Equity in bilingual education: socioeconomic status and content and language integrated learning in monolingual Southern Europe

Journal:	*Applied Linguistics*
Manuscript ID	APPLING-19-11-370.R2
Manuscript Type:	Article
Keyword:	equity, bilingual education, CLIL, socioeconomic status, European language planning, European language policies

Equity in bilingual education: socioeconomic status and content and language integrated learning in monolingual Southern Europe

Previous research has raised concerns that equity may be compromised in Content and Language Integrated Learning (CLIL) education, creating schisms in otherwise fairly egalitarian education systems. In Andalusia (southern Spain), where bilingual education has expanded, this paper aims to analyse the difference between CLIL bilingual education and traditional monolingual education in terms of student equity indicators.

A sample of over 3,800 students representing the four socioeconomic status levels (SES 1 to SES 4), selected by stratified random sampling, was analysed with correlational statistics to determine their performance levels at CLIL and non-CLIL schools, according to their competence in Spanish L1, English L2 and history. Results point to certain egalitarian effects of CLIL education: while a staircase pattern is constantly present in the performance of non-CLIL students (with those from higher social classes obtaining better results), all CLIL students seem to obtain equally high results regardless of their socioeconomic status.

Keywords: equity; bilingual education; CLIL; socioeconomic status; European language planning, European language policies.

Introduction

The emergence of CLIL education and research

Preceded by former elite European schools, the European Commission supported and popularised a new model for second language education that was expected to improve learning deficits revealed by continental language surveys (Beardsmore 1993; De Mejia 2002; Housen 2002; Lorenzo 2007; European Commission 2012). The aim of the new content and language integrated learning (CLIL) model is to extend languages to society at large as part of the political shaping of a new learning citizen different from the alternative model humorously dubbed 'Anglo-American human capitalist' (Dale & Robertson 2006). The rationale behind this was that 'languages are for all' and that only through their real command would it be possible to ensure the compliance of the civil rights established by European laws: the right of free movement of persons, services and capital.

Over the years, EU policymakers have supported and financed the introduction of new vehicular languages in school systems under the CLIL label (Council of Europe 2005, 2006). In doing so, they have embraced immersion models as an alternative to mainstream foreign language programmes that were often inefficient and which, in some cases (such as in this research context), had produced *generations of non-communicators* (Lorenzo et al 2011). As a result, CLIL thrived, not only being introduced by the EU Member States but also worldwide as a default immersion model that supplemented mainstream foreign language programmes in many countries (as regards Australia, see, among others, Turner 2012; as to Japan, see Tsuchiya & Pérez 2015). Around the world, CLIL is now, alongside immersion education, an alternative form of educational bilingualism (concerning equivalent terms for bilingual programmes, see Cenoz et al

2014; Dalton-Puffer et al 2014).

However, the implementation of CLIL programmes has triggered social concerns. The extensive media coverage—in mainstream newspapers such as *El País*—of benchmark bilingual education research soon challenged aspects like the effects on learning, school selection, teachers' language competence and what was mellifluously called the *shadows of bilingualism* (see, for example, Sanmartín 2013; Ballesteros 2015; Martín-Arroyo 2017; Torres 2018a, 2018b). Student admission to the programme was the most sensitive issue of all since any social bias would have defeated the purpose of the new policy. After all, bilingual implementation in the form of CLIL was meant to spread multilingualism across state school networks in line with democratic, egalitarian and inclusive continental language policies.

Besides general research on CLIL, specific aspects like discourse and language classroom description, effects on L2 and L1 competence or new professional roles have been studied (see, for example, the special issues of *International Journal of Bilingual Education and Bilingualism*, in 2019; and *Language Learning Journal*, in 2014). Attention has also been paid to egalitarianism: to the eligibility of students from different walks of life, their overt or covert selection and the gradual attrition of the less privileged (Madrid 2005; Huettner & Smit 2014; Relaño 2015; Pérez et al 2016). Many studies have taken a critical stance, decrying the model as *evangelical* (adopted in a dogmatic manner), *neoliberal* (considering languages as a commodity), *elitist* (catering to a select few) and *segregative* (creating artificial divides in the student population) (see, respectively, Banegas 2011; Relaño 2015; Paran 2013; Broca 2016). As these legitimate critiques have mostly taken the form of argumentations based on anecdotal evidence, it is now urgent to test the actual egalitarianism (or lack thereof) of the new multilingual model in empirical studies.

Thus, the aim of this paper is to contribute to the present state of the art by considering the mutual effects between the socioeconomic profiles of CLIL students and their learning achievements. To this end, the three major learning areas—to wit, Spanish as the mainstream language in society (L1), English as an additional vehicular language at these bilingual schools (L2) and disciplinary content (history, one of the courses taught in the L2 at all bilingual schools)—affected by CLIL in the research context (nine million Andalusians) were analysed. This study employed official data disclosed by the administration by virtue of a research agreement with the Andalusian Education Assessment Agency (AGAEVE).

Socioeconomic status in CLIL programmes

Language research is increasingly more aware of the fact that 'findings concerning bilingualism and the effects of bilingual education are not necessarily transferrable across social class boundaries' (Politzer 1981: 4, in Block 2014: 110). The understanding of the social effect on language competence is ambiguous not only due to the shortcomings of class as an operational variable, but also because 'applied linguists do not have a coherent theory of class' (Butler & Le 2018: 2). Accordingly, sociological research has turned to socioeconomic status (SES) as an adequate measurable proxy for class. SES is an aggregate of indicators relating to material possessions, income, occupation, neighbourhood character, spatial relations, cultural capital and family reproduction modes (see Loos 2000). This explicit social indicator is, therefore, usually considered as one of the major contextual variables in large-scale education surveys such as PISA. More precisely, SES encompasses the following (Butler & Le 2018):

(1) Parent, family and neighbourhood character: parental education, occupation, household income and race and ethnicity.

(2) Parent, family and neighbourhood factors (child-rearing beliefs, parenting styles and efficacy beliefs).

(3) Specific parental beliefs: parental expectations for their children's performance and their perception of their abilities.

(4) Specific parental behaviour: time spent with their children, teaching strategies and the number of books.

With this conceptualisation, SES has proved to be positively related to the English performance of students in monolingual systems (Rascón & Bretones 2018), while the effects on immersion programmes like CLIL have not often been tested. In fact, the scant research performed hitherto in this regard has yielded conflicting results. Two aspects of the competence/SES dichotomy have been considered for CLIL. One strand focuses on the actual performance of students from different social backgrounds, in order to determine whether bilingual education in the form of CLIL is suitable for the least privileged or whether they experience learning deficits. While the intention of the second strand is to discover whether students of all SES levels participate in bilingual programmes or whether the school populations with the lowest SES levels are being side-lined intentionally or unintentionally.

Regarding the first research strand, Anghel et al (2016) noted that low SES levels correlate negatively with achievement in bilingual programmes. They described a clear negative effect on learning the subject taught in English for children whose parents had a low SES as measured by their educational background. More precisely, the bilingual programme had a negative effect on the content exam results of children with parents without higher education. They also observed that, when other variables like teacher expertise were included in the equation, the differences became less significant. In cohorts

receiving 'quality teaching', the differences were neutralised for students whose parents held some sort of degree (Anghel et al 2016).

Similarly, in a series of studies performed in a northern monolingual region in Spain, the authors found significant differences across SES levels (Sanjurjo et al 2017). More precisely, they argued that CLIL students from more disadvantaged backgrounds obtained worse results than their non-CLIL peers of a similar socioeconomic standing. As a matter of fact, socially-disadvantaged students learning content in an additional language were the only ones who did not meet the academic standards of the primary education curriculum in the research sample. This last assertion seriously compromises the programme, both for the learning deficits reported and for its contribution to mounting dropout rates, a sensible issue in Spain, where the dropout rate in secondary education is high compared with other EU Member States. However, social bias was not detected by Rascón and Bretones (2018) or by Admiraal et al (2006), who found no significant differences in achievement between the participating students as regards their SES, although the research context and the educational level (secondary as opposed to primary) were different.

With respect to the second research strand, Van Mensel et al (2019) compared the composition of the CLIL and non-CLIL cohorts involved in a large-scale project undertaken in French-speaking Belgium in terms of a number of background factors. They discovered that bilingual tracks attracted a more privileged student population as measured by SES levels. Nonetheless, they also found that those CLIL strands using English as a vehicular language happened to be less selective than those using Dutch. This was the case both in primary and secondary education. Mediavilla et al (2019) also researched the social composition of the student population in bilingual schooling in yet another monolingual area in Southern Europe (Madrid, Spain), where CLIL programmes

have thrived. Their findings suggest that CLIL enhances the social stratification of the state education system, on the one hand, and confirm cream-skimming of a social nature in bilingual tracks, on the other. In fact, they identified some SES-related variables that correlated with participation in the bilingual strands: fewer immigrant students, the higher educational and occupational level of the parents and a greater availability of cultural items in the household. To this should be added the language screening procedure, introduced in the region of Madrid, consisting of a minimum B1 level certificate requirement for the participation in the bilingual programme at the first stage of secondary education (when students are 12-13 years old). This language filter may exclude less competent second language students who are likely to come from more disadvantaged social backgrounds.

Since contextual factors are key in complex systems and multilingual education is certainly a complex system, we cannot offer any general considerations on social bias in bilingual education (Dalton-Puffer et al 2014). Nevertheless, the implementation of policies pertaining to different starting conditions, recruiting procedures, educational levels and evaluation strategies can result in higher or lower levels of social inclusion.

Research design and methodology

Research context

The bilingual network in Andalusia (Spain) was assessed one decade ago, thus constituting an early evaluation of bilingual programmes under the CLIL label in Europe (Lorenzo & Moore 2009, Lorenzo et al 2011). From its inception, the network has expanded over the years and now incorporates 1,000 schools, 300,000 students and 8,000 teachers, including schools located in more disadvantaged rural and suburban areas.

CLIL in Andalusia is currently regulated by the Order of 28 June 2011 of the Andalusian regional government (Junta de Andalucía 2011). By virtue of this legislation, all bilingual secondary schools are obliged to teach two content subjects in an L2 at every level, and the L2 exposure should amount to 30 per cent of the weekly teaching hours (including L2 as a subject). A whole-school plan is required before permission is granted by the administration for bilingual implementation. This involves a considerable amount of teamwork programming in the form of a language-across-the-curriculum plan, with the involvement of L1 Spanish teachers, L2 English teachers, CEFR qualified content teachers and native language assistants.

Regarding access, students were first allowed to choose between a bilingual or monolingual strand at the onset of the implementation. Nevertheless, as parents perceived the bilingual programme as a feature of quality schooling, the number of applicants exceeded the vacancies. Accordingly, the admission rules were changed in 2011 to avoid mounting parental discontent. Schools were now either fully bilingual or monolingual and the initial in-school split between monolingual and bilingual strands was discontinued. In other words, admission to the bilingual school network is now across the board (with no screening procedures whatsoever), with all enrolled students following CLIL programmes. This was so because of the regional concern that social disadvantages might be transferred to the education system, which features a high social inclusion index (76.71 per cent), seven points above the national average (69.29 per cent) and 15 above other monolingual areas like the state capital (61.82 per cent) (López-Rupérez et al 2019: 8).

Further sociological information on the region may help to understand local CLIL policies. With a population of nine million people, the region's income per capita accounts for 73.8 percent of the national total (€18,219 vs. €24,703), whereas the average

domestic income is a little over half of that of the rest of Spain (€582 vs. €1,036) (Pérez et al 2018). In a country like Spain, with the second lowest language competence level in the European Union (European Commission 2012), major domestic investment has tried to make up for this deficit, mainly through a solid network of costly private extracurricular language schools beyond the means of the more disadvantaged. In regions with a low average domestic income, the state bilingual network was supposed to bridge income gaps, for which reason the first bilingual schools were established in rural areas and less affluent urban districts. Very tellingly, charter schools were not permitted at first to offer bilingual schemes in order to prevent them from attracting affluent students and thus widening educational divides.

On a more positive note, and according to the Organisation for Economic Co-operation and Development (OECD), Spain has high educational equity levels. In PISA 2012, there was a difference of 26 points between the best and worst schools, compared to the 71-point average for OECD countries. Moreover, the effect of SES on educational outcomes is much lower among Spanish students than in the OECD as a whole (Ruiz et al 2018: 86). Therefore, these figures should be borne in mind when interpreting the effects of CLIL.

Research design

Against this educational backdrop, the following research questions were posited:

Q1. How do bilingual schools from different socioeconomic backgrounds (SES levels) perform in second language achievement tests (English) in relation to non-bilingual schools?

To address this question, the correlations between SES and L2 scores in standardised L2

tests were analysed. Despite the ample evidence that CLIL furthers L2 competence (Lorenzo et al 2019), the intention was to discover whether CLIL only benefitted high achievers or exhibited a SES bias. Given the parallels between L2 competence and SES, the study of L2 dynamics in relation to SES should be reviewed.

Q2. How do bilingual schools from different socioeconomic backgrounds (SES levels) perform in first language achievement tests (Spanish) in relation to non-bilingual schools?

To address this question, the correlations between SES and L1 scores in standardised L1 tests were analysed. There is a social concern that CLIL schemes may hamper Spanish as an L1 and that learning deficits occur in the students' mother tongue. We contend that, if this is indeed the case, there may be a bias caused by the students' socioeconomic background as measured by SES levels. Therefore, the aim here is to inform on the impact of bilingual models on L1 competence and related factors like bi-literacy development. Even though enriched L2 bilingual programmes have proved not to affect the L1 (Navarro & López 2019), group arrangements or other organisational factors may lead to different learning conditions and social bias, a major concern among education stakeholders.

Q3. How do bilingual schools from different socioeconomic backgrounds (SES levels) perform in the history course test in relation to non-bilingual schools?

To address this question correlations between SES levels and scores in standardised history tests are observed. At bilingual schools, the fact that history is taught in the L2 has raised concern that students may suffer content learning deficits. Indeed, research is unclear as to whether or not learning content in an L2 is detrimental to students from different social backgrounds, with authors offering conflicting results as noted above (on

CLIL history, see, for example, Dallinger et al 2016).

Q4. How do L2, L1 and history competence levels occur with different SES levels?

The purpose here is to determine if the bilingual programme varies significantly with learning outcomes for students with different SES levels, i.e. the parallels between CLIL implementation and the competence of socially diverse student populations. In order to operationalise the question, the net and percentage-wise differences between non-bilingual and bilingual groups were analysed.

Methodology

As in all the 36 member countries of the OECD, Spanish students sit international tests for L1 competence (PISA and PIRLS), subject-content competence (TIMMS) and institutional L2 tests, run by the European Commission (2012). This international assessment is often supplemented by regional and national tests that monitor progression on an annual basis and inform new policies. In the research context, the Andalusian Education Assessment Agency (AGAEVE) designs and administers annual regional diagnostic tests (*pruebas de diagnóstico* in Spanish) in the final year of primary and compulsory secondary education (involving students aged 12 and 16, respectively). These tests measure L1 and L2 competence, content performance in science for primary education students and in history for compulsory secondary education students. Additionally, the tests include questionnaires on contextual variables, like the participation in bilingual programmes and the SES of students.

As is the case with international tests like PISA and PIRLS, each one of the schools is assigned a SES index. This index is measured globally for schools. According to national and regional legislation (Spanish Ministry of Education, Culture and Sport,

2017; Andalusian Department of Education, 2017), the SES of students is calculated on the following basis:

- Use of information resources (books, press, encyclopaedias, computers, tablets and the Internet) by household members.
- Number of information and communication technology (ICT) devices (computers, tablets, smartphones, smart TVs, e-readers, etc.) at home.
- Number of household members.
- Number of books at home.
- Parents' highest qualification.
- Parents' occupation.

Drawing from this information, the SES of schools is distributed in four quartiles, from SES 1, corresponding to the schools with the least favourable socioeconomic conditions, to SES 4, the highest level. This index is therefore an aggregate figure that serves as a benchmark to describe the overall SES of all the students attending a school.

Furthermore, other contextual variables include the students' gender, the number of times that they have retaken a year, the age at which they began schooling, their level of absenteeism, their immigrant status, the time that they devote to homework, their satisfaction with the school and their school-related use of ICTs. However, this information is not used for calculating the SES index.

This study is based on an analysis of the data gathered through the aforementioned diagnostic tests and stored by the AGAEVE. Thanks to an institutional agreement with this agency, databases were disclosed to the authors of the project funding this research, including the assessment information of bilingual and non-bilingual students and the SES of their schools. Thus, as opposed to most of the research to date, limited by sampling

restrictions, results could be obtained on the basis of large samples representing a national system (over 3,800 students). This allowed for comparisons between bilingual and non-bilingual students who had completed the external evaluations at the end of primary and compulsory secondary education.

In order to analyse the differences between the mean scores obtained in L1, L2 and history by the students in each strand (bilingual and monolingual) depending on their socioeconomic status, we performed a one-way ANOVA test (Appendix 1), prior to a Bonferroni test for multiple comparisons (Appendix 2). The significance level was established at 0.05. Spearman's rho (ρ) was also calculated in order to observe correlations. Finally, a reliability measure based on Cronbach's alpha (α) was also calculated, the result (α=0.73) demonstrating the instrument's consistency. Due to space constraints, only the results for compulsory secondary education students are presented here.

Sample

In the academic year 2016-2017, the AGAEVE's test protocols required a random sample of 184 secondary schools across all provinces and school types (public, charter and private) to perform external evaluations. Twenty-nine of those secondary schools were bilingual. The sample was based on stratified random sampling: all the schools were selected in terms of their regional distribution and the proportional representation of all four SES quartiles (from SES 1 to SES 4). The resulting sample was composed of over 3,800 students. The bilingual students had received CLIL training since primary education, as envisaged by the regional education regulations. The total number of students who sat each one of the tests in June 2017 is shown in Table 1, below. The slight variance from one test to another is simply due to school attendance (tests were

administered on different days).

[Table 1 near here]

Measures

The AGAEVE's external diagnostic tests for compulsory secondary education students measure their knowledge and skills in *Spanish language and literature*, *first foreign language* (English), *mathematics* and *history*. The tests were completed independently over a two-day period (two hours per day, one hour for each test with a half-hour break in between). Beforehand, the students had performed sample introductory activities to prepare them for the actual tests assessing the knowledge and skills that they were expected to have acquired by the end of the stage. They were also required to fill in a background questionnaire, as explained above. All the tests, background questionnaires and scoring guidelines can be consulted online (Lorenzo, 2019). In order to achieve reliability, open-ended sections followed analytical scoring with competence bands based on descriptors, a technique which minimised score variance and ensured reliability in composition scoring. Also, double scoring was used to increase the validity of the L2 test in order to establish correlates with CEFR levels.

Spanish language and literature

Students were tested in two skill areas: reading comprehension and written production. Reading comprehension was measured by means of multiple-choice questions based on two narrative texts and one journalistic text. Written production was assessed on the basis of two open writing tasks: (1) a structured summary (a minimum of 80 words) of one of the narrative texts used in the reading comprehension test, and (2) an argumentative text (a minimum of 150 words) relating to the theme of the journalistic text used in the reading

comprehension test. Assessment was based on task fulfilment, cohesion and coherence, grammatical range and accuracy, and lexical range and accuracy. The results were presented on a 44-point scale for individual skills: reading and writing, in this case. SES factors were then analysed for each skill, as well as for the total score.

First foreign language: English

Students were tested in three skill areas: listening comprehension, reading comprehension and written production. Listening comprehension was tested by means of multiple-choice questions about a recording of a radio talk show. Reading comprehension was measured by means of multiple-choice questions about a series of job advertisements, and written production was evaluated via two open writing tasks: students were required to write (1) a short opinion piece (20-30 words) and (2) a short essay (80-95 words). Assessment was based on task fulfilment, cohesion and coherence, grammatical range and accuracy, and lexical range and accuracy. The contexts for these writing tasks incorporated the respective themes of the listening and reading comprehension tests. The results were presented on a 30-point scale for individual skills: reading and writing, in this case. SES factors were then analysed for each skill, as well as for the total score.

For each exam, the results were subsequently converted to the scale of the Common European Framework of Reference for Languages (CEFR), thus yielding a score distribution for the proficiency levels pre-A1 to B2, according to the multilevel nature of the test.

History

In this test, which was taken in Spanish by both CLIL and non-CLIL students, subjects were assessed in critical text analysis and advanced expository text comprehension

relating to political philosophy and modern and contemporary history (colonisation, invasions, industrialisation, the structure of societies, the organisation of the state, etc.). The results were presented on a 36-point scale. History was selected here in order to study the effects of bilingualism on the content areas and its relation to socioeconomic status. As it is the only compulsory subject covered in an L2 at all bilingual schools, it offers a more balanced overview of bilingual factors.

Results

Q1. SES significance on L2 competence: bilingual vs. non-bilingual schools

As can be seen in Table 2, there are evident competence differences between the bilingual and non-bilingual groups, as would be expected if only because of the higher exposure that CLIL programmes provide. As the line graphs in Figure 1 and Figure 2 clearly show, a recurrent pattern emerges in non-bilingual groups in the form of a staircase. In other words, the higher the SES level, the higher the competence level, until a maximum level of achievement in SES 4. A logical inference from this staircase effect is the decisive impact of SES on L2 performance for non-bilinguals. Moreover, as can be seen in Figure 3, this effect was consistent with all of the skills measured (listening, reading and writing). However, this peculiar stepwise picture did not hold for bilingual groups. Students with a lower SES outperformed others with higher levels, as reflected in the uneven distribution of scores across socioeconomic levels. As a result, the visual representation of each step up in the SES scale is disrupted. This interpretation is supported by the significance levels in the statistical appendix: in non-bilingual groups, all the differences between social levels were significant, whereas in bilingual groups the results were not statistically significant for SES 1 and SES 2 or for SES 3 and SES 4, respectively. Furthermore, a statistically significant value of $\rho=0.345$ was obtained for

Spearman's rank correlation coefficient.

[Table 2 near here]

[Figure 1 and Figure 2 near here]

[Figure 3 near here]

On the other hand, the net difference in competence levels between bilingual and non-bilingual SES 4 groups was very limited (only 0.45 points more, on average), this level apparently benefitting less from the programme. Even though the study design does not allow us to infer the cause, the upper classes might have benefited from a very solid network of private extracurricular language schools and study abroad programmes, for which high-income families have been a loyal customer base, a recurrent theme in the research literature. This may account for the fact that the non-bilingual SES 4 students usually perfect their L2 skills elsewhere.

Q2. SES significance on L1 competence: bilingual vs. non-bilingual schools

For the L1, students were tested for reading and writing, thus providing a snapshot of literacy levels. As already noted, low literacy levels are a serious issue in the research context, especially in terms of poor reading comprehension. Consequently, SES is a clear statistically significant factor in reading skills in non-bilingual environments, as illustrated in Figure 4. Of special note is the performance gap between non-bilingual SES 1 (3.96) and SES 4 (6.78) students in written composition inasmuch as the score of the former was little more than half of that of the latter. Even though the aim here is to describe bilingual education environments, what these results show instead are semi-literate levels in broad sectors of the non-bilingual student population and that class factors in the form of SES variables are crucial correlates. A statistically significant value of $\rho=0.303$ was obtained for Spearman's rank correlation coefficient.

[Table 3 near here]

[Figure 4 near here]

As before, the stepwise pattern in Figure 4 is a representation of competence levels across SES levels in non-bilingual groups and, yet again, CLIL environments seem to cut across them and to be more resilient to social determinants (e.g. no statistically significant differences were found between SES 1 and SES 3-SES 4, as can be seen in the appendix). This picture was consistent across all skills. In both reading comprehension and written production, SES levels correlated with performance scores in the non-CLIL groups, while this effect dwindled in their CLIL counterparts. As can be seen in Figure 4, in the aggregate for production, for instance, CLIL SES 1 (12.29) students obtained better scores than their CLIL SES 2 (10.77) and CLIL SES 4 (11.71) peers.

Q.3 SES significance on history: bilingual vs. non-bilingual schools

The results also highlight the differences between the CLIL and non-CLIL groups as regards content areas, history in this case. The literature has contended that when language competence is practised across the curriculum and scaffolding is incorporated in content courses, learning conditions are enhanced (Nikula et al 2016). As has occurred in other secondary education CLIL settings, in this study CLIL did not affect content learning. No significant differences between the bilingual and non-bilingual strands can be observed in Figure 1 and Figure 2. This goes to show that even when a substantial proportion of the course is taught in an L2, students assimilate advanced history content like that described in the tests above. The fact that it was a history course is relevant too. History relies on advanced language structures to express historical meanings: causation, counterfactuals and complex narrative frameworks (Lorenzo 2017, De Alba et al 2018). The students in bilingual strands seemed to perform at similarly high levels as those in

non-bilingual strands (see mean scores in Table 4). A statistically significant value of ρ=0.256 was obtained for Spearman's rank correlation coefficient.

[Table 4 near here]

But once again, when SES factors are accounted for, the results show unequal distributions. The *staircase effect* persists, proving to be a consistent trait of non-bilingual education. In the CLIL groups, however, the results were even for all the SES levels. Furthermore, while the endpoints in monolingual groups showed a difference of almost one third between SES 1 and SES 4 (14.61 and 21.05, respectively), in the CLIL groups the differences were of less than one full score and even in an inverted order (SES 2 ranked the highest with 17.91 and SES 3 the lowest with 17.32). Another feature illustrated in the graph is not new: the performance of CLIL SES 4 students was lower than that of their non-bilingual peers (i.e. non-CLIL SES 4 students), which may hint at a side effect of bilingual education that merits further interpretation.

Q.4 How do L2, L1 and history competence levels occur with different SES levels?

As has been briefly described above, in spite of the seemingly egalitarian effect of CLIL education, the performance of students from different SES levels may vary in the bilingual programme. If the net scores of the bilingual and non-bilingual groups are compared, the net and percentage-wise differences of the former can be quantified.

[Table 5 near here]

As can be seen in Table 5, the socioeconomic level with the highest CLIL/SES differences is SES 1. This is particularly the case as regards L1, thus proving that CLIL not only does not hinder the L1 competence of students, but also may reinforce it. The SES 2 and SES 3 CLIL students also show meaningful differences with respect to non-CLIL individuals, with averages of 8 and 13 per cent, respectively. However, the history

learning of the SES 3 students was discrete, thus suggesting that the bilingual programme may not have the same effects on students from the higher social orders, especially in content courses.

This assumption is borne out when the differences of the CLIL/non-CLIL SES 4 students are analysed. Differences were not so evident between the two. This gap was small in the L2, while being particularly prominent in history, pointing to a lagging effect of CLIL on SES 4 students. Unfortunately, it is impossible to determine here why the CLIL SES 4 students obtained more discrete results than their non-CLIL peers.

Discussion

Several conclusions can be drawn from the results. The difference that first meets the eye is the *staircase effect* in non-bilingual groups, which proves that in ordinary monolingual schools, performance tends to replicate tiered social structures in a very noticeable manner. The results give credence to the assertion in the critical sociology of language that language structures mirror social structures and point to the social origin of language capital distribution. They also suggest a pronounced *Matthew effect*. The application of this *dictum* to non-bilingual education—the mainstream model—would suggest that students with more language capital (SES4) will gain more, as well as acquiring ample multilingual resources, while those with the least capital (SES 1) will even be deprived of what they have (for a social interpretation of cultural and language capital, see Bernstein 1973; Bourdieu 1992; Blommaert 2010). This is what the results for the non-CLIL groups point to. Furthermore, this stepwise pattern holds for all the competencies measured—linguistic (L1 and L2) and non-linguistic (content courses like history)—which in a way highlights the central role of language in other courses, in this case history.

Of course, emphasis must be placed on the reverse side of the coin, namely, the disruption

of social/performance parallels in bilingual groups. As already observed, no causal explanation can be offered, but the consistency of the correlations between the dependent and independent variables deserves further attention. CLIL implementation seems to transform several aspects of the school experience: the attitudes of the participants, curriculum organisation, teacher dynamics and methodology. This lesson has been learned from the literature: changing the language regime implies changes across the board. Even if admission bias persists, the results seem to show that CLIL creates a liveable atmosphere for lower-class students who, in fact, are the ones who benefit most from the programme.

No language theory would support, however, that the changes reported may occur solely as a result of the language regime at bilingual schools, i.e. due to the fact that two vehicular languages, instead of one, are used. The results point to organisational features that may be distinctive in bilingual schools and which give rise to quality schooling (on quality standards and bilingual education, see Brisk 1998). Of particular note is the inclusion of language-across-the-curriculum tools on which the administration insists as a prerequisite for bilingual implementation and which involve a language focus on all courses. These are genre maps, integrated language units and school language projects, among other tools, which relate to integration, a keystone in CLIL and further bilingual education (Nikula et al 2016). Lessons are planned in a co-operative manner by content and language teachers, and this teamwork brings to light language aspects otherwise unnoticed (Lorenzo & Trujillo 2017, Lorenzo & Meyer 2018). It may be concluded that these ongoing innovations enhance language awareness and, over the past few years, may have reinforced the language competence of those students with less out-of-school literacy support, usually the most underprivileged SES groups.

Likewise, the fact that L1 competence yields similar results across all CLIL SES levels is of particular significance. That bilingual students present similar literacy levels, as shown in the graphs, contrasts with the view held in some school communities that competence in a second vehicular language may increase at the expense of the mother tongue. But it also shows that the intervention of an L2 generates pedagogical dynamics that ultimately benefit overall literacy levels. Indeed, many literacy programmes aimed at the disadvantaged implement procedures similar to the language enhancement techniques employed in integrated lesson plans: Disadvantaged Schools Programme (DSP), Language and Social Power and Write it Right and Reading to Learn (Rose & Martin 2012). That bilingual schooling may result in higher biliteracy levels is a significant leap forward in language planning, not the least for the EU policy objective to reduce the number of low achievers in reading to below 15 per cent by 2020 (Council of Europe 2011).

As already noted, however, this study is only correlational. Causes will surely be multifaceted and of a varied nature and, in order to overcome this paper's limitations, the direction and relevance of each one of them will need to be addressed in future research. They may be related to the concurrence of many success factors whose efficiency has been underscored in the literature: teacher training, the self-worth dynamics of the participants and cohesion in planning and implementation. Inference analyses—beyond the scope of this study—need to be conducted to explore and identify the different causes and their preponderance. Notwithstanding its limitations, this study presents findings that are very important for reducing social differences in educational outcomes by means of language policies favouring multilingual education. It could be that multilingual policies facilitate arrangements that ultimately lead to the reduction of socioeconomic differences as determinants of competence. Egalitarianism is a holy grail in contemporary education

and the fact that language innovation of the sort described above disrupts conventional social determinants would meet two essential language policy targets.

Finally, bilingual CLIL-type programmes are now widespread and will continue to gain ground for the aforementioned reasons. Nonetheless, education systems have taken different paths with regard to their implementation. In some, admission to secondary bilingual strands depends on obtaining a pass in Cambridge KET or PET tests, thus giving rise to elite bilingual strands. In others, totally bilingual schools have been created that admit all students, irrespective of level tests and SES profiles. In yet others, the intention has been to implement a middle-of-the-road solution with the organisation of flexible bilingual strands based on parental decision (on educational quality and egalitarian educational structures, see Hattie 2009: 89; Cummings & Bain 2014).

In the context of this study, there is concern that differentiated streams may cause a schism in the state education system. Teaching staff feel that a second vehicular language might make school life worse for underachievers by creating a sink or swim atmosphere in which the less able will struggle. In line with this assertion, egalitarianism in bilingual education may backfire, proving to be detrimental to those students who have not reached L2 instruction competence, that is, those who find it beyond their means to learn in a language for which they simply do not have the sufficient competence level (Rolstad 2015). As has been the case in other immersion models, one-size-fits-all bilingual programmes may be more harmful to the less advantaged (Dalton-Puffer et al 2014).

Conclusion

In light of the results, we take a personal stance against screening procedures that select students beforehand. The results show that bilingual education may provide advantages

for underprivileged SES students that outweigh L2 competence factors. Also, since we believe that uninformed universal bilingual schooling across the board would be prejudicial to the programme and many individuals, we espouse flexible arrangements in all-bilingual schools where a number of factors determine the courses, the amount of L2 contact and the methods used. Many classroom-based decisions need to be made in order to achieve more ecological CLIL programmes. With this study we hope to have cast some light on the synergies of class and competence dynamics, thus encouraging further research clearly showing the effects of alternative variables that permit to identify causation.

Acknowledgments

This research was funded by the Spanish Ministry of Economy & Competitiveness (research grant ref. FFI2016-74885-P EU ERDF) and by the Spanish Ministry of Science (research grant ref. FPU18/02673).

We would like to express our gratitude to the three anonymous reviewers for their comprehensive feedback. We would also like to thank Conecta13 (https://conecta13.com, University of Granada) for supporting this study.

Appendix

Appendix 1. ANOVA test results

Table 6. P-values and effect size in the one-way ANOVA test
Dependent variable: total test scores

Independent variable: SES level

		P-value	Effect size
L2 - English	Non-bilingual	.000	.160
	Bilingual	.000	.044
L1 - Spanish	Non-bilingual	.000	.095
	Bilingual	.000	.032
History	Non-bilingual	.000	.116
	Bilingual	.715	.001

Appendix 2. Bonferroni tests' results

L2 English

Table 7. Multiple comparisons. SES levels in non-bilingual groups. L2 English

Dependent variable: total test scores

Bonferroni test

(I) SES	(J) SES	Mean difference (I-J)	Standard error	Sig.	95%. Confidence interval	
					Lower bound	Upper bound
SES 1	SES 2	-2.041*	.359	.000	-2.99	-1.09

	SES 3	-3.214*	.374	.000	-4.20	-2.23
	SES 4	-7.415*	.343	.000	-8.32	-6.51
SES 2	SES 1	2.041*	.359	.000	1.09	2.99
	SES 3	-1.173*	.368	.009	-2.14	-.20
	SES 4	-5.373*	.336	.000	-6.26	-4.49
SES 3	SES 1	3.214*	.374	.000	2.23	4.20
	SES 2	1.173*	.368	.009	.20	2.14
	SES 4	-4.201*	.352	.000	-5.13	-3.27
SES 4	SES 1	7.415*	.343	.000	6.51	8.32
	SES 2	5.373*	.336	.000	4.49	6.26
	SES 3	4.201*	.352	.000	3.27	5.13

Table 8. Multiple comparisons. SES levels in bilingual groups. L2 English

Dependent variable: total test scores

Bonferroni test

(I) SES	(J) SES	Mean difference	Standard error	Sig.	95%. Confidence interval

		(I-J)			Lower bound	Upper bound
SES 1	SES 2	-.585	.641	1.000	-2.28	1.11
	SES 3	-3.304*	.584	.000	-4.85	-1.76
	SES 4	-2.256*	.588	.001	-3.81	-.70
SES 2	SES 1	.585	.641	1.000	-1.11	2.28
	SES 3	-2.718*	.533	.000	-4.13	-1.31
	SES 4	-1.671*	.537	.011	-3.09	-.25
SES 3	SES 1	3.304*	.584	.000	1.76	4.85
	SES 2	2.718*	.533	.000	1.31	4.13
	SES 4	1.047	.468	.153	-.19	2.28
SES 4	SES 1	2.256*	.588	.001	.70	3.81
	SES 2	1.671*	.537	.011	.25	3.09
	SES 3	-1.047	.468	.153	-2.28	.19

L1 Spanish

Table 9. Multiple comparisons. SES levels in non-bilingual groups. L1 Spanish
Dependent variable: total test scores

Bonferroni test

(I) SES	(J) SES	Mean difference (I-J)	Standard error	Sig.	95%. Confidence interval	
					Lower bound	Upper bound
SES 1	SES 2	-1.200*	.219	.000	-1.78	-.62
	SES 3	-1.636*	.229	.000	-2.24	-1.03
	SES 4	-3.486*	.211	.000	-4.04	-2.93
SES 2	SES 1	1.200*	.219	.000	.62	1.78
	SES 3	-.436	.226	.325	-1.03	.16
	SES 4	-2.286*	.208	.000	-2.84	-1.74
SES 3	SES 1	1.636*	.229	.000	1.03	2.24
	SES 2	.436	.226	.325	-.16	1.03
	SES 4	-1.850*	.218	.000	-2.43	-1.27
SES 4	SES 1	3.486*	.211	.000	2.93	4.04
	SES 2	2.286*	.208	.000	1.74	2.84

	SES 3	1.850*	.218	.000	1.27	2.43

Table 10. Multiple comparisons. SES levels in bilingual groups. L1 Spanish

Dependent variable: total test scores

Bonferroni test

(I) SES	(J) SES	Mean difference (I-J)	Standard error	Sig.	95%. Confidence interval	
					Lower bound	Upper bound
SES 1	SES 2	1.522*	.430	.002	.39	2.66
	SES 3	-.464	.383	1.000	-1.48	.55
	SES 4	.576	.388	.825	-.45	1.60
SES 2	SES 1	-1.522*	.430	.002	-2.66	-.39
	SES 3	-1.986*	.363	.000	-2.95	-1.03
	SES 4	-.946	.368	.062	-1.92	.03
SES 3	SES 1	.464	.383	1.000	-.55	1.48
	SES 2	1.986*	.363	.000	1.03	2.95
	SES 4	1.040*	.313	.005	.21	1.87

SES 4	SES 1	-.576	.388	.825	-1.60	.45
	SES 2	.946	.368	.062	-.03	1.92
	SES 3	-1.040*	.313	.005	-1.87	-.21

History

Table 11. Multiple comparisons. SES levels in non-bilingual groups. History

Dependent variable: total test scores

Bonferroni test

(I) SES	(J) SES	Mean difference (I-J)	Standard error	Sig.	95%. Confidence interval	
					Lower bound	Upper bound
SES 1	SES 2	-2.546*	.353	.000	-3.48	-1.61
	SES 3	-3.280*	.367	.000	-4.25	-2.31
	SES 4	-6.435*	.343	.000	-7.34	-5.53
SES 2	SES 1	2.546*	.353	.000	1.61	3.48
	SES 3	-.735	.364	.261	-1.69	.23
	SES 4	-3.890*	.339	.000	-4.78	-2.99

SES 3	SES 1	3.280*	.367	.000	2.31	4.25
	SES 2	.735	.364	.261	-.23	1.69
	SES 4	-3.155*	.354	.000	-4.09	-2.22
SES 4	SES 1	6.435*	.343	.000	5.53	7.34
	SES 2	3.890*	.339	.000	2.99	4.78
	SES 3	3.155*	.354	.000	2.22	4.09

Table 12. Multiple comparisons. SES levels in bilingual groups. History

Dependent variable: total test scores

Bonferroni test

(I) SES	(J) SES	Mean difference (I-J)	Standard error	Sig.	95%. Confidence interval	
					Lower bound	Upper bound
SES 1	SES 2	-.516	.655	1.000	-2.25	1.21
	SES 3	.073	.596	1.000	-1.50	1.65
	SES 4	-.268	.605	1.000	-1.87	1.33
SES 2	SES 1	.516	.655	1.000	-1.21	2.25

	SES 3	.589	.548	1.000	-.86	2.04
	SES 4	.248	.558	1.000	-1.23	1.72
SES 3	SES 1	-.073	.596	1.000	-1.65	1.50
	SES 2	-.589	.548	1.000	-2.04	.86
	SES 4	-.341	.487	1.000	-1.63	.95
SES 4	SES 1	.268	.605	1.000	-1.33	1.87
	SES 2	-.248	.558	1.000	-1.72	1.23
	SES 3	.341	.487	1.000	-.95	1.63

References

Admiraal, W., **G. Westhoff**, and **K. de Bot**. 2006. 'Evaluation of bilingual secondary education in The Netherlands: Students' language proficiency in English,' *Educational Research and Evaluation* 12/1: 75-93.

Andalusian Department of Education. 2017, May 10. *Administrative resolution of 10 May, by the Andalusian Department of Education, regulating the diagnostic tests for Primary Education and Compulsory Secondary Education in the academic year 2016/17.* Retrieved from: https://www.adideandalucia.es/normas/resoluc/Resolucion10mayo2017EvaluacionesFinalesMuestral.pdf

Anghel, B., **A. Cabrales**, and **J. M. Carro**. 2016. 'Evaluating a bilingual education programme in Spain: The impact beyond foreign language learning,' *Economic Inquiry* 54/2: 1202-1223.

Ballesteros, X. 2015, March 11. 'Las claves de una enseñanza bilingüe de calidad,' *El Mundo*. Retrieved from: https://www.elmundo.es/

Banegas, D. L. 2011. 'Content and language integrated learning in Argentina 2008–2011,' *Latin American Journal of Content & Language Integrated Learning*, 4/2: 33-50.

Beardsmore, H. B. (ed.). 1993. *European Models of Bilingual Education*. Clevedon: Multilingual Matters.

Bernstein, B. 1973. *Class, Codes and Control Theoretical Studies towards a Sociology of Language*. New York: Routledge.

Block, D. 2014. *Social Class in Applied Linguistics*. New York: Routledge.

Blommaert, J. 2010. *The Sociolinguistics of Globalization*. Cambridge: Cambridge University Press.

Bourdieu, P. 1992. *Language and Symbolic Power*. Cambridge: Polity Press.

Brisk, M. E. 1998. *Bilingual Education: From compensatory to quality schooling*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Broca, A. 2016. 'CLIL and non-CLIL: Differences from the outset,' *ELT Journal* 70/3: 320-331.

Butler, Y. G. and **V. Le**. 2018. 'A longitudinal investigation of parental social-economic status (SES) and young students' learning of English as a foreign language,' *System* 73: 4-15.

Cenoz, J., **Genesee, F.** and **D. Gorter**. 2014. 'Critical anlysis of CLIL: Taking stock and looking forward,' *Applied Linguistics* 35/3: 243-262.

Council of Europe. 2005, April 20. *Luxembourg Presidency Position on Plurilingual Education*. Retrieved from: http://www.eumonitor.nl/9353000/1/j4nvgs5kjg27kof_j9vvik7m1c3gyxp/vi7jgt07sazv/f=/8269_05.pdf.

Council of Europe. 2006. *Content and Language Integrated Learning at School in Europe*. Brussels: Eurydice.

Council of Europe. 2011. *Teaching Reading in Europe: Contexts, Policies and Practices*. Brussels. Eurydice.

Cummings, W. K. and **O. Bain**. 2014. 'Educational quality and egalitarian educational structures: A multi-nation multi-variate analysis,' *Educational Studies in Japan: International Yearbook* 8: 33-48.

Dale, I. R. and **S. L. Robertson**. 2006. 'Homo Sapiens Europaeus or Homo Atlanticus Quaestuosus? European learning citizen or Anglo-American human capitalist? The case of the UK' in M. Kuhn & R. Sultana (Eds.), *Homo Sapiens Europaeus. The European Learning Citizen* 21–46. Frankfurt: Peter Lang.

Dallinger, S., **J. Jonkmann**, **J. Hollm**, and **C. Fiege**. 2016. 'The effect of content and language integrated learning on students' English and history competences,' *Learning and Instruction* 41: 23-31.

Dalton-Puffer, C., **Llinares, A.**, **Lorenzo, F.** and **T. Nikula**. 2014. '"You can stand under my umbrella": Immersion, CLIL and bilingual education. A response to Cenoz, Genesee & Gorter (2013),' *Applied Linguistics* 35/2: 213-218.

De Alba, V., **Vázquez, G.** and **F. Lorenzo**. 2018. **'**El desarrollo del español académico en los programas para estudiantes universitarios internacionales: las funciones discursivas superiores,' *Journal of Spanish Language Teaching* 5/2: 115-124.

De Mejia, A. M. 2002. *Power, Prestige and Bilingualism*. Bristol, UK: Multilingual Matters.

European Commission. 2012, June. *First European Survey of Language Competence* [Executive Summary]. Retrieved from: http://ec.europa.eu/assets/eac/languages/library/studies/executive-summary-eslc_en.pdf

Hattie, J. A. C. 2009. *Visible Learning: A Synthesis of over 800 Meta-Analysis Relating to Achievement*. London: Routledge.

Housen, A. 2002. 'Second language achievement in the European School system of multilingual education' in D. So & G. Jones (eds.), *Education and Society in Plurilingual Contexts* 6-127. Brussels: VUBpress.

Huettner, 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, 587–597,' *System* 44: 160-167.

Junta de Andalucía. 2011. 'Order of 28 June, on the regulation of bilingual education in schools in the region of Andalusia,' *Boletín Oficial de la Junta de Andalucía* 135: 6-19.

Loos, E. 2000. 'Language choice, linguistic capital and symbolic domination in the European Union,' *Language Problems & Language Planning* 24/1: 37-53.

López-Rupérez, F., **Expósito-Casas, E.** and **I. García-García**. 2019. 'Equal opportunities and educational inclusion in Spain,' *RELIEVE* 25/2: 1-22.

Lorenzo, F. (ed.). 2019. *Educación Bilingüe en Andalucía. Informe de Gestión, Competencias y Organización*. AGAEVE. Consejería de Educación. Junta de Andalucía. DOI: 10.13140/RG.2.2.14996.76167.

Lorenzo, F. 2007. 'An analytical framework of language integration in L2-content based courses: the European dimension,' *Language and Education* 21/6: 502-514.

Lorenzo, F. 2017. 'Historical literacy in bilingual settings: Cognitive academic language in L2 history narratives,' *Linguistics and Education* 37: 32-41.

Lorenzo, F. and **F. Trujillo**. 2017. 'Languages of schooling in European policymaking: Present state and future outcomes,' *European Journal of Applied Linguistics*, 5/2: 177-197.

Lorenzo, F. and **O. Meyer**. 2018. 'Languages of schooling: Language competence and educational success,' *European Journal of Applied Linguistics* 6/1: 1-3.

Lorenzo, F. and **P. Moore**. 2009. 'European language policies in monolingual Southern Europe,' *European Journal of Language Policy* 1/2: 121-136.

Lorenzo, F., **Granados, A.** and **I. Ávila**. 2019. 'The development of cognitive academic language proficiency in multilingual education: Evidence of a longitudinal study on the language of history,' *Journal of English for Academic Purposes* 41: 100767.

Lorenzo, F., **Moore, P.** and **S. Casal**. 2011. 'On complexity in bilingual research: the causes, effects, and breadth of content and language integrated learning—A reply to Bruton (2011),' *Applied Linguistics* 32/4: 450-455.

Madrid, D. 2005. 'Bilingual and plurilingual education in the European and Andalusian context,' *International Journal of Learning* 12/4: 177-186.

Martín-Arroyo, J. 2017, October 8. 'Las sombras del bilingüismo,' *El País*. Retrieved from: https://elpais.com/

Mediavilla, M., **M . J. Mancebón**, **M . J. Gómez-Sancho**, and **L. Pires**. 2019. *Bilingual Education and School Choice: A Case Study of Public Secondary Schools in the Spanish Region of Madrid.* IEB Working. Retrieved from

http://diposit.ub.edu/dspace/bitstream/2445/134081/1/IEB19-01_Mediavilla%2bet.al.pdf

Navarro, M. and **Y. López**. 2019. The effects of CLIL on L1 competence development in monolingual contexts, *The Language Learning Journal* 48: 18-35

Nikula, T., **Dafouz, E.**, **Moore, P**. and **U. Smit**. 2016. *Conceptualising Integration in CLIL and Multilingual Education*. Bristol: Multilingual Matters.

Paran, A. 2013. 'Content and language integrated learning: Panacea or policy borrowing myth?,' *Applied Linguistics Review* 4/2: 317-342.

Pérez, A., **F. Lorenzo**, and **V. Pavón**. 2016. 'European bilingual models beyond 'lingua franca'. Key findings from CLIL French programs,' *Language Policy* 15: 485-504.

Pérez, F., **L. Serrano**, and **E. Uriel** (dirs.), **Hernández, L.**, **S. Mollá**, **J. Pérez**,and **A. Soler**. 2018. *Diferencias Educativas Regionales 2000-2016: Condicionantes y Resultados*. Bilbao: Fundación BBVA.

Politzer, R. 1981. Social class and bilingual education: Issues and contradictions. Bilingual Education Paper Series 5/2: 1-24.

Rascón, D. and **C. Bretones**. 2018. 'Socioeconomic Status and its Impact on Language and Content Attainment in CLIL Context,' *Porta Linguarum* 29: 115-135.

Relaño, C. 2015. 'The commodification of English in "Madrid, Comunidad Bilingüe": insights from the CLIL classroom', *Language Policy* 14: 131-152.

Rolstad, K. 2015. 'Second language instruction competence,' *International Journal of Bilingual Education and Bilingualism* 20: 497-509.

Rose, D. and **J. Martin**. 2012. *Learning to Write, Reading to Learn: Genre, Knowledge and Pedagogy of the Sydney School*. London: Equinox Publishing.

Ruiz, M. A., **M. A. