The Language Learning Journal

ISSN: 0957-1736 (Print) 1753-2167 (Online) Journal homepage: https://www.tandfonline.com/loi/rllj20

The effects of CLIL on content knowledge in monolingual contexts

Stephen P. Hughes & Daniel Madrid

To cite this article: Stephen P. Hughes & Daniel Madrid (2019): The effects of CLIL on content knowledge in monolingual contexts, The Language Learning Journal, DOI: 10.1080/09571736.2019.1671483

To link to this article: https://doi.org/10.1080/09571736.2019.1671483

Published online: 29 Oct 2019.

THE LANGUAGE LEARNING JOURNAL
https://doi.org/10.1080/09571736.2019.1671483

The effects of CLIL on content knowledge in monolingual contexts

Stephen P. Hughes and Daniel Madrid

Universidad de Granada, Granada, Spain

ABSTRACT

The adoption of Content and Language Integrated Learning (CLIL) in Spain has affected educational spending and introduced modifications in learning and assessment procedures for students. While positive effects are noticeable in terms of language learning, evidence concerning the impact on non-linguistic subject areas has been less forthcoming. This study, which forms part of the wider MON-CLIL investigation into CLIL instruction in monolingual areas of Spain, aims to contribute to the developing body of knowledge by comparing academic performance in science in primary education and natural science in compulsory secondary education. In addition to comparing school results for a sample of 93 public, charter (semi-private) and private primary students and 139 public and charter secondary students (total = 232), this study also involved a stepwise discriminant analysis with other potential intervening variables in order to determine the extent to which differences might be due to instruction type or to other factors. Significant differences were detected between certain school types in favour of non-CLIL groups in primary education, while differences also existed between public secondary groups in favour of CLIL. These findings seem to indicate that CLIL instruction had no detrimental effects on the science learning of the secondary learners participating in the study.

KEYWORDS

Content and language integrated learning (CLIL); language learning; science; primary education; secondary education

Introduction

Content and Language Integrated Learning (CLIL) focuses by definition on two central areas: the teaching and learning of academic contents from non-linguistic areas (e.g. science, technology, history, etc.), and the foreign/second language (L2) in which non-language subject matter is imparted. In order to gauge the impact of a CLIL program, then, it is necessary to investigate relevant skills and knowledge associated with both of these fields.

Several studies already point to the positive effects of CLIL on communicative and linguistic competence. From the research, we could highlight Coyle (2006), Muñoz (2007), Lasagabaster (2008), and papers included in edited publications by Ruiz de Zarobe and Jiménez Catalán (2009), and Madrid and Hughes (2011).

The influence of CLIL and other forms of bilingual instruction (for clarification in terms see Coyle 2006) on non-language subjects, however, is another matter; here research is scarcer and rather less conclusive. A number of studies, discussed below, include Jäppinen (2005), Bergroth (2006), Seikkula-Leino (2007) in mathematics, along with Madrid (2011) and Wode (1999) in the area of natural and social sciences, as well as Marsh, Hau and Kong (2000) in science, mathematics, history and geography.

The first major wave of schools to adopt CLIL in Spain came over a decade ago. Since then, the number of schools with CLIL-type instruction has increased considerably. It would appear, then,

CONTACT Stephen P. Hughes shughesugr.es

that sufficient time has elapsed in order to make it both possible and desirable to ascertain the potential educational benefits and cost-effectiveness of this type of education.

With a particular focus on natural science, this study examines the results obtained from two related research projects (see Acknowledgements) that have investigated the effects of CLIL on students at the final stages of primary and secondary education in Spain. While other authors (Lorenzo, Casal and Moore 2009; Madrid 2011) have explored this area to a certain degree within a similar geographical and socio-cultural context, the research designs in each case appear to have a series of limitations since the samples involved were not particularly homogenous and fundamental factors may not have been controlled for; these factors may have contributed to variability in results (see Pérez Cañado 2016).

The present investigation attempts to improve upon the statistical design of previous work and has made efforts to address a number of limitations present in earlier research. Among other considerations, outliers have been removed in order to attain a homogenous sample and facilitate comparisons. At the same time, however, it should be made clear that the main data pertaining to learner attainment have been extracted from individual teacher assessments at the end of primary and compulsory secondary stages. While this type of result has limitations in terms of comparability and generalizability, it may provide a useful insight into important transitory stages of academic development which, to a large extent, determine whether or not students move forward to the next academic level.

In addition to focusing on the effect of CLIL programs in relation to the learning of content in non-linguistic subjects taught through the medium of English, this work examines other independent variables which are linked to academic performance and which are often absent from other similar studies. The specific variables in question include verbal intelligence, out-of-class exposure to English, motivation towards learning, anxiety, lack of interest in studying, and 'self-demand' or realistic expectations of one's own learning. This investigation, then, aims to examine potential variability in levels of academic success in non-linguistic subjects arising not only from CLIL but also from other cognitive, contextual or affective variables.

