Foreign language competence and content and language integrated learning in multilingual schools in Catalonia: An ex post facto study analysing the results of state key competences testing.

* Dr. Josep Coral is a lecturer at the Faculty of Education, Autonomous University of Barcelona. Address: *Physical Education Teaching Unit, Department de Didàctica de l'Expressió Musical, Plàstica i Corporal, Facultat de Ciències de l'Educació, Edifici G6, despatx 154, 08193 Bellaterra (Cerdanyola del Vallès) Spain* Tel. +34 93 586 31 93 (Email: Josep.Coral@uab.cat)

Dr. Teresa Lleixà is a lecturer in physical education at the Faculty of Education, University of Barcelona. Address: *Departament de Didàctica de l'Expressió Musical i Corporal, Facultat d'Educació, Campus Mundet, Edifici Llevant, 3a planta, despatx 318, Passeig del Vall d'Hebron 171, 08035 Barcelona, Spain* Tel. +34 93 403 51 42 (Email: teresa.lleixa@ub.edu)

Dr. Carles Ventura is a lecturer at the National Institute of Physical Education of Catalonia, Barcelona, Address: *INEFC, Department of Health and Sport Sciences, Avinguda de l'Estadi, 12-33 (Anella Olímpica), 08038 Barcelona, Spain* Tel. +34 425 54 45 (Email: carlesventura@gencat.cat)

* Corresponding author.

Foreign language competence and content and language integrated learning in multilingual schools in Catalonia: An ex post facto study analysing the results of state key competences testing.

Abstract

The member states of the European Union have funded many initiatives supporting the teaching and learning of foreign languages. Content and language integrated learning is one of the experimental language programmes that have been introduced in Catalonia, in the north-east of Spain. The aims of this study are to analyse the results achieved by students on the state test of English language competence during the period 2009–2012 at (1) 1175 Catalan primary schools, (2) a purposive subset of 85 primary schools and (3) a school of reference, which applies a programme integrating physical education and language. The analysis indicates that the improvement of English competence occurs regardless of whether schools have participated in the experimental English language programmes. It is also showed that slightly better results can be observed in schools that apply the content and language integrated approach and a notable improvement in the school of reference.

Keywords:

Foreign language competence
CLIL
Key competences testing
Physical Education
Multilingualism

1. Introduction

In 1995 the European Commission's White Paper on Education and Training *Teaching and learning – Towards the learning society* included as a goal to "develop proficiency in three European languages" (European Commission 1995, 1). Thus, the European Union encouraged language learning and proposed that "it is necessary to make proficiency in at least two foreign languages at school a priority" (13). Since then, most members of the European Union have funded numerous initiatives in support of teaching and learning foreign languages (Dalton-Puffer 2011; Lorenzo, Casal and Moore 2011) and the strength of such proposals has been reinforced by the inclusion of "Communication in foreign languages" in the Key Competences for Lifelong Learning (European Parliament 2006). Moreover, in 2011, the Civil Society Platform on Multilingualism recommended conducting further research on language education pedagogy and on the impact of language testing from a multilingual point of view, meaning L1 plus two foreign languages. In the same document the Platform named Content and Language Integrated Learning (CLIL) as one of the approaches to language education that "need to be explored, both to increase motivation and effectiveness" (Civil Society Platform on Multilingualism 2011, 16). In fact, according to Coyle, Hood and Marsh (2010), since its adoption in 1994 the CLIL approach has been consistently promoted.

1.1. CLIL implementation in European countries

It should be noted that countries such as the Netherlands and Spain, amongst others, have made a considerable investment in CLIL implementation (Dalton-Puffer 2011) and that English is, to a large extent, the target language. With regard to language learning, "there is increasing evidence that CLIL programmes are more successful in developing foreign language competence than traditional language classes" (Ioannou Georgiou 2012, 501). This assertion is supported by Lasagabaster (2008), Navés (2009), and Lorenzo et al. (2009), while Coyle (2007, 558) advocates "unifying a range of research opportunities: scientific research and classroom enquiry, top-down and bottom-up approaches, qualitative and quantitative". Cenoz, Genesee and Gorter (2013, 258) insist on "conducting research that is generalizable, meaningful and useful". Other studies such as Sylvén (2013) discuss the discrepancies in CLIL outcomes in four European countries while Bruton (2011) expresses his doubts about the analysis and conclusions of some CLIL research.

Note that, whereas the foreign language establishes its own pace when taught in foreign language classes, CLIL programmes are usually planned as content lessons of different subjects taught in the target language. CLIL thus complement traditional language lessons because they offer additional exposure to the target language. With regard to learning outcomes, Dalton-Puffer (2011, 186) points out that "most of the research on outcomes is in the area of attainment in the CLIL language", usually among students attending the same school. However, research on the effects of CLIL on school outcomes is always challenging. Faubert (2009) identified two difficulties in measuring the impact: data availability and inference. The former refers to the relationship between the existence of data and the researcher's ability to identify the impact of a particular policy. The latter takes into consideration the difficulty of measuring the impact of the programmes that have been implemented without a 'control & treatment group'. We should also remember that if national tests are implemented they test the same school-age year after year and therefore always examine different students.

