

Is CLIL so beneficial, or just selective? Re-evaluating some of the research

Anthony Bruton

Departamento de Filología Inglesa (Lengua Inglesa), Universidad de Sevilla, 41004 Sevilla, Spain

Received 15 September 2010; revised 4 August 2011; accepted 9 August 2011

Available online 19 October 2011

Abstract

A number of studies on CLIL, particularly from Spain, which is familiar to this author, will be analysed to show that there are numerous anomalies not only in the research, but in the analysis, and doubts about the conclusions drawn. CLIL instruction is not always necessarily that beneficial, and there is every reason to believe some students may be prejudiced by CLIL, and that not only academic, but also institutional, interests may be taking precedence over some students' interests in the state educational sector. Some research issues are covered in the detailed analysis of one study before a plea is made for ensuring that disinterested research is carried out into the *overall* effects of CLIL initiatives in state educational institutions and systems, so that the welfare of all state-school students is recognised and respected.

© 2011 Elsevier Ltd. All rights reserved.

Keywords: Content and Language Integrated learning (CLIL); Communicative Language Teaching (CLT); Communicative Approach; ELT syllabus and methodology

1. Preamble

It seems to be characteristic of many of the CLIL initiatives in the EU and much of the published research that CLIL necessarily produces encouraging outcomes, very often in terms of quantitative scores. Given the rationale for CLIL, that by integrating foreign language (FL) and content learning there will be improved FL learning with at least no cost to content learning (D. Marsh),¹ this is not altogether surprising. However, a closer look at some of the research conducted into CLIL and content learning in an L2 suggests that such initiatives do not necessarily produce better results than the alternatives they compete with, even in FL development terms. Some of the recent quantitative research into CLIL and related initiatives will be the major consideration here, hopefully dispelling any expectation that there are necessarily cause-effect relationships between the initiatives and increased language learning outcomes.

CLIL is sometimes defended as an alternative to Communicative Language Teaching (Coyle et al., 2010) and sometimes as an extension of it (Dalton-Puffer, 2007; Lasagabaster and Sierra, 2010). Either way, CLIL shares many aspects of CLT, while emphasising academic content as the substance of the communication, which is supposed to

E-mail address: ab@us.es.

¹ This reference to Marsh is not dated as it appears on Internet without dates.

make this communication more relevant and purposeful, by fulfilling more immediate needs (Lorenzo et al., 2010: 429). However, it should be recognised that really the CLIL defended on paper, reflected for example in Barwell's (2005) definition, "Language and content integration concerns the teaching and learning of both language and subject areas (e.g. science, mathematics, etc.) in the same classroom, at the same time" (p.143), is rarely the CLIL in practise. In many cases the FL-language content classes are supplemented with additional FL classes (Dalton-Puffer, 2007), which more recently seems to be justified by Coyle et al. (2010) as allowing more exposure to the FL. In fact, there seem to be three possible CLIL variations: 1) Learn the FL separately, in order to learn the content through the FL; 2) Learn the FL through the content, which has already been learnt in the L1; 3) Learn the FL and the content together – see Lucietto (2008) for a discussion. Most of the arguments justifying CLIL suggest option (3), but Mehisto (2008) found that it rarely occurred in practise, so it is likely that (1) or (2) actually prevail. As far as Dalton-Puffer (2008) is concerned, CLIL classes should actually supplement the language of the FL classes, which is somewhat different from acquiring the language through the content instruction. A bilingual option might also be contemplated as in the Serra (2007) study, but in this study the students were not learning FLs so much as minority local languages, with bilingual teachers.

2. Some potentially contrary outcomes

There are a number of questions about CLIL which will be enumerated briefly, since they will be discussed in more depth below. It is assumed that making academic content the goal of the learning in the FL will increase motivation, when actually the opposite could be possible – see Seikkula-Leino (2007) on the possible negative effects of CLIL on self-esteem or Lasagabaster and Sierra (2009) below on the drop in interest in CLIL. What is more, academic content may not stimulate more oral interaction on the part of the students (Dalton-Puffer, 2007), but actually less if the subject matter is novel. Furthermore, if the content is conceptually difficult, the FL medium will make it even more difficult to assimilate, and the content in turn could complicate the FL development – see Seikkula-Leino (2007:338) on the amount of language above the students' current competence. Sometimes when the content becomes too difficult, students resort to their L1, especially in peer work (Coonan, 2007). Add to this that the content teachers' English might be limited and the possible benefits of CLIL can begin to look dubious. Partly for these reasons, it seems that many CLIL initiatives, as with immersion ones (Marsh et al., 2000), have been limited to students who have been selected in some way or other. Furthermore, in immersion contexts, there are cases of 20% drop-out rates (Netten and Germain, 2009). There is the question of cultural content as well, which might be ignored in the instruction of other content, and which has always been one of the reasons for including FLs in state-school curricula.