Review of the literature

There is a growing body of evidence which suggests that training in CLIL contexts has a positive effect on a wide range of components related to L2 communicative competence. Examples of significant differences in favour of CLIL may be found for vocabulary acquisition (Jiménez Catalan and Ruiz de Zarobe 2009), listening comprehension (Dallinger et al. 2016), oral production and interaction (Admiraal et al. 2006; Nieto Moreno de Diezmas 2016), reading (Pérez-Vidal and Roquet 2015) and writing (Martínez 2017). Additionally, a number of papers (e.g. Lasagabaster 2008; Villoria et al. 2011) cover a more comprehensive range of skills in which CLIL students outperform non-CLIL learners in several language skills.

Early investigations in terms of academic performance initially suggested that, over time, bilingual education was not detrimental to academic results. Swain and Lapkin (1982), for instance, stated that in the long-term students who received bilingual training obtained similar scores to those who followed monolingual classes in the mother tongue.

More recent research has corroborated this trend and, despite the potentially negative aspects of CLIL (discussed below), several researchers have highlighted the successes or at least have pointed to non-detrimental effects on non-language subjects. Bergroth's (2006) comparative analysis of results by immersion students in university matriculation examinations in Finland, for example, concludes that CLIL did not appear to have a negative impact on the mother tongue or on the learning of non-linguistic subject areas studied through another language; however, immersion did seem to favour improved confidence and performance in L2. Likewise, for Jäppinen (2006), Admiraal et al. (2006), Vázquez (2007) and Wode (1999), CLIL programs are not seen as detrimental to student performance in relation to content knowledge. Similarly, in the domains of cognition and affect, in Rodenhauser and Preisfield's (2015) quasi-experimental study, conducted with upper secondary

biology students in an out-of-school lab in Germany, there were no significant differences between bilingual and non-bilingual groups in terms of cognitive development or in perceived levels of interest and talent either in relation to the L2 or science.

In the subject area of mathematics, Ouazizi's (2016) small-scale quasi-experimental study suggests that CLIL training might even be more effective than traditional forms of education in terms of helping students to attain higher levels of both L2 and mathematical competence. Among the potential factors contributing to this success, Ouazizi mentions the potential underlying connections between learning patterns for language and mathematics, enhanced motivation in learning environments and the existence of teaching approaches that are more effective than those traditionally employed.

In a much larger study, Anghel, Cabrales and Carro (2016) examined performance in mathematics, general knowledge and reading (in L1 Spanish) at the end of primary education, in part to ascertain if there were any negative effects of bilingual education on key subject areas taught in L1. Although no significant differences were detected between CLIL and non-CLIL students, differences were found between students whose parents had educational qualifications below upper secondary education level and those whose parents were qualified above this level.

While several of the above-mentioned studies point to positive effects of CLIL in terms of language learning and, at least, a non-detrimental effect in relation to content learning, other research conducted provides evidence to suggest that results, particularly in non-language subjects, are not always positive. For example, Marsh et al.'s (2000) large-scale, longitudinal study conducted in Hong Kong found that while instruction through English in non-subject areas had a moderately beneficial effect on Chinese and English achievement, it had a small negative effect on mathematics and a very large negative impact on science, history and geography. It is important to mention here, perhaps, that the sample of the study included only students in late-immersion education (secondary level), and in this particular context, English was seen as a particularly difficult language for learners.

In Europe, several studies which incorporate primary level or primary and secondary level students present somewhat contrasting results. Jäppinen's (2005) quasi-experimental study in Finland with 669 students between 7 and 15 years of age, found that while younger students had difficulties with abstract scientific topics and that CLIL environments were initially more demanding, over time, CLIL had a beneficial effect on learners' cognitive development. Also in Finland, in Seikkula-Leino's (2007) study, CLIL students in mathematics were more likely to be 'achievers' (i.e. their achievement was in line with their level of intelligence); however, there were significantly higher numbers of overachievers in non-CLIL groups. In the case of Spain, test results in Madrid's (2011) study with participants from CLIL and non-CLIL science classes showed that public primary school learners had not developed their linguistic proficiency sufficiently and were thus at a potential disadvantage compared with monolingual students studying in the mother tongue. Some studies, then, point to less than optimum levels of performance among primary students who study non-language subject areas in L2.

In addition to these results in performance, further issues may arise. In Otwinowska and Foryś' (2017) study in upper primary education in Poland, primary CLIL learners experienced negative affect and intellectual helplessness. Concerns also arise in relation to teaching practices and among the teaching professionals themselves. In Rodríguez-Sabiote et al. (2018), a number of surveyed in-service teachers reported that the teaching and learning processes were more time-consuming in CLIL classes and that there was a greater need for repetition, emphasising and simplifying of explanations in order for the program to be successful. In the same study, teachers admitted to having limitations in their own language proficiency and recognised that content teaching was constrained as a result.

Concerning classroom practice, Esquierdo and Almaguer (2012) point to the fact that CLIL science teachers often focus on language development rather than on science literacy, while Moore and Lorenzo (2015) highlight the lack of available materials for CLIL content teachers in general. Furthermore, the uptake of CLIL programs in schools has created new difficulties and workloads for L2 and

CLIL teachers alike, as higher levels of collaboration are required in addition to increased hours of training (see Dale, Oostdam and Verspoor, 2017). CLIL, then, is not without its challenges, not only in terms of performance and affect, but also in relation to teacher constraints, practices and other additional burdens.