1.2. The testing of communication in foreign language

School evaluation systems should take, as a point of reference, the European Reference Framework of key competences for lifelong learning (European Parliament 2006) to assess the different kinds of essential knowledge, skills and attitudes included in each competence. Faubert (2009) reports that the evaluation systems applied by school establishments to measure student learning vary considerably among European countries. However, she observes that "the majority of approaches to evaluation concerned with outcomes rely on the results obtained by pupils in national tests and examinations, as a means to evaluate the performance of individual schools" (Faubert 2009, 9). In these evaluations, it is important to consider not only the levels of performance attained by each school, but also the equity of results within schools. Particularly, key competences are assessed using a "single" or "separate" format (Prats 2009). The former generally means a test that is administered just once, at the end of compulsory education. Most of the Nordic countries adopt this format. With the latter format, a clear distinction is made between primary and lower-secondary education. This is the format for Germany, Holland and Spain, amongst others. Despite the limitations of this type of test, Cook and Weaving (2013) affirm that standardised tests can contribute to the assessment of key competences if they include questions with "structure and content that reproduce real-life contexts authentically", "multiple steps

requiring a chain of reasoning and a range of competences" and "a range of formats allowing responses that require different competences" (43).

Prats (2009) observes that competences in language and mathematics are evaluated across the EU and some countries include foreign language competences. By the same token, Gordon et al. (2009) report that communication in a foreign language is commonly assessed in national standardised tests. In general, the foreign language component of such tests is based on the Common European Framework of Reference for Languages (CEFR). This is the case of Austria, for example, where the standards for language competence are built on the curriculum (European Commission 2012a). A different case is Cyprus, where foreign language assessment is more knowledge-based than competence-based. Similarly, in Germany the tests assess subject-specific competences (Gordon et al. 2009).

However, most of our information about the foreign language proficiency of students in Europe comes from secondary education and is the work of the First European Survey on Language Competences (European Commission 2012b), which evaluated students from 16 countries. The Final Report of the survey states that many educational systems show high levels of achievement. Nevertheless, with regard to foreign language, in six educational systems at least 20% of students do not achieve A1 in one or more than one skill. But what is most important about this survey is the contextual information collected to make more productive comparisons between language policies and language teaching methods.

1.3. Foreign language programmes and school testing: the case of Catalonia

Catalonia, in the north-east of Spain, is a bilingual community where Catalan and Spanish are used equally thanks to an immersion programme initiated by the Catalan government during the 1980s. Currently, these two languages coexist with others as a result of recent heavy immigration. However, the opportunity to use English in everyday life remains limited and, in contrast to northern European countries, the degree of exposure to English outside school is very low (Sylvén 2013).

Foreign language policies have changed dramatically in Catalonia as a result of the international convergence that is part of globalisation, the Catalan commercial and industrial tradition and the understanding that children and young adults need more opportunities. The 2009 Education Law of Catalonia (Llei 12/2009) provided a solid

basis for educational reform and for a strengthening of multilingualism at schools by mandating a high standard in the acquisition of foreign languages.

Although a few early CLIL projects were conducted in Catalonia prior to 2004 (Marsh 2012), the generalized implementation of CLIL came about as a result of the 2005 *Pla Experimental de Llengües Estrangeres* (Resolution EDC/1329/2005) (Experimental Plan for Foreign Languages, hereafter, PELE). During the period 2006–2012, a total of 273 CLIL-oriented PELE projects were implemented in primary education throughout Catalonia and teachers involved in such projects were able to receive training in both Catalonia and abroad (Lorenzo and Piqué 2013). Unfortunately, the severe financial crisis that impacted Catalonia starting in 2008 led to drastic cuts in government funds and most of the teacher training abroad programmes had disappeared by 2011.

In Catalonia, at the end of primary education (i.e., at age 10–11 years) students are assessed by a test which is designed and validated by a council (hereafter, ECES) which acts as a consultancy division of the Catalan Department of Education and which is charged with analysing and assessing all the pre-university stages of the Catalan education system (Decret 305/1993). Testing procedures are tightly controlled by the ECES, with tests administered and marked at schools by external personnel. These tests are intended to assess the level of achievement of certain key competences commonly associated with mainstream school subjects. It is a formative and learning-oriented process which does not determine access to secondary education but provides schools and parents with meaningful information. The four components of primary school testing cover Catalan, Spanish, Foreign Languages and Mathematics, reflecting the importance attached to language achievement by the Catalan government in response to repeated observations that school success is often hampered by limited language proficiency (Stanat et al. 2012).