Despite the arguments of D. Marsh and Lorenzo et al. (2010) that CLIL is egalitarian, the truth is many of the potential pitfalls which CLIL might encounter are actually avoided by selecting for these programmes students who will be academically motivated to succeed in the FL, as in other subjects – see Ruiz de Zarobe and Lasagabaster (2010:25). But even with this provision, the research results are not particularly encouraging, as the critical assessment later will hope to show. In the next section, we will consider a study into content teaching in the L2 and a CLIL case. This will be followed by an extended section on CLIL research in Spain, before considering an extended research project on CLIL in Andalusia (Spain) in more detail, since it demonstrates how CLIL research can be deceptive. Some points will be supported with references to studies of CLIL outside Spain.

3. Some initial doubts: a Hong Kong study

Unfortunately for the CLIL argument, the results of the study by Marsh et al. (2000) over three years of secondary school in Hong Kong are rather disturbing. In their conclusions, the authors state:

In summary, Hong Kong high school students were very disadvantaged by Instruction in English in geography, history, science, and, to a lesser extent, mathematics. The size of this disadvantage was reasonably consistent across the first three years of high school. (Marsh et al., 2000: 337)

The least disadvantaged were those who had higher levels of English to begin with and who were in English language classes where English was most used. Unless the variable had been isolated, the results would have been contaminated by the fact that many of the academically bright students attended the English-medium schools, thus raising the scores of the English-medium classes. However, the academically bright with low initial English proficiency were probably the most

prejudiced. The nature of the different subjects taught in English, the English received in the English language classes, and the initial level of the students' English therefore do matter. In addition, these authors refer to two other interesting factors. First, despite these typical results in both content and language, many parents in Hong Kong persevere with English-medium education, because they think that their children will have more opportunities in the future if they can speak English. Second, the authors conjecture that if the students have not reached a threshold language level, they cannot cope, and as the subject matter content becomes more complicated so the language threshold rises, which might also be true of the English of the content teachers. There have been some questions asked about the level of English of some of the teachers, apparently, but that would be an issue in any similar teaching situation.

Similar outcomes occur in a study conducted by [Seikkula-Leino \(2007\)](#) in a comprehensive school, in Finland, in which the CLIL groups were selected, and 40–70% of the CLIL maths classes were conducted in English. On the maths test, in the distribution of students in terms of under-average-high-achievers according to their IQ potential, there were far fewer over-achievers in the CLIL group than in the normal group (30%–10%), which seems to suggest that the CLIL group might have been disadvantaged, despite the pre-selection. The CLIL groups were also shown to be more highly motivated, but with low self-esteem on FL ability.

4. Reinterpreting some CLIL results in Spain

The first study for consideration of CLIL in the Basque Country in Spain is by [Ruiz de Zarobe \(2007\)](#), with only 24 subjects in total. Two groups, a 'traditional' and CLIL group, with twelve years of English, were compared, the major difference being that the CLIL group took two science courses in 3rd and 4th year of secondary in addition to their English classes. The results on an oral retelling task are summarised in [Table 1](#).

There were only marginal differences on all counts, although there was a significant difference on the utterance length. The first thing to notice is that it is not so much that the CLIL students received 210 more hours of English, but that these were concentrated in the last two years, so the students received 105 h per year more English than the other group prior to the test, without including homework etc. – this would be the equivalent of double the hours per week in the last two years. The author then concludes that:

The results presented in this study indicate that, although the CLIL group performed better in most categories analysed, there are no overall significant differences between both groups (traditional versus CLIL) in relation to oral proficiency. This could be due to the fact that the difference in the amount of hours (210) is not sufficient to obtain significantly better results. (p. 51)

But first of all, the results do not show that the CLIL group's performance was better on most counts – and even if they did with such a small sample, the results would be of dubious validity. Secondly, the results show that, despite 210 or more hours of exposure to English, the differences were negligible – see the effect of 200 h of EFL in [Muñoz \(2006\)](#) below. In this case it could be argued that hour-for-hour the results of the CLIL group were actually considerably *worse*, an interpretation that could also be applied to the difference between the minor advantage of the English-medium over the L1-medium groups in the [Marsh et al. \(2000\)](#) study. Furthermore, there is no assurance that there was any actual improvement in both groups since there was no pretest. Finally, it is probable that the 'traditional' group would perform better on the science content in their L1, all other things being equal, which they probably were not.