While there seems to be evidence to suggest that CLIL favours L2 development, certain discrepancies in the research do appear in relation to non-language subjects. Several studies point to the non-detrimental effects, others to possible affective or cognitive benefits and yet others which indicate negative results, particularly when learners have had smaller periods of exposure to this type of education. The scant research on the effects on content knowledge, coupled with the fact that available scholarship tends to focus on single educational levels (e.g. specific primary or secondary levels) and that available results are inconclusive would appear to justify further study into performance in CLIL programs in non-language subject areas at more than one level.

The study

This study forms part of the larger MON-CLIL[1] research project into Content and Language Integrated Learning in monolingual communities in Spain. This particular investigation examines the effects of CLIL in science at the end of primary education and natural science at the end of compulsory secondary education and the research questions for this investigation are as follows:

(1) Are there significant differences between CLIL and non-CLIL students' school results in science at the end of primary education and natural science at the end of compulsory secondary education?
(2) Are there significant differences between students from public, private and privately owned state-funded charter schools in science at the end of primary education and at the end of compulsory secondary education?
(3) Which of the following additional variables contribute towards explaining the differences between performance in science between CLIL and non-CLIL students: school, socio-economic level, verbal reasoning ability, persistence in study, level of anxiety, level of interest, self-demand, weekly hours of exposure to English?

The subsections below will provide details of the sample, instruments and data collection procedures and analysis.

Sample

The sample for this study was obtained from a larger population of 472 students from six primary (a total of 10 groups) and seven secondary schools (a total of 16 groups) in Andalusia (see Table 1). These students took part in a wider research project into the effects of CLIL in terms of language and non-language subjects. The present investigation was concerned only with students in the subject area of science in primary education and natural science in secondary education. Additionally, efforts were made to ensure there was homogeneity between CLIL and non-CLIL students through the matching of learners by means of preliminary tests in verbal intelligence, motivation and levels of English (discussed below). To this end, outliers (students with lowest or highest scores in the preliminary tests) were removed so that higher levels of homogeneity and, hence, comparability could be guaranteed. Furthermore, those students who did not complete all parts of the study were also eliminated from the study sample. A total of 232 students remained after this process (Table 2).

Table 1. Number and characteristics of participating classes.

	Public bilingual	Public non-bilingual	Private bilingual	Charter non-bilingual	Total
Primary	4	3	1	2	10
Secondary	7	6	0	3	16

Table 2. General characteristics of participants in the study.

Students	Primary ($n = 93$)	Secondary ($n = 139$)
Male	37	66
Female	58	73
Rural	47	59
Urban	46	80
Public school	47	101
Private school	23	0
Charter school	23	38
Non-bilingual	38	74
Bilingual	55	65

The first group in the study were finishing their final year in primary education (ages 11–12), while the second was in the process of completing the final stage of compulsory secondary education (ages 15–16). It is perhaps worth mentioning at this point that CLIL in primary and secondary education in the public sector is generally limited to a maximum of four non-linguistic subject areas (primarily history, but also subjects such as science, music and physical education), and a minimum of 50% of class time is conducted in L2. In addition, as occurs in this study, in some public schools, CLIL groups exist alongside non-CLIL groups within the same school. In contrast, private CLIL schools are not limited to the number of subjects imparted in English (with the exception of L1), and usually spend more time in the non-linguistic subject areas working in the foreign language, with most if not all of the subject completed in L2.

It should also be stated that in Spain, students from fully private and semi-private, grant-aided charter schools often obtain better academic results than public school students. This fact is consistently evidenced in central government data, for example, MECD (2017), which shows levels of successful student completion of all subjects in compulsory education range from 89.6% in private schools, 74.6% in charter schools and 68.6% in public schools; the report also shows higher proportions of immigrant and special needs students in public schools. This fact is likely to have some effect on overall student outcomes, although this difference would also appear to be due to the more favourable socio-economic backgrounds of families (Mancebón and Muñiz 2007). Similarly, as Escardíbul and Villaroya's (2009) study indicates, virtually the same family traits (including socio-economic and cultural make-up) influence the families' preference for charter and private schools over public schools.

In our study, the majority of students attended public schools ($n = 148$) and the partially state-funded charter schools ($n = 61$), while a smaller number participated from the private education sector ($n = 23$). This distribution according to school type somewhat reflects the current situation in Spain (see MECD 2017: 423), where the majority of the national population in compulsory education (i.e. non-special or adult education) attend public schools (59.4%), a smaller number go to charter schools (22.4%) and a minority attend totally private institutions (5.8%).

Variables and instruments

Three groups of variables are present in this study, including the following:

- Dependent variable: scores in science in primary and secondary education
- Independent variables: program type (CLIL or non-CLIL); school type (public, private, charter)
- Moderating variables: socio-economic level, verbal reasoning skills, motivation/persistence in study, anxiety, level of interest, levels of self-demand, exposure to English outside class.