Schools are categorised by the Inspectorate of Education of Catalonia, which distinguishes between three categories, A, B and C intended to quantify the level of educational challenge that each school faces (Moral 2012). The three specific criteria that are used to determine this (Resolution ENS/906/2014) are disadvantaging contexts, specific educational needs and the proportion of immigrant children in the student population. From 2009 to 2013, a total of 1,173 level B schools participated in the key competences test. This group range covers a wide variety of educational institutions that

may differ considerably. Recently, some schools have dropped from level B to level C, and the number of schools in this latter group has increased with the emergence of newly deprived areas as a result of the financial crisis. On the other hand, the chief feature of schools categorised as level A is that they tend to be located in wealthy areas.

1.4. A Physical Education-in-CLIL programme as a point of reference

Although Physical Education (hereafter, PE) is often mentioned as a subject included in CLIL programmes, specific research on potential benefits of PE-in-CLIL is scarce. However, of the several examples of such programmes in Europe (e.g. Machunsky 2008; Rottmann 2007), one has been underway since 2008 at a level B school in Catalonia. The school's PE-in-CLIL programme has been validated through research and (like all public schools in Catalonia) the school also has been participating in the key competences testing. These two conditions made the school an ideal school of reference (hereafter, SoR) for the present study not only because its results could be compared with the other PELE programmes but also because it would constitute a felicitous context for purposive sampling. In addition, this particular PE-in-CLIL programme has been the subject of extensive research (Coral 2010, 2012, 2013a, 2013b, 2013c; Coral and Lleixà 2013, 2014, 2016), whose results provide clear proof that the programme fully meets the conditions required (Navés 2009; Florit 2013) for the effective implementation of CLIL programmes, to wit:

a) "CLIL programmes support and respect learners' L1 and home culture" (Florit 2013, 128). Although the use of the CLIL language is encouraged at all times in the SoR's PE-in-CLIL programme, learners are allowed to use their L1 as well as code-switch (Coral 2012).

b) "CLIL programmes have multilingual or bilingual teachers that master students' L1" (Florit 2013, 128). In the SoR, teachers are multilingual and the process of learning in PE was supported by language teaching techniques as detailed elsewhere (Coral 2013a).

c) "CLIL programmes are part of a bilingual/trilingual/multilingual integrated curriculum" (Florit 2013, 129). The SoR decided to foster a long-term CLIL programme without segregating students from mainstream classes. The programme was focused on PE, where the target language of instruction is included in its multilingual long-term

plan along with Catalan and Spanish. In the sixth year of primary education the time exposure is increased by three hours per week through CLIL.

d) "CLIL programmes are long-term projects and applied by stable teaching staff" (Florit 2013, 129). In the SoR, the subject teacher involved in the project, the foreign teachers that supported the CLIL programme and the school board are the same people who started the project in 2007.

e) "CLIL programmes count on the support and involvement of parents" (Florit 2013, 130). In the SoR, parent representatives were well informed about and fully supportive of the programme (Coral 2012).

f) "CLIL programmes are developed with the joint effort of all parties involved" (Florit 2013, 130). The SoR's CLIL programme has the support of the teaching staff and school board. The Sport Research Group of the University of Barcelona's Faculty of Teacher Training also participated in the research (Coral 2012).

g) "CLIL programmes have a teaching staff trained in the content, the language and the appropriate methodology" (Florit 2013, 130). All PELE-CLIL programmes have been applied by multilingual trained teachers. The content teachers receive language training and both content and language teachers participate in in-service teacher training programmes (Lorenzo and Piqué 2013).

h) "CLIL programmes start with high expectations and are under a continuous evaluation and revision" (Florit 2013, 131). In 2010, the SoR's main medium-term goal was to significantly reduce the percentage of students that didn't achieve A1. As described in recent publications (Coral 2012; Coral and Coral 2013, 2014), the PE-in-CLIL programme has been under evaluation for a period of three years through an Action Research process and student outcomes have been assessed through the national key competence test system.

i) "CLIL programmes are provided with quality materials that have been designed *ad hoc*" (Florit 2013, 132). Specific PE-in-CLIL materials for this programme (Coral 2013a, 2013b, 2013c) were created, improved and validated.

2. Aim and research questions

The utility of conducting retrospective studies to determine the advantage of CLIL and to provide policy-makers with useful data has been suggested by Dalton-Puffer and Smit (2013) and Pérez-Cañado (2011). However, to our knowledge, no study has yet investigated the relationship between the results of a state key competences test, national-level plans to promote foreign language acquisition and CLIL programmes.

The aims of this study are to analyse the results of the state test of English language competence in Catalonia during the period 2009–2012, and then, within that context, to quantitatively analyse the results of the SoR. Specifically, the study explores the following research questions (RQ):

RQ1: In the results of the state test of English language competence in the sixth year of primary school, did the schools that applied PELE programmes obtain better results than those that did not?

RQ2: Considering the PELE programmes, did the schools that applied CLIL-based programmes obtain better results than others that did not? If so, what subjects were taught through CLIL? And what what was the exposure time of students to the target language?

RQ3: Did the SoR, which applied a PE-in-CLIL programme, obtain better results than the rest of the schools analysed? If so, what characteristics of the SoR CLIL programme might explain this difference?

We answer these RQs by:

- describing the evolution of the official foreign language test results in level B schools in Catalonia from academic years 2010 to 2012;
- describing the evolution of the foreign language test results of a purposive sample of schools that carried out a PELE programme over the same period;
- describing the evolution of the foreign language test results of a purposive sample of schools classified according their participation or non-participation in the PELE project and a CLIL programme;
- identifying in the purposive sample the subjects that were taught through CLIL;
- analysing the foreign language test results of the SoR compared with other level B schools and the purposive sample.

3. Method

This study used an ex post facto descriptive research design to analyse the results of testing in foreign language competence in Catalonia.

3.1. Participants

The analysis is based on data from a state-wide key competences test for sixth-year students carried out in Catalonia from 2010 to 2012. The participants in the study were students from 85 primary schools categorised as level B schools. One of these schools was the SoR, which also served as the context for a purposive sampling procedure. The total number of pupils for each academic year was 3,923 (2009-2010), 3,856 (2010-2011) and 3,883 (year 2011-2012). In the SoR, the number of students per year was 43 (2009-2010), 29 (2010-2011) and 24 (2011-2012).

3.2. Instrument

Data were gathered through the foreign language tests administered and marked by the Catalan Ministry of Education which assessed reading and listening comprehension. In particular, the test comprised closed items about two short listenings and two short readings. In both cases students needed to demonstrate their ability to identify the main idea, and locate and deduce specific and also non-explicit information. The results are organised in three levels: low, medium and high. In the low level the mean of the student's answers was under the threshold of achievement, in the medium level it crossed the threshold and in the high level it clearly surpassed it. Although the levels cannot be matched precisely with CEFR standards because the student's writing and speaking competences are not tested, any student who receives a low-level grade has a level that may be regarded as equivalent to pre-A1.

3.3. Procedure and method of analysis

The detailed results of the state competence tests are not in the public domain, although schools are naturally given access to the results obtained by their own students as well as mean grade achieved by the entire test-taking population to facilitate the interpretation of their results. For the purposes of this study, however, the ECES gave us access to this data on the condition that it should remain confidential. Although we examined a period of five years, only the data from the academic years 2009-2010,

2010-2011 and 2011-2012 were comparable because the test criteria were altered in the year 2013-2014. The data were made available to us in an Excel spreadsheet and were later imported to SPSS Statistical Analysis Software.

In the study, a purposive sample criterion was followed. As explained above, a SoR was chosen on the merit of previous research that had demonstrated the success of a CLIL programme and had been widely reported and published in journals (Coral & Lleixà, 2013, 2014, 2016). The SoR is a level B school where 44.2% of the students scored in the low level in the year 2009-2010. Thus, the sample criteria specifically considered level B schools where the percentage of students in the low level ranged between 39% and 49% in year 2009-2010. The resulting sample was of 85 schools including the SoR.

The next step was to identify which of these schools completed a PELE project during the 2009-2012 period. After cross-checking the list of PELE projects with the schools in the sample, we learned that 40 schools had participated in the PELE. Next, we queried the schools to find out how many were applying a CLIL programme with the result that 19 schools answered positively and the two schools that did not answer were withdrawn from the study. This process of sampling provided us with three categories (NO-PELE; PELE-NO-CLIL; PELE-CLIL) that could then be compared with the SoR.

The study is based on the percentages of the students with a low, medium and high score per year and per school, following the sampling criteria described above. We use a descriptive presentation of the overall performance obtained by the students. Data are analysed according to whether a school participates in the PELE-NO-CLIL and PELE-CLIL projects. At the same time, we will present the performance obtained for all the students from level B schools who participated in the testing as well as the results from the SoR.

The inferential statistical test used to analyse the relationship between the participation in each category and the student's results is the statistical chi-square (χ^2) test with a predetermined level of significance of .05 (p-value<.05). Cramer's V is used to determine the strengths of statistical associations. Values above .6 indicate a strong association (Rea & Parker, 1992).

4. Results

With schools sorted into the three categories (NO-PELE, PELE-NO-CLIL and PELE-CLIL), Table 1 shows a slight reduction in the percentage of students in the low-scoring level from year 2009-2010 to year 2011-2012. Nevertheless, the statistical analysis of the relationship between the three categories in each school year regarding scores on the English test does not show significant statistical differences: for year 2009-2010, χ^2 (4)=1.439 (p>.05); for year 2010-2011, χ^2 (4)=3.899 (p>.05); and for year 2011-2012, χ^2 (4)=5.159 (p>.05). On the other hand, significant statistical differences can be observed when, category by category, the relationship between school year and test scores is analysed, even though the association is weak: for NO PELE schools, χ^2 (4)=206.31 (p<.000), Cramer's V=.128; for PELE-NO –CLIL schools, χ^2 (4)=77.94 (p<.000), Cramer's V=.123; and for PELE-CLIL schools, χ^2 (4)=104.86 (p<.000), Cramer's V=.145.

[Table 1 near here]

In relation to all the level B schools in Catalonia, the evolution of results presented in Table 2 shows data that is similar to Table 1, meaning a reduction in the percentage of students in the low level that matches an increase in the percentage of students in the medium level, and stability in the high level. In this case, statistical analysis shows a significant relationship between the school year and test scores (χ^2 (4)=2495.02 p<.000) yet with a weak association (Cramer's V=.123), thus reinforcing the idea of the influence of non-controlled variables, which will be discussed later.

[Table 2 near here]

On the other hand, the results obtained by the SoR (*Table 3* and *Figure 1*) show a much more noteworthy tendency. In this case, the percentage of students in the medium level increases by 21.6% (32.6% in 2009-2010 compared with 54.2% in 2011-2012) in conjunction with a similar (19.2%) decrease in the percentage of students in the low level (44.2% in 2009-2010 compared with 25% in 2011-2012). However, the relationship is not statistically significant (χ^2 (4)=3.498 p>0.05), though this may not be surprising, given the small number of pupils enrolled in the SoR.

[Table 3 near here]

[Figure 1 near here]

With regard to the subjects involved in the PELE-CLIL category, four were identified: Art and Crafts (8 schools), Social and Natural Sciences (5 schools), Mathematics (3 schools) and PE (3 schools). In each subject, students were exposed to content instruction in English for differing lengths of time, with 1 hour of target language instruction in Art and Crafts, 1 hour in Social and Natural Sciences, an average of 1.3 hours in Mathematics and an average of 2.3 hours in PE.

5. Discussion

In answer to RQ1 on the improvement in English competence in Catalonia's level B schools, the results indicate that improvement occurs regardless of whether schools have participated in PELE programmes or not (*Table 2*). In answer to RQ2, slightly better results are obtained in the CLIL schools (*Table 1*), while with regard to RQ3, a marked improvement is observed in the SoR (*Table 3*), which shows better results than the overall test scores of the rest of the sample and level B schools. The factors that seem to influence SoR results will be discussed below.

Within the sample of schools that were carrying out PELE programmes, the reduction in the proportion of low level students was more pronounced in the PELE-CLIL category than in those categories that did not apply any type of PELE programme in the same period. In fact, NO-PELE results were very similar to results in those schools that applied other types of experimental language programmes (PELE-NO-CLIL). This suggests that such small differences might be directly related not with PELE programmes but with other factors such as the type of test, differences in the student population from one year to another and the time of exposure to the PELE programme.

Regarding the PELE-CLIL (*Table 1*) category and the SoR (*Table 3*), the reduction in the percentage of students within the lower range runs parallel to a clear increase in students within the medium level but no gain in the number of students with a high

level. This significantly broader band of students in the medium level is consistent with the conclusions of Dalton-Puffer (2008).

The results for the population of level B schools in general (*Table 2*) show that even though there is an improvement, there is less difference in the reduction of the low level (from 39.37% to 34.27%) compared with the SoR and the sample schools, even though the starting point is lower (44.2% in the SoR, 43.83% in the PELE-NO-CLIL category compared with 39.37%, in Catalonia's level B schools overall). The reduction in the low level is comparatively greater in the SoR than in the rest of the schools. Although the results are not meant to be interpreted as the direct and unique consequence of the implementation of CLIL programmes, the trend supports the conclusions of San Isidro (2011) and Lasagabaster (2008) on bilingual societies where the target foreign language is hardly ever used outside the school setting. It also supports the results of studies conducted in monolingual societies with a very low extramural exposure to English, such as Hungary (Várkuti 2010) and Andalusia (in the south of Spain) (Lorenzo, Casal and Moore 2009).

On the other hand, the nature of the SoR and the sample confirms that, in the Catalan context, CLIL is not a discriminatory teaching approach. This supports the arguments provided by Hüttner and Smit (2014), in opposition to Bruton's critique (2011). In the same way, Catalan CLIL programmes do not replace foreign language teaching, confirming the assertion by Hüttner and Smit (2014, 163) that "CLIL is typically an additional element of FL instruction, and does not replace dedicated language classes". Indeed, the information provided by the schools in the PELE-CLIL category reinforces the fact that in Catalonia CLIL is complementary (Lorenzo & Piqué, 2013). Likewise, the profile of such programmes confirms the variety, in time exposure and subjects involved, reported by Coyle et al. (2010) and Sylvén (2013), amongst others. It seems that two models (Coyle et al. 2010) are present, with either the subject being entirely taught in English (the case of PE and Art and Crafts) or the subject being taught only part of the time in English (Mathematics and Social and Natural Sciences). The co-presence of these two modalities is explained by the fact that in Catalonia, as in all the countries of Europe, the education authorities in question apply the CLIL approach on their own terms (Ruiz de Zarobe 2013). At any rate, according to Ruiz de Zarobe and Celaya (2011) it is almost impossible to ascertain which factor has more influence, the CLIL approach itself or the additional time exposure that it entails.

We agree with Ruiz de Zarobe and Celaya (2011, 211) that "the positive results that were obtained using a CLIL approach could sometimes be attributable to both the educational approach and to the fact that under this approach, students had a higher amount of exposure to the foreign language". However, there is not enough evidence to understand the full extent to which each factor has influenced the language improvement shown in this PELE-CLIL sample. We must also ask how it is that the percentages of achievement demonstrated by students enrolled in NO-PELE and PELE-NO-CLIL are similar.

A further question also arises: why does the PE-in-CLIL programme show better results? Two factors seem to be crucial: time exposure and methodology. Because increasing time exposure is not enough (De Graaff 2007) substantial methodological changes were included in the teaching of PE. Thus, language teaching techniques have been combined with specific PE teaching styles that incorporate tasks that require high order thinking skills. Cooperative learning was also included in order to increase student talking time and language was embedded into the motor activities through meaningful hands-on tasks (Coral and Lleixà 2016) accompanied by the necessary language scaffolding (Coral 2013b, 2013c). Furthermore, it is significant that the PE-in-CLIL programme developed in the SoR takes into consideration all the theoretical assumptions regarding effective teaching performance directed at language acquisition in CLIL contexts described by De Graaff et al. (2007) and also meets the conditions required for the effective implementation of CLIL programmes, as these are explained by Navés (2009) and Florit (2013).

6. Conclusions

Our study confirms the improvement of English competence in Catalonia's level B schools in the period 2009–2012. No significant differences in improvement were detected between schools that did not participate in any type of PELE programme and schools that participated in the PELE-NO-CLIL category. Related to the schools that completed a PELE programme, slightly better results were obtained in schools that implemented a CLIL programme (PELE-CLIL category). Furthermore, a clear difference was observed in our school of reference (SoR), which uses CLIL in PE and rigorously applies the characteristics of effective CLIL programmes as described by different experts. These findings have various implications for educational policies,

suggesting that education authorities should promote multilingual curriculums, ensure that CLIL projects follow the features of successful CLIL programmes and assess the four skills in order to be comparable with the CEFR rubrics. One of the limitations of the present study is the absence of speaking and writing skills in the state test. The year-to-year variability of the students also needs to be controlled for, along with the periods of time in which students have been continuously exposed to a CLIL programme. Despite such limitations and because the study is based on real tests that have been applied by governmental agencies in natural contexts, the findings have ecological validity, at least in the Catalan context. However, further research will determine whether CLIL programmes maintain the same positive tendencies with new testing system which includes the assessment of writing and speaking skills.

7. Acknowledgements

The authors acknowledge the support provided by the Evaluation Council of the Education System of Catalonia (ECES), the project 'Key Competences within Quality Physical Education' R&D Project; Ref. DEP 2012-33296, the research group 'Language and Education' (LED); Ref. 2014 SGR 1190 and the National Institute of Physical Education of Catalonia (INEFC).

8.- References

Coral, J. 2010. "L'aprenentatge de l'anglès a través de l'educació física: el programa 'Mou-te i aprèn.' [English Language Acquisition through Physical Education: The Programme 'Move your body and learn']". *Temps d'Educació,* 39: 149–170.

Coral, J. 2012. "Aprenentatge integrat de continguts d'educació física i llengua anglesa: educació física en AICLE al cicle superior de primària [Physical Education and English Integrated Learning: PE-in-CLIL in the fifth year of Primary School.]." PhD diss., University of Barcelona. Retrieved from http://www.dart-europe.eu/full.php?id=680451.

Coral, J. 2013a. "Physical education and English integrated learning: How school teachers can develop PE-in-CLIL programmes." *Temps d'Educació*, 45: 41–65. Retrieved from http://www.raco.cat/index.php/TempsEducacio/article/view/274635.

Coral, J. 2013b. *Physical Education Year 5. P.E. World 5, Student's book.* Zaragoza: Edelvives - Grupo Editorial Luis Vives.

Coral, J. 2013c. *Physical Education Year 6. P.E. World 6, Student's book.* Zaragoza: Edelvives - Grupo Editorial Luis Vives.

Coral, J., and T. Lleixà. 2013. "Las tareas en el aprendizaje integrado de educación física y lengua extranjera (AICLE). Determinación de las características de las tareas mediante el análisis del diario de clase. [Tasks in PE-in-CLIL. Determining the characteristics of tasks through the analysis of teacher diaries]." *Retos. Nuevas Tendencias en Educación Física, Deporte y Recreación*, 24: 79–84.