In what is apparently part of the same study published by the same author ([Ruiz de Zarobe, 2008](#)) across three levels of secondary school (3rd, 4th and 5th years), the conclusion is that the two CLIL groups at each level outperform the single non-CLIL groups "in every single category of the speech production test" (p. 70) and that "this study serves as evidence that CLIL can be more effective than traditional foreign language teaching in promoting proficiency in the foreign language." (p.70). This position is reasserted in [Ruiz de Zarobe and Lasagabaster's \(2010\)](#) summary of some

Table 1
Results of story telling task in [Ruiz de Zarobe \(2007\)](#) study.

Groups	Level	Hours of FL	Results: features better than other group
Traditional	4 ESO (15–16)	1148 (12 years)	Types, Tokens, Utterance Length, Pronunciation
CLIL	4 ESO (15–16)	1358 (12 years)	Vocabulary, Grammar, Fluency

CLIL research in the Basque Autonomous Community, where the authors recognise that admission into CLIL classes is voluntary, but selective, so the average ability and motivation is probably higher in both English and the other subjects. On closer inspection of the (selected) results, in Table 2, the following facts emerge, however. First, the CLIL groups had received over two hundred, three hundred and four hundred hours more of the FL by the end of each level than the non-CLIL groups. Although in Secondary 3 there seem to be visible differences between the average scores of the CLIL and non-CLIL groups, suggesting that the non-CLIL students were less proficient at this stage, the same is not true at Secondary 4, where one of the CLIL groups actually scores lower than the non-CLIL group on grammar (and also fluency and content), and the other CLIL group scores similarly on these features. The scores of the non-CLIL group at Secondary 4 have actually increased significantly more than the two CLIL groups, and look very similar to the results of the CLIL groups at Secondary 3, but with fewer additional hours. Given the huge number of extra hours of English for the CLIL groups (460 for the Pre-U students), the results cannot be very encouraging — Ruiz de Zarobe and Lasagabaster (2010) also mention that the non-CLIL group “did not receive any extra-English classes outside school” (p.15), suggesting that the others did.

In a further study in the Basque Country, Alonso et al. (2008) compared three levels (Compulsory Secondary Education 1 and 3, and Post-CSE 1) of CLIL and non-CLIL groups over two years. The students took different levels of Cambridge ESOL tests at the end-of-year 1 and year 2. Apart from some rather misleading graphs, the authors conclude that the CLIL groups always outperform the non-CLIL groups in English, based on the results reproduced in Table 3. They state that “The PE [plurilingual education] increases the rate of learning the vehicular language and the improvement in linguistic and communicative competence appears to be substantial” (p. 47), and at no expense to content learning.

However, although their conclusion is true, and although the authors state that the control group students are in the same grade and have similar overall academic achievement according to previous grades, it is clear from the 2004 results that the groups do not start with the same percentage averages, and the control group scores are much lower. The scores also show that the differences between the two groups widen, but this increase diminishes drastically to 2% in Bach1-2 — see Table 4. Although the tests are different in the first and second years, the percentages increase from CSE1-2, are level at CSE3-4 and drop for Bach1-2, suggesting that the increases diminish progressively at each stage, after the initial jump for the CSE1-2 experimental groups. Despite some limitations in the research, it is obvious that the non-CLIL students are the less proficient to start with and remain that way, so may be selection was the operative variable — see the reference to Admiraal et al. (2006) below for a similar outcome.

Among other results in the Villarreal Olaizola and García Mayo (2009) study, again in the Basque country, a CLIL group of 27 students produce significantly fewer errors of omission on inflectional 3rd person singular -s and simple past -ed than another group of 29 non-CLIL students, although crucially there was no pretest — otherwise there is little difference due to a floor effect. However, the CLIL students had over 300 h more English than the non-CLIL

Table 2
Selected results from Ruiz de Zarobe (2008).

Groups	Variables	SEC 3	SEC 4	PRE-U
Non-CLIL	Number	29	28	?7
	Age	14–15	15–16	17–18
	Hours of instruction	695	792	990
	Vocabulary	4.4	6.3	6.3
	Grammar	4.0	6.6	6.1
CLIL 1	Number	24	16	—
	AGE	14–15	15–16	17–18
	Hours of instruction	875 (+180)	1120 (+328)	—
	Vocabulary	6.5	6.4	
	Grammar	6.5	6.1	
CLIL 2	Number	36	17	14
	Age	14–15	15–16	17–18
	Hours of instruction	910 (+215)	1155 (+363)	1453 (+463)
	Vocabulary	6.2	6.9	7.4
	Grammar	6.5	6.9	7.1

Table 3

Overall results of the control and experimental groups in the [Alonso et al. \(2008\)](#) study (with experimental-control differences added in brackets).