A total of three instruments were used to measure the moderating variables. The first test employed was the verbal reasoning test taken from the Factorial Assessment of Intellectual Abilities (Santamaría et al. 2016). This test has two different versions that have been adapted to the sixth grade in primary education and fourth grade in secondary education. In both cases, students were required to choose the

correct option in a multiple-choice test, answering as many questions as possible in the space of five minutes.

Student motivation was measured using Pelechano's (1994) test. This instrument includes 35 items and estimates: (a) desire to work and self-esteem (10 items); (b) anxiety in exams (9 items); (c) lack of interest in studying (9 items); and (d) realistic personal expectations of self or 'self-demand' (7 items).

The questionnaire which measured out-of-class exposure to English was based on the diary devised by Sundqvist and Sylvén (2014). This instrument required students to reflect upon their exposure to the L2 outside the class and to report on the number of hours of exposure per week (e.g. through television, digital gaming, etc.).

In addition to these tests, scores in science and natural science were recorded for each student. As required in the Andalusian education system, these scores range from 1 to 10 points. It is necessary to mention again at this point that scores are not based on a standardised assessment, but on the evaluation of the individual teachers in each class; they do, however, represent scores that are taken into consideration in enabling students to move on to the next educational stage.

Statistical procedures

Data obtained from the different instruments mentioned above were analysed using the 21.0 version of SPSS. Once the mean scores and standard deviations were calculated, *t*-tests were performed in order to compare the scores between CLIL and non-CLIL students in the area of science. In addition, Cohen's *d* was calculated in order to quantify effect size; here, we base our determination of small, medium and large effects on our review of Cohen (1988), Plonsky and Oswald (2014) and Rosenthal (1996). Following these procedures, a stepwise discriminant analysis was performed in order to see which variables were significantly useful to discriminate between CLIL and non-CLIL students.

Results

The subsections below provide results and discussions for the main research questions in terms of differences between CLIL and non-CLIL students in science in primary and secondary education. Differences based on school-type and differences in other potential areas are also examined. In each bivariate analysis, Cohen's *d* is considered only where significant differences are present.

Results based on type of program

Table 3 shows the overall differences in performance between all CLIL and non-CLIL students, as well as differences for primary and secondary groups.

When primary and secondary groups are considered together, the mean CLIL group score is slightly higher than the non-CLIL score, but the difference is not significant. In the case of primary education, while again there is no significant difference in performance in general between CLIL and non-CLIL

Table 3. Bivariate analysis for CLIL and non-CLIL primary and secondary groups.

Group	*n*	Mean	SD	Cohen's *d*	*p* value
Non-CLIL and CLIL primary and secondary groups					
Non-CLIL	112	6.70	1.926	N/A	.120
CLIL	120	7.09	1.932		
Non-CLIL and CLIL primary groups					
Non-CLIL	38	7.97	1.49	N/A	.078
CLIL	55	7.36	1.70		
Non-CLIL and CLIL secondary groups					
Non-CLIL	74	6.04	1.79	−0.42	.013
CLIL	65	6.86	2.09		

groups, scores are slightly higher for students who receive instruction in the mother tongue. As discussed below, higher levels of performance in primary education are to be expected from learners studying content and abstract concepts in their own language. In contrast to the primary groups, CLIL students in secondary education outperform non-CLIL pupils, a difference which is statistically significant. The Cohen's *d* (−0.42) value indicates that belonging to one group or the other has a medium size effect. These differences require further analysis and more discrete levels of comparison between school types (discussed below).

Results based on type of program and school

This section presents the bivariate analyses based on school type (public, private and charter) and type of program (CLIL and non-CLIL). There are no data for non-CLIL groups in private education because all groups were CLIL; similarly, there are no CLIL groups in charter primary and secondary education because these schools did not implement CLIL education. Table 4 provides data for primary schools.

In the case of public primary schools, we found no significant difference between scores in non-CLIL (mean = 6.86) and CLIL (mean = 7.12) groups. Likewise, no significant differences were found between public CLIL and private CLIL, on the one hand, or between public non-CLIL and private CLIL on the other. The latter finding follows other previously mentioned studies, which show small but non-significant differences in favour of non-CLIL groups in primary education. Significant differences do arise, however, between public non-CLIL and charter non-CLIL groups and between public CLIL and charter non-CLIL; in both comparisons, Cohen's *d* indicates a large to very large effect (−1.37 and −1.21, respectively), confirming that students in the non-CLIL charter schools perform better in science, even against students in the private CLIL schools.

It would be plausible to attribute these differences to the variable of socio-cultural and socio-economic background rather than, directly, 'type of school' per se. Thus, we can see a clear difference between non-CLIL students in charter schools (relatively higher socio-economic status) who obtain much higher scores than non-CLIL students from public schools (relatively lower socio-economic status). However, the non-CLIL students in charter schools also outperform the CLIL learners from private schools, who would typically be considered to have the higher socio-economic status. Cohen's *d* (−0.83) here indicates a large effect. The implication, then, is that primary students may attain higher levels in science in non-CLIL instructional settings even when socio-economic status may be slightly lower. This finding would be in line with previously mentioned studies that showed better content results in non-CLIL contexts at primary level.

Table 4. Bivariate analysis for primary school types.

Group	*n*	Mean	SD	Cohen's *d*	*p* value
Public CLIL and non-CLIL primary groups					
Public non-CLIL	15	6.86	1.80	N/A	.64
Public CLIL	32	7.12	1.75		
Public non-CLIL and private CLIL primary groups					
Private CLIL	23	7.69	1.60	N/A	.16
Public non-CLIL	15	6.86	1.80		
Public non-CLIL and charter non-CLIL primary groups					
Public non-CLIL	15	6.86	1.80	−1.37	.001
Charter non-CLIL	23	8.69	0.55		
Public CLIL and charter non-CLIL primary groups					
Public CLIL	32	7.12	1.75	−1.21	.000
Charter non-CLIL	23	8.69	0.55		
Charter non-CLIL and private CLIL primary groups					
Charter non-CLIL	23	8.69	0.55	−0.83	.008
Private CLIL	23	7.69	1.60		
Public CLIL and private CLIL primary groups					
Public CLIL	32	7.13	1.78	N/A	.224
Private CLIL	23	7.69	1.60		

Table 5. Bivariate analysis for secondary school types.

Group	*n*	Mean	SD	Cohen's *d*	*p* value
Public CLIL and non-CLIL secondary groups					
Public non-CLIL	36	5.75	1.67	−0.58	.007
Public CLIL	65	6.86	2.09		
Secondary public non-CLIL and charter non-CLIL groups					
Public non-CLIL	36	5.75	1.67	N/A	.17
Charter non-CLIL	38	6.31	1.37		
Public CLIL and charter non-CLIL secondary groups					
Public CLIL	65	6.86	2.09	N/A	.18
Charter non-CLIL	38	6.31	1.87		

Table 5 presents results for the different types of school and programs at secondary level.

In contrast to primary groups, there is a significant difference in scores between public secondary CLIL and non-CLIL students; here, Cohen's *d* (−0.58) indicates a medium effect. Public CLIL students also marginally outperform charter non-CLIL learners, although the difference is not significant. There is also no significant difference between non-CLIL students in public and charter schools. The importance of the variable 'socio-economic status' seems to diminish here as there are no significant differences between students in the public schools (both CLIL and non-CLIL) and those in charter schools. However, it should be noted that there were no data for private schools, where students would typically have the highest socio-economic background. The significance of socio-economic status and its effect or otherwise on content learning in CLIL versus non-CLIL programs at secondary level thus merits further investigation.

Discriminant analysis

A series of stepwise discriminant analyses were conducted to determine which of the moderating variables exerted greater levels of influence on the differences detected between CLIL and non-CLIL groups in primary and secondary education. These discriminant analyses were performed for: (a) CLIL versus non-CLIL groups in general; (b) CLIL versus non-CLIL groups in primary education; and (c) CLIL versus non-CLIL groups in secondary education. These analyses identify the variables that contribute to distinguishing between the groups considered ($p < .05$) and allow us to discard variables that have little bearing on group distinctions.

In relation to the first of these discriminant analyses, only three variables significantly discriminated between the CLIL and non-CLIL groups. These variables were: (a) rural or urban school setting ($p = .000$), (b) socio-economic level ($p = .000$), and (c) weekly hours of exposure to English ($p = .000$). These three variables appeared to influence the scores obtained by the different groups of students in this study.

The results of the second analysis indicated that no variables significantly discriminated between CLIL and non-CLIL groups in primary education. However, in the third analysis, relating to students in secondary education, the variables discriminating between CLIL and non-CLIL groups were: (a) rural or urban settings ($p = .001$); (b) score in natural science ($p = .000$); (c) self-demand ($p = .000$); and (d) weekly hours of exposure to English ($p = .000$).

Essentially, the combination of the variables for the entire sample, on the one hand, and those identified for secondary education, on the other, can be used to predict whether a student belongs to a CLIL or non-CLIL group. From the data available, no further inferences can be made in terms of differences between groups.

Discussion

We have to reiterate that the performance results in this study stem from individual teacher assessments at the end of primary and compulsory secondary education and, in that sense, we cannot say with any security that one particular system favours another in terms of the acquisition of content

knowledge. What we can tentatively conclude, however, is that academic success at the end of these two educational stages may be affected to a greater or lesser extent by whether learners are in CLIL groups or not, and that certain tendencies found in this study mirror findings from previous investigations.

In primary education we have seen that learners in non-CLIL classes obtained marginally, but not significantly, better end-of stage scores in science than CLIL groups (7.97 for non-CLIL, 7.36 for CLIL: $p = .078$). This is similar to Anghel et al.'s (2016) investigation into results in mathematics, general knowledge and reading, which also indicated higher scores – but not significantly so – for non-CLIL learners.

In secondary education, public CLIL students significantly outperformed both public and charter non-CLIL learners. This finding differs from Madrid's (2011) study, in which no significant differences were found between CLIL and non-CLIL students in non-language subjects, and contrasts sharply with Marsh et al.'s (2000) study which showed a large detrimental effect on science learning among students of late-immersion instruction in secondary education. There are several possible explanations for these differences. Firstly, the groups of students in this particular study were not late-immersion students and had received CLIL education from the early stages of primary education. This may account for a certain reversal in trends, given that it generally takes students up to six or seven years to develop an acceptable level of Cognitive Academic Language Proficiency which enables them to deal with more complex and demanding academic tasks (see Cummins 1984). One of the significant differences observed was between public CLIL and public non-CLIL learners in secondary education (see Table 5). In this particular context, we would have expected similar results at best rather than clear outperformance by CLIL students. This leads us to ask whether or not there may still be hidden factors behind this success, particularly in those cases in which CLIL is a deliberate choice of students. We could surmise that the voluntary opting for CLIL education might indicate higher levels of motivation or, indeed, contribute to a classroom culture that is more conducive to success.

Other factors that seem to be linked to science achievement include weekly hours of exposure to English, rural versus urban setting and socio-economic level. While weekly hours of exposure and socio-economic level might seem to have a natural correlation with the fact that a student is in a bilingual group or not, the settings (urban versus rural) might not be immediately apparent. It is true, however, that the majority of bilingual schools in Andalusia are situated in major towns and cities; there is also evidence, seen for example in Diaz (2017), that there are lower levels of receptiveness towards L2 learning in rural areas as well as a scarcity of resources for bilingual learning, particularly in smaller schools. This lack of receptiveness might well affect learner motivation not only in terms of language learning but also in content learning through L2; this point however, goes beyond the current scope of this paper and would require further investigation.

In terms the discriminating factor of 'weekly hours of English', it should be noted that it is common in Spain for learners in CLIL programs to participate in external examinations for English (e.g. Cambridge, Trinity, etc.). The extent to which this extra support might further improve CLIL students' level of English is, perhaps, self-evident; the degree to which this could improve content level beyond that of peers in non-CLIL classes, however, is not. With regards to socio-economic level, we can see a general tendency at primary level for better results from private and charter schools over public schools, regardless as to whether or not the students are in CLIL classes. Yet by the time students reach the end of secondary education, there appear to be no significant differences between the science scores of students in public school and those of students in charter schools. Where we do see a significant difference is between the CLIL and the non-CLIL groups in public schools. These findings do suggest that learners in CLIL groups are not at a disadvantage in their science learning when compared to non-CLIL learners; indeed, within the public schools and at secondary level, they may even be advantaged. Again, we need to stress the caveat that the science scores gathered here were those given by individual teachers at the end of an educational stage, but these are the results which are important contributing factors in determining promotion to the next educational level.

Conclusions

In this study, we have examined differences between student performance in science in primary and natural science in secondary education in CLIL and non-CLIL programs, taking into account a number of contextual variables. In general terms, there was no significant difference between the science scores of primary school students who had studied in the L1 and those who had studied in the L2 as part of a CLIL program, although mean scores were slightly higher for the non-CLIL group. This indicates that possible detrimental effects of CLIL education on science learning at primary level were not seen. In secondary education, CLIL students overall received significantly better natural science scores than non-CLIL students ($p = .013$); in particular, CLIL students significantly outperformed non-CLIL students in public schools. Again, these results suggest that any detrimental effects of CLIL on science learning are minimal.

In terms of school type (public, charter and private), there were no significant differences between public and private CLIL primary schools, although differences were found between non-CLIL public and non-CLIL charter schools. In the latter case, the standard of living of families who send their children to charter schools is likely to be higher and this might be a key factor in results (see Madrid and Barrios 2018), but it is also possible that other factors related to the functioning of charter schools are involved. These differences, however, disappear by the end of compulsory secondary education.

The examination of other moderating variables has also shed some light on results obtained. In general terms, rural or urban school settings, socio-economic level and weekly hours of exposure to English appear to play a role. As students advance, however, external factors, such as socio-economic level, appear to become less important predictors of participation in CLIL programs, and other factors, such as self-demand, start to become important.

The final conclusion for this study, then, is that several potentially intervening variables are at play and these may affect the success of students studying science in the target language. The implementation of CLIL in the primary classroom may initially have a slightly negative effect on students' scores in content areas; this is perhaps to be expected, given the fact that CLIL learners are obviously at a linguistic disadvantage when they begin their studies in L2. However, in this study, this situation appears to be mitigated by the time students reach secondary education.

Further investigation needs to be conducted in order to establish this trend conclusively, but this study certainly confirms other research that shows the tendency for lower content scores for primary CLIL students giving way to similar, or even higher, content scores as non-CLIL learners by the end of compulsory secondary education. Given the multiple benefits that CLIL programs are likely to afford students (cognitively, socially, academically and professionally), the short-term lowering of performance could be both understood and tolerated as a natural part of the development of future plurilingual citizens. Nevertheless, it is important for teachers, particularly at the primary stage, to be continually aware of the potential negative consequences of lower levels of performance in content areas as a result of CLIL instruction.