Coral, J., and T. Lleixà. 2014. "La enseñanza de la educación física mediante el enfoque educativo CLIL: La resolución de los dilemas profesionales surgidos durante un proceso de investigación-acción. [Teaching PE through CLIL: solving professional dilemmas that emerged during an action research project]." *Movimento* 20 (4): 1423–1445. Retrieved from http://www.seer.ufrgs.br/index.php/Movimiento/article/view/46307.

Coral, J., and T. Lleixà. 2016. "Physical education in content and language integrated learning: successful interaction between physical education and English as a foreign language." *International Journal of Bilingual Education and Bilingualism* 19 (1):108-126. doi:10.1080/13670050.2014.977766.

Bruton, A. 2011. "Is CLIL so beneficial, or just selective? Re-evaluating some of the research." *System* 39 (4): 523–532.

Cenoz, J., F. Genesee, and D. Gorter. 2013. "Critical analysis of CLIL: Taking stock and looking forward." *Applied Linguistics* 35 (3): 243–262.

Civil Society Platform on Multilingualism. 2011. *Policy Recommendations for the Promotion of Multilingualism in the European Union*. Retrieved from http://ec.europa.eu/languages/information/documents/summary-civil-society_en.pdf

Cook, R., and H. Weaving. 2013. *Key Competence development in school education in Europe: KeyCoNet's review of the literature: A summary.* Brussels: European Schoolnet. Retrieved from http://keyconet.eun.org/c/document_library/get_file?uuid=3a7a093c-4c8f-473c-8702-f38ed86bb730&groupId=11028.

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

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

Dalton-Puffer, C. 2008. “Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe.” In *Future Perspectives for English Language Teaching* edited by W. Delanoy, and L. Volkmann, 139–157. Heidelberg: Carl Winter.

Dalton-Puffer, C. 2011. “Content and Language Integrated Learning: From Practice to Principles?” *Annual Review of Applied Linguistics* 31: 182–204. doi: 10.1017/S0267190511000092.

Dalton-Puffer, C., and U. Smit. 2013. “Content and Language Integrated Learning: A research agenda.” *Language Teaching* 46 (4): 545–559. doi:10.1017/S0261444813000256.

De Graaff, R., G. J. Koopman, Y. Anikina, and G. Westhoff. 2007. “An Observation Tool for Effective L2 Pedagogy in Content and Language Integrated Learning (CLIL).” *International Journal of Bilingual Education and Bilingualism* 10 (5): 603–624. doi:10.2167/beb462.0.

Decret 305/1993, de 9 de desembre, de creació del Consell Superior d'Avaluació del Sistema Educatiu. [Decree 305/1993, of 9 November, creating the High Council for the Assessment of the Catalan Educational System], Diari Oficial de la Generalitat de Catalunya (Official Journal of the Government of Catalonia, DOGC) 1837 (1993).

European Commission. 1995. *White Paper on Education and Training Teaching and Learning – Towards the learning society.* Retrieved from http://ec.europa.eu/white-papers/index_en.htm#block_13

European Commission. 2012a. *Education and Training 2020 Work programme. Thematic Working Group ‘Assessment of Key Competences’ Literature review, glossary and examples.* Retrieved from http://ec.europa.eu/education/policy/school/doc/keyreview_en.pdf

European Commission. 2012b. *First European Survey on Language Competences: Final Report.* Retrieved from http://ec.europa.eu/languages/policy/strategic-framework/documents/language-survey-final-report_en.pdf

European Parliament, Council of the European Union. 2006. *Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning. DO L 394 of 30.12.2006.* Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006H0962.

Faubert, V. 2009. *School Evaluation: Current Practices in OECD Countries and a Literature Review. OECD Education Working Paper No. 42.* Paris: OECD. doi: 10.1787/218816547156.

Florit, C. 2013. "Pràctica docent efectiva AICLE [Effective teaching in CLIL]." *Temps d'Educació* (45): 115–141.

Gordon, J., G. Halasz, M. Krawczyk, T. Leney, T. Michel, D. Pepper, E. Putkiewicz, and J. Wiśniewski. 2009. *Key competences in Europe: Opening doors for lifelong learners across the school curriculum and teacher education, CASE Network Reports, No. 87.* Retrieved from http://hdl.handle.net/10419/87621.

Hüttner, J., and U. Smit. 2014. "CLIL (Content and Language Integrated Learning): The bigger picture. A response to: A. Bruton. 2013. CLIL: Some of the reasons why ... and why not. System 41 (2013): 587-597." *System* 44: 160–167. doi:10.1016/j.system.2014.03.001.

Ioannou Georgiou, S. 2012. "Reviewing the puzzle of CLIL." *ELT Journal* 66 (4): 495–504. doi:10.1093/elt/ccs047.

Lasagabaster, D. 2008. "Foreign Language Competence in Content and Language Integrated Courses." *The Open Applied Linguistics Journal* 1: 31–42. doi:10.2174/1874913500801010030.

Llei 12/2009, del 10 de juliol, d'educació [Law 12/2009 of 10 July, on Education], Diari Oficial de la Generalitat de Catalunya (Official Journal of the Government of Catalonia, DOGC) 5422 (2009).