	2004 (% difference)		2006 (% difference)	
	Experimental	Control	Experimental	Control
CSE1-to-2	57%	49% (–8%)	82%	60% (–22%)
CSE3-to-4	77%	59% (–18%)	79%	55% (–24%)
Bach1-2	72%	55% (–16%)	61%	43% (–18%)

(selected from [Alonso et al., 2008](#)).

groups, and the authors admit: “Unlike in the first group [non-CLIL], we could not control for the ‘extra-English’ variable – most learners (78%) attended English classes after school” (p. 163). However, they do go on to say that the correlation between extra hours and outcomes was not strong, despite the fact that they could not apparently control for these hours, the samples were very small (15 and 12 for the CLIL groups) and there was the floor effect mentioned. Even so, the CLIL students had extra private English classes, which suggests that these students tend to be of a higher socio-economic level than the non-CLIL ones. They probably opted for CLIL as well, which might have been evident if there had been a pretest. The authors conclude that the results are in line with other studies “advocating for the benefits of using the language as the means of instruction rather than as just the goal of instruction” (p. 171), although they admit that this would have to be confirmed in the longer term. Such a conclusion seems surprising, given the tendencies over time.

Interestingly, in another previous study not on CLIL with bilingual Basque/Spanish participants, the second author and other colleagues ([García Mayo et al., 2005](#)) show that older learners pass through developmental stages much faster than younger learners with the same number of class-time hours within a communicative approach, but make the following statement about the instruction:

It should be borne in mind, however, that the way the input is presented in formal settings varies according to the age range of the learners (more metalinguistically oriented instruction as they grow older). Besides, the cognitive development of the learners is a factor that cannot be disregarded when the way they approach new input is considered. (p. 471)

[Muñoz \(2006\)](#) shows that the older group in her study with Catalan students of English with 200 h of exposure demonstrated higher accuracy rates than the younger groups with the same number of hours of exposure, but in this study it is possible that the older students opted for English. The point of these two references is to make the point that there are other significant variables to be accounted for in making learning-outcome comparisons, such as age and exposure, cognitive development and instruction, to which [Rothman and Guijarro-Fuentes \(2010\)](#) add the variable of type of classroom input, varying between naturalistic, SL and FL. The [Muñoz \(2006\)](#) study also shows that adult students, her D1 group, can make considerable accuracy gains on various morphosyntactic functors in a short period of time, and the [García Mayo et al. \(2005\)](#) study shows some positive effects of the communicative approach with only 3 h per week.

[San Isidro \(2010\)](#) offers a very brief summary of a CLIL study in Galicia, another autonomous region of Spain. In his conclusions, he states, “This study provides empirical evidence that the CLIL approach is successful and helps to improve students’ foreign language competence, ...” (p. 75). However, he previously admits:

Results must be treated with caution, since there were language competence differences between both groups prior to the experimental group entering the CLIL programme. The CLIL students who took part in the testing enrolled in the CLIL programme voluntarily. Their degree of motivation and their English language competence were possibly higher than those of their non-CLIL counterparts. (p. 74)

Table 4

% change in differences between the control and experimental groups in the [Alonso et al. \(2008\)](#) study.

	2004–2006	
CSE1-2	22%–8%	=+14%
CSE3	24%–18%	=+6%
Post-CSE	18%–16%	=+2%

So, the author suggests that non-CLIL students were lower on average as a group at the beginning of the study, verified in Table 5, which shows that the percentage profiles of the non-CLIL was initially much lower than the CLIL group on the students 2006–7 end-of-year scores. Since the 2009 assessment instrument is completely different from the 2006–7 instruments, and the overall percentage scores are lower with the modal groups dropping by one category in both cases, there is actually nothing reliable to show that “the CLIL approach is successful”. Finally, we are given no evidence of what the CLIL stream supposed in terms of subjects, hours, activities, and so on. The same is true of the Hüttner and Rieder-Bünemann (2007) report of their study in Austria, which showed CLIL advantages on accuracy of verb forms, but with the serious limitation of no pretest, although the authors admit the CLIL students were probably more motivated as they could opt for the programme.

Navés and Victori (2010) mention various studies which seem to show that neither early start nor more hours of formal exposure in an FL are advantageous for any significant language improvement, while there are indications that CLIL can be. However, they do not recognize that most CLIL groups, in Spain at least, are either overtly or covertly selective in terms of both initial proficiency and motivation. In the two studies they describe in Catalonia, basically the non-CLIL (EFL) groups start off at a disadvantage at 5th grade in both general English proficiency and in writing proficiency and this disadvantage is maintained, but not in all cases. It is only possible to refer to 5th, 7th and 9th grade subjects as no 10th grade EFL students are included and no 11th or 12th grade CLIL ones. The authors conclude that the CLIL students are often a grade level or two ahead of their non-CLIL counterparts, but they warn that “implementing CLIL programmes at any cost, does not necessarily guarantee that those gains will take place” (p. 48). A similar result was obtained by Admiraal, Westhoff and de Bot (2006) with the bilingual English students in Holland, who started with an advantage on the English vocabulary scores which remained parallel thereafter – the comprehension and oral scores may have reflected the same tendency, but the initial scores are not given. Three of the bilingual schools were for ‘International Education’ and the other two were in the same school as the regular students, so they may have been selective. In this case it seems that the bilingual students began with an advantage which the CLIL programme, with an expected 50% of the classes in English, did not alter.