Finally, having recognised the specific limitations of this study, particularly in terms of teacher-dependent assessments, we would suggest that future avenues of investigation might incorporate a longitudinal dimension which traces student development from primary through secondary education with standardised tests. We consider that this type of evaluation would be particularly useful if comparisons between public CLIL and non-CLIL examined more fully any additional factors that may lead students to opt for L2 content instruction over mainstream L1 education.

Note

1. For more on MON-CLIL, see http://monclil.com/index.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Spanish Ministry of Economy and Competitiveness, under Grant FFI2012-32221, and by the Junta de Andalucía, under Grant P12-HUM-23480.

ORCID

Stephen P. Hughes http://orcid.org/0000-0001-6457-8512

References

Admiraal, W., G. Westhoff and K. de Bot. 2006. Evaluation of bilingual secondary education in the Netherlands: students' language proficiency. *English Educational Research and Evaluation* 12, no. 1: 75–93. DOI:10.1080/13803610500392160.

Anghel, B., A. Cabrales and J.M. Carro. 2016. Evaluating a bilingual education program in Spain: the impact beyond foreign language learning. *Economic Inquiry* 54, no. 2:1202–23. DOI:10.1111/ecin.12305.

Bergroth, M. 2006. Immersion students in the matriculation examination three years after immersion. In *Exploring Dual-Focussed Education. Integrating Language and Content for Individual and Societal Needs*, ed. S. Björklund, K. Mård-Miettinen, M. Bergström and M. Södergard, 123–43. Vaasa: University of Vaasa. http://www.uwasa.fi/materiaali/pdf/isbn_952-476-149-1.pdf.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edition. New York, NY: Lawrence Erlbaum Associates.

Coyle, D. 2006. Content and language integrated learning: motivating learners and teachers. *Journal Scottish Languages Review* 13: 1–18. https://www.scilt.org.uk/Portals/24/Library/slr/issues/13/SLR13_Coyle.pdf.

Cummins, J. 1984. Wanted: a theoretical framework for relating language proficiency to academic achievement among bilingual students. In *Language Proficiency and Academic Achievement*, ed. C. Rivera, 21–38. Clevedon: Multilingual Matters.

Dale, L., R. Oostdam and M. Verspoor. 2017. Searching for identity and focus: towards an analytical framework for language teachers in bilingual education. *International Journal of Bilingual Education and Bilingualism*, 21, no. 3: 366–83. DOI:10.1080/13670050.2017.1383351.

Dallinger, S., K. Jonkmann, J. Holm and C. Fiege. 2016. The effect of content and language integrated learning on students' English and history competences – killing two birds with one stone? *Learning and Instruction* 41: 23–31. DOI:10.1016/j.learninstruc.2015.09.003.

Díaz-López, M.R. 2017. Análisis de la implantación del bilingüismo en centros educativos públicos en contextos rurales de la provincia de Málaga desde un enfoque de educación inclusiva. Doctoral thesis, Málaga: UMA Editorial.

Escardíbul, J.M. and A. Villaroya. 2009. The inequalities in school choice in Spain in accordance to PISA data. *Journal of Education Policy* 24, no. 6: 673–96. DOI:10.1080/02680930903131259.

Esquierdo, J.J. and I. Almaguer. 2012. Developing science literacy for bilingual learners: a framework for effective learning. *Teacher Education and Practice* 25, no. 1: 38–51.

Jäppinen, A.K. 2005. Thinking and content learning of mathematics and science as cognitional development in Content and Language Integrated Learning (CLIL): teaching through a foreign language in Finland. *Language and Education* 19, no. 2: 148–69.

Jäppinen, A.K. 2006. CLIL and future learning. In *Exploring Dual-Focussed Education: Integrating Language and Content for Individual and Societal Needs*, ed. S. Björklund, K. Ma°rd-Miettinen, M. Bergström and M. Södergard, 22–37. Vaasa: Vaasan Yliopiston Julkaisuja. http://www.uva.fiurl/materiaali/pdf/isbn_952-476-149-1.pdf.

Jiménez Catalán, R.M. and Y. Ruiz de Zarobe. 2009. The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL instruction. In *Content and Language Integrated Learning: Evidence from Research in Europe*, ed. Y. Ruiz de Zarobe and R.M. Jimenez Catalan, 81–92. Bristol: Multilingual Matters.

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

Lorenzo, F., S. Casal and P. Moore. 2009. The effects of content and language integrated learning in European education: key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics* 31, no. 3: 418–42. DOI:10.1093/applin/amp041.

Madrid, D. 2011. Monolingual and bilingual students' competence in social sciences. In *Studies in Bilingual Education*, ed. D. Madrid, and S. Hughes, 195–222. Bern: Peter Lang.

Madrid, D., and E. Barrios. 2018. CLIL across diverse educational settings: comparing programs and school types. *Porta Linguarum* 29: 29–50.

Madrid, D. and S. Hughes, eds. 2011. *Studies in Bilingual Education*. Bern: Peter Lang.

Mancebón, M.J. and M.A. Muñiz. 2007. Private versus public high schools in Spain: disentangling managerial and program efficiencies. *The Journal of the Operational Research Society* 59, no. 7: 892–901. DOI:10.1057/palgrave.jors.2602427.

Marsh, H.W., K. Hau and C. Kong. 2000. Late immersion and language instruction in Hong Kong high schools: achievement growth in language and non-language subjects. *Harvard Educational Review*, 70, no. 3: 302–46.

MECD. 2017. *Informe 2017 sobre el Estado del Sistema Educativo: Curso 2015-2016*. Madrid: Secretaría General Técnica.
Moore, P. and F. Lorenzo. 2015. Task-based learning and content and language integrated learning materials design: process and product. *The Language Learning Journal* 43, no. 3. DOI:10.1080/09571736.2015.1053282.
Muñoz, C. 2007. CLIL: some thoughts on its psycholinguistic principles. *Revista Española de Lingüística Aplicada* Extra 1: 17–26.
Martínez, A.C. L. 2017. Analysis of the effect of CLIL programs on the written competence of secondary education students. *Revista de Filología* 35: 169–84.
Nieto Moreno de Diezmas, E. 2016. The impact of CLIL on the acquisition of L2 competences and skills in primary education. *International Journal of English Studies* 16, no. 2: 81–101. DOI:10.6018/ijes/2016/2/239611.
Otwinowska, A. and M. Foryś. 2017. They learn the CLIL way, but do they like it? Affectivity and cognition in upper-primary CLIL vlasses. *International Journal of Bilingual Education and Bilingualism* 20, no. 5. DOI:10.1080/13670050.2015.1051944
Ouazizi, K. 2016. The effects of CLIL education on the subject matter (mathematics) and the target language (English). *Latin American Journal of Content and Language Integrating Language* 9, no. 1: 110–37. DOI:10.5294/laclil.2016.9.1.5.
Pelechano, V. 1994. Prueba MA. *Análisis y Modificación de la Conducta* 20: 71–2.
Pérez Cañado, M. 2016. From the CLIL craze to the CLIL conundrum: addressing the current CLIL controversy. *Bellaterra Journal of Teaching and Learning Language and Literature* 9, no. 1: 9–31. DOI:10.5565/rev/jtl3.667.
Pérez-Vidal, C. and H. Roquet. 2015. CLIL in context: profiling language abilities. In *Content-based Language Learning in Multilingual Educational Environments*, ed. M. Juan-Garau and J. Salazar-Noguera, 237–55. Educational Linguistics, vol. 23. Berlin: Springer, Cham.
Plonsky, L. and F.L. Oswald. 2014. How big is big? Interpreting effect sizes in L2 research. *Language Learning* 64, no. 4: 878–912. DOI:10.1111/lang.12079.
Rodenhauser, A. and A. Preisfeld. 2015. Bilingual (German-English) molecular biology courses in an out-of-school lab on a university campus: cognitive and affective evaluation. *International Journal of Environmental and Science Education* 10, no. 1: 99–110. DOI:10.12973/ijese.2015.233a.
Rodríguez-Sabiote, C., D. Madrid, J.L. Ortega-Martín and S.P. Hughes. 2018. Resultados y conclusiones sobre la calidad de los programas AICLE en España. In *Influencia de la Política Educativa de Centro en la Enseñanza Bilingüe en España*, ed. J.L. Ortega, S. Hughes and D. Madrid, 141–59. Madrid: MECD.
Rosenthal, J.A. 1996. Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research* 21, no. 4: 37–59.
Ruiz de Zarobe, Y. and R. M. Jimenez Catalan, eds. 2009. *Content and Language Integrated Learning: Evidence from Research in Europe*. Clevedon: Multilingual Matters.
Santamaría, P., D. Arribas, J. Pereña and N. Seisdedos. 2016. *EFAI. Evaluación Factorial de las Aptitudes Intelectuales*. Madrid: TEA Ediciones.
Seikkula-Leino, J. 2007. CLIL learning: achievement levels and affective factors. *Language and Education* 24, no. 1: 328–41. DOI:10.2167/le635.0.
Sundqvist, P. and L.K. Sylvén. 2014. Language-related computer use: focus on young L2 English learners in Sweden. *ReCALL* 26, no. 1:3–20. DOI:10.1017/S0958344013000232.
Swain, M. and S. Lapkin. 1982. *Evaluating Bilingual Education*. Clevedon: Multilingual Matters.
Vázquez, G. 2007. Models of CLIL: an evaluation of its status drawing on the German experience: a critical report on the limits of reality and perspectives. *Revista Española de Lingüística Aplicada* 1: 95–111. https://dialnet.unirioja.es/servlet/articulo?codigo=2575498.
Villoria, J., S. Hughes and D. Madrid. 2001. Learning English and learning through English. In *Studies in Bilingual Education*, ed. D. Madrid and S. Hughes, 157–94. Bern: Peter Lang.
Wode, H. 1999. Language learning in European immersion classes. In *Learning Through a Foreign Language. Models, Methods and Outcomes*, ed. J. Masih, 16–25. London: Centre for Information on Language Teaching and Research.