the reading aloud of a passage about the Atacama Desert in Chile, and a more extemporaneous task consisting in telling a story about a bank robbery. The tasks had been previously piloted with a group of 8 students from a third school to test whether the language level and the procedures were appropriate. The students were recorded in a quiet room by one of the members of the research group using an Olympus LS-10 Linear PCM recorder with a built-in microphone. They were allowed a few minutes to read the text and examine the picture story before they were recorded.

2.1.3 Measurement Procedures

Following the trends of recent research on L2 learner speech (Derwing et al. 2004; Kormos 2006; de Jong and Wempe 2009; Segalowitz 2010), the fluency measurements used in the present study combined temporal aspects of speech production with phonological error analysis. The temporal measurements were obtained from the extemporaneous speech samples (story-telling task) and they included four dependent variables:

1. Speech rate 1. This was calculated by dividing the total number of syllables by the total time required to produce the speech sample, including pauses, hesitations and fillers.
2. Speech rate 2. This was calculated by dividing the total number of English syllables by the total time required to produce the speech sample, including pauses, hesitations and fillers.
3. Speech rate 3. This was calculated by subtracting self-corrections, self-repetitions, false starts, non-lexical filled pauses and asides, from the total number of English syllables. The resulting “pruned syllables” were then divided by the total time required to produce the speech sample.
4. Silent pauses per minute. These were defined as the total number of pauses over 0.2 s divided by the total amount of time spent speaking.
5. Mean length of pauses. This was defined as the total length of pauses above 0.2 s divided by the total number of pauses above 0.2 s.

All sound files were imported and annotated with the *Praat* program (Boersma and Weenink 2013) and transcribed orthographically. Speech rate was measured with a *Praat* script that automatically detects syllable nuclei through peaks in intensity (dB) that are preceded and followed by dips in intensity (de Jong and Wempe 2009). The total number of syllables calculated by the script was subsequently checked by hand. Pauses were also detected automatically with the *Praat* “Textgrid to silence” option. The minimum silence interval duration was set at 0.2 s.

For the read-aloud speech, the target words were marked in a separate annotation tier. A native speaker of English experienced in EFL speech, listened to the recordings in counter-balanced order and coded vowel identity using the symbols of the International Phonetic Alphabet. Only the vowel errors of a selection of words from the reading aloud passage were transcribed (see Table 1).

Table 1 Target words used in the phonological error analysis

/i:/	/ɪ/	/u:/	/ʊ/	/e/	/ɜ:/	/ə/	/ɔ:/	/æ/	/ʌ/	/ɑ:/	/ɒ/
week	big	moon	look	best	earlier	amazed	water	happy	some	hard	fog
sea	building	food		very	word	shower	small	fact	nothing	last	top
									but	stars	

2.2 Results

The mean values and the standard deviations of the four fluency measurements are tallied in Table 2. At T1, the FI learners showed a slight advantage relative to their CLIL peers. In order to test whether this advantage was statistically significant, five independent-sample *t*-tests, assuming equal variance, were run. No significant differences were found in any of the three speech rate measurements: syllables/min (speech rate 1) [$t(41) = -0,814$ $p = .42$], English syllables/min (speech rate 2) [$t(41) = -1,2$ $p = .23$], or pruned syllables/min (speech rate 3) [$t(41) = -0,199$ $p = .32$]. In contrast, both learner groups differed significantly in terms of pauses/min [$t(41) = -3,19$ $p < .05$] and in the duration of the pauses [$t(40) = 2,27$ $p < .05$]. This calls for caution in interpreting learners' gains from T1 to T3.

2.2.1 Speech Rate

The mean speech rate measures obtained for each of the 44 subjects were submitted to three (2) Group and (2) Time two-way ANOVAs. No significant effects of group were found for the total number of syllables uttered in a minute (speech rate 1) [$F(1, 80) = 0.969$ $p = .328$], the number of English syllables per minute (speech rate 2) [$F(1, 80) = 2.02$ $p = .159$] or the number of pruned syllables per minute (speech rate 3) [$F(1, 80) = 2.59$ $p = .111$]. The main effect of time was significant for the three speech rate measures, speech rate 1 [$F(1, 80) = 14.553$ $p < .001$], speech rate 2 (English syllables only) [$F(1, 80) = 17.26$ $p < .001$] and speech rate 3 (pruned syllables) [$F(1, 80) = 15.13$ $p < .001$]. The Time \times Group interaction yielded no significant results for any of the speech rate measures, speech rate [$F(1, 80) = 0.056$ $p = .814$], speech rate 2 [$F(1, 80) = .10$ $p = .747$] or speech rate 3 [$F(1, 80) = 0.04$ $p = .835$].

2.2.2 Pauses

Overall, at T3, both learner groups paused more often than at T1. However, the length of the pauses was shorter. The mean number of silent pauses per minute and the mean duration of pauses were submitted to two additional two-way ANOVAs. A marginal effect of group was found for the number of silent pauses per minute [$F(1, 80) = 5.62$ $p < .05$] and also for the mean length of pauses [$F(1, 80) = 3.98$ $p < .05$].

Table 2 Mean fluency measures calculated for the CLIL and FI learner groups at times 1 and 3

Group	Time	Syllables/min	Syl/min (English)	Pruned syl/min	Pauses/min	Pause duration
CLIL	1	1.28 (0.711) ^a	0.98 (0.729)	0.73 (0.607)	0.38 (0.184)	1.867 (1.398)
	3	1.80 (0.655)	1.61 (0.753)	1.24 (0.806)	0.61 (0.081)	0.865 (0.387)
FI	1	1.43 (0.560)	1.23 (0.624)	0.93 (0.667)	0.52 (0.097)	1.163 (0.443)
	3	1.90 (0.427)	1.77 (0.478)	1.50 (0.450)	0.60 (0.133)	0.877 (0.255)

^aStandard deviations are in parentheses

The main effect of time was significant for both number of silent pauses per minute [$F(1, 80) = 28.71$ $p < .001$] and mean length of pauses [$F(1, 80) = 22.17$ $p < .001$]. Finally, marginally significant Group \times Time interactions were also found for silent pauses/min [$F(1, 80) = 6.88$ $p < .05$] and mean pause duration [$F(1, 80) = 4.43$ $p < .05$]. Four additional one-way ANOVAs examining the main effect of time on both groups of learners were run. The main effect of time was significant for both the CLIL group [silent pauses/min $F(1, 38) = 24.68$ $p < .001$; mean pause duration $F(1, 39) = 14.66$ $p < .001$] and the FI group [silent pauses/min $F(1, 42) = 4.97$ $p < .05$; mean pause duration $F(1, 42) = 6.909$ $p < .05$], but significance levels for the latter only reached the .05 alpha decision level, indicating that, after 2 years of CLIL instruction, learners progressed significantly faster than their FI peers, who made only modest progress after 2 years of FI.

2.2.3 Pronunciation of vowels

In order to quantitatively assess the pronunciation of the English vowels by both groups of learners, a variable labelled “accuracy” was created. If the pronunciation of the target vowel was correct, it was considered a hit, if not, it was considered an error. The total number of vowel productions was submitted to a two-way ANOVA examining the main effects of Time and Group and the two-way interaction. The ANOVA yielded no significant effects of group [$F(1, 1,485) = 0.056$ $p = .814$] and no significant effects of time [$F(1, 1,485) = 1.462$ $p = .227$]. The two-way interaction was not significant either [$F(1, 1,485) = 0.357$ $p = .550$]. These results indicate that CLIL instruction had no effect on learners’ pronunciation of English vowels and that neither group of learners made significant improvement over time. The percentage of times that a target vowel was identified as intended averaged across learners (CLIL and FI) and times (1 and 3) for each of the target English vowels is tallied in Table 3. Overall, learners had less difficulty with the vowel phonemes that have a similar phoneme in their L1 systems. Both Spanish and Catalan sound inventories include vowels that are close to English /æ/, /e/, /ʌ/ and /ɒ/, which were identified as intended at 80 % accuracy or higher. In contrast, vowels that do not have a similar phoneme in the L1 such as /u:/, /ɜ:/, /ə/ or /ɑ:/ are more difficult to pronounce (Flege 1995; Rallo Fabra and Romero 2012).