Lasagabaster and Sierra (2009) compared the attitudes to English of 287 CLIL and ‘traditional EFL’ students at secondary 3 (SE3) and secondary 4 (SE4) levels in the Basque Country, using a 7–point semantic differential scale. They argue that CLIL has a positive impact on attitudes at both levels, given the comparative scores in favour of the CLIL groups at each level, and conclude that “Our results seem to indicate that CLIL may be a very useful approach to keep students interested” (p. 13), since they compared the attitudes of each group at each level. A short summary of this study is included in Ruiz de Zarobe and Lasagabaster (2010), who state that “The authors affirm that CLIL may be a very useful approach to keep students interested in the learning of English as a foreign language, as the attitudinal decline observed in studies undertaken in EFL (non-CLIL) contexts (...) seems to wane once CLIL is implemented” (p. 24). Although what Lasagabaster and Sierra say is true, they only make comparisons at each level, noting that the trend for CLIL to produce more positive responses “seems to diminish as the students go up in the educational ladder” (p. 11). However, consider the scores in Table 6, which shows the changes between the two years per group: CLIL SE3-to-SE4 and EFL SE3-to-SE4. What the authors do not mention is that the *change* in the CLIL scores are noticeably more negative than the (non-CLIL) EFL ones, over the limited period of a year in a cross-sectional study – especially in terms of ‘nice’, ‘pleasant’ and ‘appealing’, which hardly suggest the CLIL experience sustains more interest, which also dropped. This should be a big cause for concern. Furthermore, the students appear to be from the same schools, which would mean the CLIL students opted for that programme. This does not deter the authors from reasoning that CLIL is more meaningful, authentic and relevant than ‘traditional EFL’ instruction. The SE3 scores again suggest some form of selection, especially from the scores on the term ‘easy’, which actually rise for the SE4 EFL group and fall for the SE4 CLIL group.

The point of this mini-analysis is to show four things: (1) results of some empirical research can be interpreted and biased in various ways, depending very often on researcher interests; (2) some of the studies are very limited, and the

Table 5

Adapted percentages per mark band for the CLIL and Non-CLIL groups in San Isidro (2010).

Years	CLIL (by marks bands)				Non-CLIL (by marks bands)			
	0.0–2.5	2.5–5.0	5.0–7.5	7.5–10.0	0.0–2.5	2.5–5.0	5.0–7.5	7.5–10.0
2006–7	0	0	43.5%	56.5%	3.8%	5.4%	72.3%	18.5%
2009	0	12.3%	44.8%	42.9%	2.3%	52.6%	36.8%	8.3%

Table 6
Results from the Lasagabaster and Sierra (2009) study on attitudes to English.

	Nice	Appealing	Pleasant	Easy	Important	Useful	Interesting	Necessary
SE3: EFL	4.32	3.92	4.23	2.87	5.62	6.00	4.47	5.93
SE4: EFL	4.21	3.91	4.09	3.20	6.18	6.42	4.13	5.80
EFL change	-0.11	-0.01	-0.14	+0.37	+0.56	+0.42	-0.34	-0.13
SE3: CLIL	5.06	4.96	5.11	4.11	6.45	6.79	5.11	6.54
SE4: CLIL	4.62	4.37	4.58	3.83	6.51	6.63	4.71	6.37
CLIL change	-0.44	-0.59	-0.53	-0.28	+0.06	-0.16	-0.40	-0.17

results questionable in numerous ways, particularly in terms of pretesting, sampling, and observation data on actual instruction; (3) most of the non-CLIL groups are on average less proficient, or probably less motivated to begin with, suggesting that the CLIL groups attract the more proficient/motivated; (4) both quantitative and qualitative results such as these are not very encouraging for CLIL, especially since the CLIL groups typically start off more motivated and with higher initial scores.

5. Some research design questions: an Andalusian study as an example

In another study, in Spain, but this time in Andalusia, apparently the language results were much more favourable for the CLIL over the non-CLIL groups in the other language of the EU, namely English, French and German. In fact, the authors (Lorenzo et al., 2010) state that “When the results of the linguistic evaluation had been compiled, it emerged that the CLIL learners were clearly outperforming their mainstream peers. Global average scores were 62.1 per cent for the bilingual groups in comparison with 38 per cent for the control groups” (p. 426). Furthermore, the authors underscore the fact that “admission to the bilingual sections is open to everyone” (p. 422) and that “Bilingual sections are, therefore, essentially egalitarian (although the possibilities of corollaries between social class and parental choice cannot be ignored)” (p. 423) — a possible characteristic not considered so laudable by the same authors more recently (Lorenzo et al., 2011). Given that these enormous differences resulted after only a year and a half (p. 427), the results should have given the authors reasons to suspect that something was afoot. Even more so, given that many of the CLIL students apparently achieving OLEU language scores above some of the CLIL teachers themselves, at the end of primary at least (Lorenzo, 2008).

Actually, there were a number of factors to explain the global result — a somewhat dubiously simplified composite score in itself. The first is that although the authors state that the CLIL streams are open to everyone, it is generally acknowledged that essentially the students who opt for, and are very often encouraged into, the bilingual programmes are the highly motivated ones, whose parents are generally in the higher socio-economic classes. There is an implicit selection (see also Alonso et al., 2008; or Ruiz de Zarobe and Lasagabaster, 2010), since the middle-class parents believe that their children will be with other motivated students. Reference to this fact was voiced by a number of teachers during the event for teachers participating in the bilingual centres (*centros bilingües*) co-hosted by Lorenzo (2008) and the Andalusian education authorities — Lasagabaster (2008) also shows that parents of 65% of the CLIL group students in his study had university education, even though thereafter socio-cultural status did not affect within-CLIL scores apparently, which is not the point here. There is every likelihood that a percentage of these CLIL students would be taking extra-English classes outside school as well, as in the Villarreal Olaizola and García Mayo (2009) study. Furthermore, the control groups are the remnants from the (selected) CLIL groups in the *same* schools, which is a critical factor in the study, but selection in state schools cannot (legally) be recognized by the regional education authority. Casal and Moore (2009:39), two of the authors, confirm that the CLIL groups are not compared to ‘control’ groups of the same initial proficiency in other non-CLIL schools.

The authors did not publish any pre-programme scores, either, but there is little doubt that the control groups started with much lower average scores than their CLIL counterparts — see some of the studies above. Furthermore, there are post-programme numerical scores for the two groups, but inexplicably without any pre-programme ones. So, we do not know where either of the two groups started in relation to each other, or how much they changed, only that at the end the CLILs were on overall average much better than the non-CLILs — see the Alonso et al. (2008) and the Villarreal Olaizola and García Mayo (2009) studies, as well. It is possible, of course, that the non-CLIL groups actually deteriorated.

The other critical factor was the language assistants, who were seconded to the CLIL, but not the non-CLIL, groups. In other words, not only did the CLIL groups have native-speaker teachers, but these teachers were extra teachers, who helped with language materials as well (Lorenzo, 2010). It is very possible, and likely, that the real difference in the language improvement, if this is actually the case, was due to the added English and added teacher support of the language native-speaker assistants rather than the FL input of the content teachers – similar to the Marsh et al. (2000) study in Hong Kong. This is not what CLIL is supposedly about.

To have valid comparisons of the effect on the OLEU development in content and non-content classes, the classes would have to be similar in most other respects, including the methodology used, which is not even given consideration, the motivational and proficiency levels of the students and teachers (Alonso et al., 2008), the hours, the support, etc. Some observational data would be necessary rather than only the questionnaire data on classroom activity Lorenzo et al. (2010) offer, which is notoriously unreliable and limited in scope (Bruton, 2011). The limitation in the type of data is all the more surprising, along with the composite numerical post-programme comparative scores, given that Lorenzo et al. (2011) themselves recognise they are dealing with a complex phenomenon. In Andalusia, it would be problematic to find comparison groups, since, in non-CLIL schools, it is not permitted to select and stream state-school students, though it might be possible to assess selected students and classes. It would also be of interest to have contrasting student views (Alonso et al., 2008), and it is crucial to know what actually happens in so-called CLIL classrooms and how they are characterised (Dalton-Puffer, 2007; Mehisto, 2008), a view shared by Leung (2005), who insists that:

A key point to bear in mind is that, irrespective of the concerns of the moment, be it social integration, nation building, language and curriculum learning or language revival, the claims for or against bilingual education of any form ring hollow when there is not a clear sense of what happens inside the classroom. (p. 238)