Lorenzo, F., S. Casal, and P. Moore. 2009. "The Effects of Content and Language Integrated Learning in European Education: Key Findings from Andalusian Bilingual Sections Evaluation Project." *Applied Linguistics* 31 (3): 418–442. doi: 10.1093/applin/amp041.

Lorenzo, M.N., and I. Piqué. 2013. "Report and evaluation of the development of CLIL Programmes in Catalonia." *Temps d'Educació,* 45: 143–183.

Marsh, D. 2012. "Content and Language Integrated Learning (CLIL). A Development Trajectory." PhD diss., University of Córdoba. Retrieved from http://www.dart-europe.eu/full.php?id=698655.

Machunsky, M. 2008. *Developing Material for Physical Education Lesson in CLIL*. Munchen: GRIN Verlag.

Moral, J. 2012. *L'autonomia dels centres educatius. Elements per a l'avaluació d'una política educativa* [School autonomy. Elements for the evaluation of an educational policy]. Barcelona: Escola d'Administració de Catalunya.

Navés, T. 2009. "Effective Content and language Integrated Learning (CLIL) Programmes." In *Content and Language Integrated Learning. Evidence from Research in Europe* edited by Ruiz de Zarobe, Y., and R.M. Jiménez-Catalán, 22–40. Clevedon: Multilingual Matters.

Prats, E. 2009. "Avaluació de les Competències Bàsiques a l'educació primària. Perspectiva Europea. [Key competences assessment in primary education. European perspective]." *Quaderns d'avaluació* 13: 11-15. Retrieved from http://csda.gencat.cat/ca/arees_d_actuacio/publicacions/quaderns_avaluacio/quaderns_avaluacio_13/

Pérez-Cañado, M.L. 2011. "CLIL research in Europe: past, present, and future." *International Journal of Bilingual Education and Bilingualism* 15 (3): 315–341.

Rea, L. M., and R.A. Parker. 1992. *Designing and conducting survey research*. San Francisco: Jossey-Boss.

Resolució EDC/1329/2005, de 25 d´abril, de convocatòria de concurs públic per a la selecció de projectes de centres docents, d'educació infantil i primària i d'educació secundària de Catalunya, de titularitat del Departament d'Educació, per participar en un pla experimental de llengües estrangeres. [Resolution EDC/1329/2005 of 25 April announcing public tender for the selection of state school projects, of centres providing early childhood, primary and secondary education to participate in an experimental foreign languages plan], Diari Oficial de la Generalitat de Catalunya (Official Journal of the Government of Catalonia, DOGC) 4375 (2005).

Resolució ENS/906/2014 de 23 d'abril, per la qual es determinen els centres educatius i elsllocs de treballdocents que tenen la consideració de dificultat especial. [Resolution ENS/906/2014 of 23 April determining the schools and teaching jobs that are considered of special difficulty], Diari Oficial de la Generalitat de Catalunya (Official Journal of the Government of Catalonia, DOGC) 6613 (2014).

Rottmann, B. 2006. "Sport auf Englisch: Lerngelegenheiten im." PhD diss., University of Hamburg.

Ruiz de Zarobe, Y. 2013. “CLIL implementation: from policy-makers to individual initiatives.” *International Journal of Bilingual Education and Bilingualism* 16 (3): 231–243. doi: 10.108013670050.2013.777383.

Ruiz de Zarobe, Y., and M.L. Celaya. 2011. “AICLE a través de las lenguas: resultados de la investigación en dos comunidades bilingües [CLIL through languages: results of a research in two bilingual communities].” *Revista Teoría de la Educación: Educación y Cultura en la Sociedad de la Información* 12 (3): 200–214.

San Isidro, X. 2011. “Análisis del modelo CLIL gallego. [Analysis of a Galician CLIL model].” In *AICLE_CLIL_EMILE. Educació plurilingüe: experiencias, research & polítiques* edited by Escobar, C., N. Evnitskaya, E. Moore, and A. Patiño , 327–337. Bellaterra: Universitat Autònoma de Barcelona.

Stanat, P., M. Becker, J. Baumert, O. Lüdtke, and A.G. Eckhardt. 2012. “Improving second language skills of immigrant students: A field trial study evaluating the effects of a summer learning program.” *Learning and Instruction* 22 (3): 159–170. doi: 10.1016/j.learninstruc.2011.10.002.

Sylvén, L.K. 2013. “CLIL in Sweden – why does it not work? A metaperspective on CLIL across contexts in Europe.” *International Journal of Bilingual Education and Bilingualism* 16 (3): 301–320. doi:10.180/13670050.2013.777387.

Várkuti, A. 2010. “Linguistic Benefits of the CLIL Approach: Measuring Linguistic Competences.” *International CLIL Research Journal,* 1 (3): 67–79. Retrieved from http://www.icrj.eu/13/article7.html