A close inspection of Table 3 revealed that many of the vowel substitutions might have been triggered by spelling. Based on previous research on the influence of orthography on L2 perception (Erdener and Burnham 2005) and L2 transfer (Rafat

Table 3 Percentage of times that each target vowel was identified as intended averaged across times (1 and 3) and groups

Group	Time	/i:/	/ɪ/	/u:/	/ʊ/	/e/	/ɜ:/	/ə/	/ɔ:/	/æ/	/ʌ/	/ɑ:/	/ɒ/
CLIL	1	82	75	28	100	100	46	32	46	100	83	52	78
	3	96	71	46	86	100	39	50	61 %	98	95	52	93
FI	1	91	59	19	81	97	53	53	22	100	81	48	94
	3	91	72	16	87	100	41	50	25	100	81	50	100

Table 4 Phoneme-grapheme frequencies and classification of the target words according to phonological transparency

Vowel	Grapheme	Target word	Frequency	Transparent	Opaque
/i:/	ee	See	249		✓
	ea	Sea	245		✓
/ɪ/	i	Big	5,346	✓	
	ui	Building	16		✓
/e/	e	Best, very	3,316	✓	
/æ/	a	Happy, fact, last	4,192	✓	
/ɑ:/	a	Hard, stars	474	✓	
/ɒ/	o	Fog, top	1,558	✓	
/ʌ/	o	Nothing, some	1,723		✓
	u	But	1,509		✓
/ɔ:/	a	Small, water	165		✓
/ʊ/	oo	Look	114		✓
/u/	oo	Moon, food	173		
/ɜ:/	or	World	321		✓
	ear	Early	29		✓

Note that Fry's (2004) classification relates to American English pronunciation. For this reason, the grapheme *a* in the word *last* is classified as an example of an /æ/ pronunciation. Nevertheless, in our analysis this grapheme was interpreted as /ɑ:/ according to British English pronunciation rules

Table 5 Pronunciation of /ɪ/

	CLIL				FI			
	Time 1		Time 3		Time 1		Time 3	
/ɪ/ ^a	/ɪ/	/ui/ ^b	/ɪ/	/ui/	/ɪ/	/ui/	/ɪ/	/ui/
Building	38 %	19 %	38 %	25 %	50 %	29 %	43 %	36 %
Big	81 %	0 %	100 %	0 %	100 %	0 %	92 %	0 %

^aOnly percentages (rounded up to the nearest whole figure) for the pronunciations /ɪ/ and /ui/ have been included here

^bPhonetic representation of actual pronunciation

2010), we classified the target words as “transparent” and “opaque” on the basis of their spelling. For instance, the two words analyzed for the target vowel /ɪ/ were *big* and *building*. Of the two graphemes (*i*, *ui*), *i* was considered “transparent” or closer to Spanish-Catalan orthography, whereas *ui* was classified as opaque, since most Spanish-Catalan speakers would pronounce it as /ui/. The decision to classify a given grapheme as “transparent” or opaque was also made considering the frequency of occurrence of each grapheme as revised by Fry (2004). The classification of each grapheme is shown in Table 4.

The pronunciation problems caused by possible orthographic interference from the speaker's L1 as in the example of /ɪ/ (*building* and *big*) can be seen in Table 5. The percentage of correct pronunciations for /ɪ/ in *big* contrasts with that of *building* in that nearly 100 % were able to pronounce /ɪ/ correctly in *big*, but less than 50 % in both groups at time 1 and time 3 achieved a correct pronunciation in *building*.

Table 6 Pronunciation of /ɑ:/

	CLIL				FI			
	Time 1		Time 3		Time 1		Time 3	
/ɑ:/ ^a	/ɑ:/	/æ/	/ɑ:/	/æ/	/ɑ:/	/æ/	/ɑ:/	/æ/
Last	6 %	88 %	0 %	100 %	0 %	92 %	0 %	100 %
Hard	75 %	13 %	69 %	31 %	71 %	14 %	71 %	29 %
Stars	69 %	25 %	81 %	19 %	64 %	28 %	64 %	36 %

^aOnly percentages (rounded up to the nearest whole figure) for the pronunciations /ɑ:/ and /æ/ have been included here

An additional two-way ANOVA was run to examine whether the level of difficulty in pronouncing the target words would vary as a function of the vowel and/or as to whether the grapheme was classified as transparent or opaque. Both the main effects of vowel and spelling were significant [$F(11, 1,486) = 18.09 p < .001$, $F(1, 1,486) = 16.13 p < .001$]. However, the two-way interaction was not significant [$F(1, 1,486) = .65 p = .41$]. These analyses indicated that the chances of pronouncing a given vowel as intended depended on the target vowel and whether the grapheme was transparent or opaque. Pair-wise comparisons with Tukey's post-hoc analysis allowed us to establish an order of ease of pronunciation starting with the most difficult vowel phoneme to the easiest following this order: /u/, /ɔ:/, /ɜ:/, /ɑ:/, /ə/, /ɪ/, /i:/, /ʌ/, /ɒ/, /æ/, /e/, /ɜ/.

It was also observed that in some cases, vowel sounds appeared to be “compromise” vowels between English and Spanish. This was especially noticeable with the words *moon* and *food* where the /u:/ frequently resembled /ʊ/, /ʌ/ or the Spanish or General American English (AmE) /u/, a vowel sound that is midway between the English Received Pronunciation (RP) /ʊ/ and the /u:/. One further finding worth mentioning was the pronunciation of *last*. Table 6 reflects the pronunciation of the grapheme *a* in the words *last*, *hard* and *stars*. Nearly 100 % of the participants pronounced the grapheme as /æ/, rather than /ɑ:/ for the word *last*, which reflects a clear preference for the AmE pronunciation. Nevertheless, this case of vowel substitution appeared to be limited, as it did not outwardly affect the pronunciation of the words *hard* and *stars*, perhaps due to the fact that the vowel was followed by an *r* in both cases.

2.3 Discussion

In this study, we examined temporal measures of fluency along with phonemic errors to determine whether late partial immersion had any ameliorating effects on Spanish-Catalan EFL learners' pronunciation and fluency. Do students following CLIL-based instruction speak English more fluently and with fewer pronunciation errors? The analyses show that CLIL learners did not perform significantly better than their FI peers as far as fluency is concerned. Both groups spoke with similar speech rates and exhibited similar gains after 2 years, regardless of the amount and type of classroom-based instruction. This finding is in line with the assumption that overall exposure

to the L2 in the form of spoken language is a prerequisite for oral fluency development (Derwing et al. 2009). Furthermore, Segalowitz (2007) argues that a speaker can develop “high levels of access fluidity and attention control only through extensive exposure and practice with the target language in naturalistic communicative situations” (p. 184). The tutors in charge of CLIL instruction in the present study were not native English speakers. It follows that if learners were not massively exposed to L2 input, they were unlikely to develop higher fluency levels in a formal instruction setting.

We speculate that our outcomes might have been the consequence of task effects. There is evidence in the L2 literature that fluency is sensitive to task-type. Derwing et al. (2004) found that L2 learner speech was perceived by native English judges as being more fluent if the speaking task involved interacting with a native speaker. Fluency ratings of 20 beginner Mandarin learners of English were higher on the monologue and dialogue tasks than on the oral narratives. The authors noted that picture narratives are more cognitively demanding than tasks that offer the speakers more freedom of lexical and grammatical choice, such as a monologue or a conversation. The speech samples we used in the present study were elicited through a picture narrative. Many students showed serious limitations because they did not have the vocabulary they needed to describe the sequence of events that appeared in the pictures. This difficulty triggered many hesitations, fillers and pauses in the learners’ speech. A question that remains unanswered is whether there is a relationship between learners’ grammatical complexity and fluency measurements.

The results of the vowel error analysis showed that CLIL learners’ pronunciation of English vowels was not significantly better than their FI peers’. Neither group of learners showed any improvement after 2 years of CLIL-based instruction. These results raise the question of whether 2 years of content-based instruction are sufficient to have a direct impact on learners’ pronunciation. Furthermore, it should be highlighted that CLIL instruction in this case does not include specific activities to enhance students’ discrimination of English sound contrasts or activities aimed at practicing pronunciation. Therefore, most learners were unlikely to develop mental representations for the English vowel categories that do not exist in their L1 or that are different from the L1 closest equivalents.

Another interesting finding of the present study is that vowel pronunciation errors are highly influenced by the phoneme-grapheme correspondences. Both groups of learners made fewer vowel errors when the target words had more transparent spellings, that is, closer to Spanish-Catalan phoneme-grapheme conversion rules. These results suggest that many pronunciation errors could have been avoided if learners had not relied so much on orthography, and raises the question of whether the introduction of written language in the initial stages of EFL learning should be delayed until the basic rules of pronunciation are well established.

We should also mention the possible influence of AmE on pronunciation. Although the participants were following a British RP-based language tuition program,—the preferred educational standard throughout Europe—the findings suggest that some of the errors could have been the result of an alternative pronunciation. Not only should we acknowledge this possibility but we also need to highlight the overall influence of AmE on the English spoken in Europe (Modiano 1996).

Although this phenomenon appears to affect both CLIL and FI equally, it would be interesting to acquire more information on the phonological nature of the input of English the learners receive both inside and outside the educational environment, paying special attention to language input from the film industry and the media. Recent research has shown that EFL learners who were exposed to audio-visual mass media improved their oral performance as compared to learners who were exposed to social interaction (Bahrani and Shu Sim 2012). Other findings in Computer Assisted Language Learning (CALL) have shown that considerable gains in L2 pronunciation can be achieved if traditional classroom-based instruction is complemented with ASR-based tools aimed at giving immediate feedback on pronunciation errors. For instance, the CAPT system was developed to help foreign learners of Dutch with difficult speech sounds (Neri et al. 2006). The Euronounce project (Demenko et al. 2009), for Slavic learners of German, includes pronunciation training in both segmental and suprasegmental aspects of speech. Some of these tools are designed for young learners. This is the case of Parling (Mich et al. 2006), a word recognition program to train Primary-level Italian learners with the sound-grapheme correspondences of English phonemes. All these systems have proved beneficial for learners' oral skills, either as an alternative to traditional classroom instruction or as self-learning resources.

Conclusion

This study contributes to the literature on L2 acquisition in that it provides data concerning the development of intermediate learners' oral skills in two learning contexts, FI and CLIL. Another contribution is that it provides outcomes in regard to the development of fluency over a 2-year period of time. The lack of robust significant differences in fluency between the two learning contexts shows some limitations of the research that should be addressed in future investigation. Some of these issues include analyses of listener judgments of fluency as well as measures of L1 fluency. Other suprasegmental measurements such as stress timing or peak alignment would provide a more thorough insight into learners' oral skills. Finally, it would also be interesting to investigate whether there is a relationship between fluency and the learners' performance in other skills such as grammar and vocabulary.

The uniformity of both learner groups in terms of pronunciation achievement seriously questions the effectiveness of CLIL to enhance learners' oral skills in a foreign language. The results of this research suggest that more work needs to be done to increase the quantity and quality of the oral input students receive. On the one hand, CLIL practitioners should broaden the learning scope for learners by encouraging them to listen to English through the media and facilitating the use of English in lingua franca settings. On the other hand, education authorities need to provide the necessary support to CLIL practitioners in order for them to be in a position to offer learners all the advantages available through the CLIL teaching system.

Acknowledgements This research was funded by research grants from the Spanish Ministry of Economy and Competitiveness (HUM2007-66053-C02-01/02, FFI2010-21483-C02-02 and FFI2013-48640-C2-1/2-P) and the Catalan Government (SGR2005-01086/2009-140/2014-1563).

References

- Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effects in a continuum of bilingualism. *Bilingualism: Language and Cognition*, 15, 517–530.
- Bahrani, T., & Shu Sim, T. (2012). Informal language setting: Technology or social interaction. *The Turkish Online Journal of Educational Technology*, 11(2), 142–149.
- Basetti, B. (2009). Orthographic input and second language phonology. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 191–206). Bristol: Multilingual Matters.
- Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer* [Computer program]. Version 5.3.41. <http://www.praat.org/>. Accessed 12 Feb 2013.
- Burnham, D. K., Tyler, M., & Horlyck, S. (2002). Periods of speech perception development and their vestiges in adulthood. In A. Rohde (Ed.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 281–300). Trier: Wissenschaftlicher Verlag Trier.
- Cucchiari, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, 107, 989–999.
- Cucchiari, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, 111, 2862–2873.
- De Jong, N., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, 41, 385–390. doi:10.3758/BRM.41.2.385.
- Demenko, G., Wagner, A., Cylwick, N., & Jokisch, O. (2009). *An audiovisual feedback system for acquiring L2 pronunciation and L2 prosody* (pp. 2–5). 2nd ISCA workshop on Speech and Language Technology in Education (SLaTE).
- Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, 54, 655–679.
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31, 533–557.
- Erdener, V. D., & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, 55, 191–228.
- Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler, & W. S.-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85–101). New York: Academic.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore: York Press.
- Fry, E. (2004). Phonics: A large phoneme-grapheme frequency count revised. *Journal of Literacy Research*, 36, 85–98.
- Fullana, N. (2006). The development of English (FL) perception and production skills: Starting age and exposure effects. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 41–64). Clevedon: Multilingual Matters.
- Gallardo del Puerto, F., Gómez Lacabex, E., & García Lecumberri, M. L. (2009). Testing the effectiveness of content and language integrated learning in foreign language contexts: The assessment of English pronunciation. In Y. Ruiz de Zarobe & R. Jiménez Catalan (Eds.), *Content and language integrated learning. Evidence from research in Europe* (pp. 63–80). Clevedon: Multilingual Matters.

- Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. New York: Newbury House.
- Hammerly, H. (1991). *Fluency and accuracy. Toward balance in language teaching and learning*. Clevedon: Multilingual Matters.
- Harada, T. (2007). The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. *International Review of Applied Linguistics*, 45, 353–378.
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29, 24–49.
- Juan-Garau, M., & Salazar-Noguera, J. (2015). Learning English and learning through English: Insights from secondary education. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 105–121). Berlin: Springer.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah: Lawrence Erlbaum Associates.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387–417.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384–422.
- Mich, O., Neri, A., & Giuliani, D. (2006). The effectiveness of a computer assisted pronunciation training system for young language learners. In *Proceedings of CALL 2006* (pp. 135–143), University of Antwerp, Antwerp, 20–22 August 2006.
- Modiano, M. (1996). European English. *World Englishes*, 15, 207–215.
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46, 610–641.
- Neri, A., Cucchiari, C., & Strik, H. (2006). ASR-based corrective feedback on pronunciation: Does it really work? In *Proceedings of the 9th International Conference on Spoken Language Processing INTERSPEECH 2006* (pp. 1982–1985), Pittsburgh 17–21 September 2006.
- Rafat, Y. (2010). Orthography as a conditioning factor in L2 transfer: Evidence from English speakers' production of Spanish consonants. In *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010*. Adam Mickiewicz University, Poznan, 1–3 May 2010.
- Rallo Fabra, L. (1998). Speaking rate effects in Catalan and English. A cross-language study. *Estudios de Fonética Experimental*, IX, 243–272.
- Rallo Fabra, L., & Juan-Garau, M. (2011). Assessing EFL pronunciation in a semi-immersion setting: The effects of CLIL instruction on Spanish-Catalan learners perceived comprehensibility and accentedness. *Poznan Studies in Contemporary Linguistics*, 47, 96–108.
- Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, 40(3), 491–508.
- Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14, 357–385.
- Segalowitz, N. (2007). Access fluidity, attention control, and the acquisition of fluency in a second language. *TESOL Quarterly*, 41, 181–186.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
- Strange, W. (1995). Cross-language studies of speech perception. A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 3–46). Baltimore: York Press.
- Toivola, M., Lennes, M., Korvala, J., & Aho, E. (2010, May 1–3). A longitudinal study of speech rate and pauses in non-native Finnish. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the 6th international symposium on the acquisition of second language speech, new sounds 2010* (pp. 499–504). Poznan: Adam Mickiewicz University.