However, a preoccupying fact in Andalusia is that there is virtually no existing research on non-CLIL language classrooms and the same with respect to content-subject classrooms. Given the very low ratings for Andalusia in the PISA reports this should be a priority. According to Dalton-Puffer (2007:118) again, she discovered that most content teaching in Austria was teacher-fronted and dealt with facts, facts, and more facts, in what Barwell (2005) calls a ‘conduit’ model of teaching. However, Coonan (2007) in Italy, with a group of highly selected CLIL teachers, concludes that there was a visible move to more peer interaction, although it tended to be in the L1, as was the case in the limited peer work observed by Dalton-Puffer (2007). Unfortunately, this Italian study, which relied on teacher responses rather than actual observation like Lorenzo et al. (2010), is rather contradictory as the teachers in the programme contend that there is greater complexification and peer interaction, in contrast to Dalton-Puffer (2007), but “lament scarce participation” (p.641). Such data demonstrate the unreliability of teacher responses even on the frequency of specific class activity. Coonan (2007) does add a cautionary comment that there is the danger of oversimplification of content, on the one hand, and that “the potential for L2 development offered through complex activities runs the risk of backfiring” (p. 641), since the author maintains that the use of the L1 in the pairwork is due to the tasks being too complicated.

6. Conclusion

Apart from the research questions addressed, two crucial consequences of CLIL have to be investigated. Firstly, what is the real effect of this on the development of the content learning and the methodology of the content classrooms, for the average student? Secondly, what is the effect on the students who are not selected in a selective scenario, or on the less able if the initiative is across-the-board? It is very convenient to select and then demonstrate that the selected perform better than the non-selected with additional language exposure, even though this does not always transpire. Of course, nothing conclusive can be shown with no pre-post average scores, no valid comparison groups, and no comparable contexts, with no control over extra FL instruction outside school, and the support of additional teachers and coordination time, to name but a few fundamental considerations.

The point of this discussion is to voice some concerns about the adoption of CLIL in certain educational circumstances. There is no doubt that there are potentially beneficial effects from adopting CLIL initiatives and some positive results have been observed. Nevertheless, CLIL cannot be seen in a vacuum, and in the region of Europe where this author lives and works, there are serious questions to be asked about the quality of the education state secondary school students receive, with the local version of CLIL and without it. If the actual pedagogy that is typical in most classrooms is producing deficient results, it would seem that adding the burden of using an FL for content

instruction is perhaps overambitious, unless there is some form of student selection, which there obviously is, not only here, but in many other FL immersion and CLIL contexts – immigrant ESL contexts are very different. Such initiatives are all the more worrying with the adoption of CLIL by many teachers with very limited training in FL content teaching, and sometimes limited FL ability. Of course, if the point of reference is an existing deficit FL curriculum, almost any motivated initiative will produce some benefits in the FL for some selected students, especially if they are compared to the unselected students, but that is surely not the point. Any state educational system should ensure adequate standards in the L1 medium for all students, before spreading the FL medium across part of the curriculum for certain students, possibly to the detriment of some of the rest, who remain in the existing seemingly deficit FL scenarios. In every case, there is always a price to be paid. Even when the circumstances for such initiatives might be appropriate, it is necessary to assure the public, in state education at least, that serious *disinterested* reliable quality research is being conducted to try to ensure that expected content and language standards, both L1 and FL, are being met across the board. That is, we must expect a true reflection of the whole educational scenario. Anything less is simply compromised and irresponsible.

Acknowledgement

This publication was made possible by the funding of project FFI2010-19022 by the Spanish Ministry of Science and Innovation.

References

- Admiraal, W., Westhoff, G., de Bot, K., 2006. Evaluation of bilingual secondary education in the Netherlands: students' language proficiency in English. *Educational Research and Evaluation* 12 (1), 75–93.
- Alonso, E., Grisaleña, J., Campo, A., 2008. Plurilingual education in secondary schools: analysis of results. *International CLIL Research Journal* 1 (1), 36–49.
- Barwell, R., 2005. Critical issues for language and content in mainstream classrooms: introduction. *Linguistics and Education* 16, 143–150.
- Bruton, A., 2011. Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore. *Applied Linguistics* 32, 236–241.
- Casal, S., Moore, P., 2009. The Andalusian bilingual sections scheme: evaluation and consultancy. *International CLIL Research Journal* 1 (2), 36–46.
- Coonan, C.M., 2007. Insider views of the CLIL class through teacher self-observation-introspection. *The International Journal of Bilingual Education and Bilingualism* 10, 625–646.
- Coyle, D., Hood, P., Marsh, D., 2010. *CLIL: Content and Language Integrated Learning*. Cambridge University Press, Cambridge.
- Dalton-Puffer, C., 2007. *Discourse in Content and Language Integrated (CLIL) Classrooms*. John Benjamins, Amsterdam.
- Dalton-Puffer, C., 2008. Outcomes and Processes in Content and Language Integrated (CLIL) Learning: Current Research in Europe. In: Delanoy, W., Volkman, L. (Eds.), *Future Perspectives in English Language Teaching*. Carl Winter, Heidelberg, pp. 7–23.
- García Mayo, M.P., Lázaro Ibarrola, A., Licerias, J.M., 2005. Placeholders in the interlanguage of bilingual (Basque/Spanish) children. *Language Learning* 55 (3), 445–489.
- Hüttner, J., Rieder-Bünemann, A., 2007. The effects of CLIL instruction on children's narrative competence. *Views* 16 (3), 20–27.
- Lasagabaster, D., 2008. Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal* 1, 30–41.
- Lasagabaster, D., Sierra, J.M., 2009. Language attitudes in CLIL and traditional EFL classes. *International CLIL Research Journal* 1 (2), 4–17.
- Lasagabaster, D., Sierra, J.M., 2010. Immersion and CLIL in English: more differences than similarities. *ELT Journal* 64, 367–375.
- Leung, C., 2005. Language and content in bilingual education. *Linguistics and Education* 16, 238–252.
- Lorenzo, F., 2008. Evaluación de competencias lingüísticas en centros bilingües. Datos de incidencia. *Jornadas sobre Competencia en comunicación lingüística: El currículum integrado de lenguas Carmona, Spain*. 27–28 November, 2008.
- Lorenzo, F., 2010. CLIL in Andalusia. In: Lasagabaster, D., Ruiz de Zarobe, Y. (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training*. Cambridge Scholars Publishers, Newcastle, pp. 2–11.
- Lorenzo, F., Casal, S., Moore, P., 2010. The effects of content and language integrated learning in European education: key findings from the Andalusian sections evaluation project. *Applied Linguistics* 31, 418–442.
- Lorenzo, F., Moore, P., Casal, S., 2011. On complexity in bilingual research: the causes, effects, and breadth of content and language integrated learning – a reply to Bruton (2011). *Applied Linguistics* 32, 450–455.
- Lucietto, S., 2008. A model for quality CLIL provision. *International CLIL Research Journal* 1 (1), 83–92.
- Marsh, D., The relevance and potential of content and language integrated learning (CLIL) for achieving MT+2 in Europe. Retrieved: <http://web.fu-berlin.de/elc/bulletin/9/en/marsh.html>.
- Marsh, H.W., Hau, K.T., Kong, C.K., 2000. Late immersion and language of instruction in Hong Kong high schools: achievement growth in language and non-language subjects. *Harvard Educational Review* 70, 302–346.
- Mehisto, P., 2008. CLIL counterweights: recognising and decreasing disjuncture in CLIL. *International CLIL Research Journal* 1 (1), 93–119.

- Muñoz, C., 2006. Accuracy orders, rate of learning and age in morphological acquisition. In: Muñoz, C. (Ed.), *Age and the Rate of Foreign Language Learning*. Multilingual Matters, Clevedon, UK, pp. 107–126.
- Navés, T., Victori, M., 2010. CLIL in Catalonia: an overview of research studies. In: Lasagabaster, D., Ruiz de Zarobe, Y. (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training*. Cambridge Scholars Publishing, Newcastle, pp. 30–54.
- Netten, J., Germain, C., 2009. The future of intensive French. *Canadian Modern Language Review* 65, 757–786.
- Rothman, J., Guijarro-Fuentes, P., 2010. Input quality matters: some comments on input type and age-effects in adult SLA. *Applied Linguistics* 31 (2), 301–306.
- Ruiz de Zarobe, Y., 2007. CLIL in a bilingual community: similarities and differences with learning English as a foreign language. *View[z]* 16 (3), 47–52.
- Ruiz de Zarobe, Y., 2008. CLIL and foreign language learning: a longitudinal study in the Basque Country. *International CLIL Research Journal* 1 (1), 60–73.
- Ruiz de Zarobe, Y., Lasagabaster, D., 2010. CLIL in a bilingual community: the basque autonomous community. In: Lasagabaster, D., Ruiz de Zarobe, Y. (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training*. Cambridge Scholars Publishing, Newcastle, pp. 12–29.
- San Isidro, X., 2010. An insight into Galician CLIL: provision and results. In: Lasagabaster, D., Ruiz de Zarobe, Y. (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training*. Cambridge Scholars Publishing, Newcastle, pp. 55–78.
- Seikkula-Leino, J., 2007. CLIL learning: achievement levels and affective factors. *Language and Education* 21 (4), 328–341.
- Serra, C., 2007. Assessing CLIL in primary school: a longitudinal study. *The International Journal of Bilingual Education and Bilingualism* 10, 582–602.
- Villareal Olaizola, I., García Mayo, M.P., 2009. Tense and agreement morphology in the interlanguage of Basque/ Spanish bilinguals: CLIL versus non-CLIL. In: Ruiz de Zarobe, Y., Jiménez Catalán, R.M. (Eds.), *Content and Language Integrated Learning: Evidence from Research in Europe*. Multilingual Matters, Bristol, pp. 157–175.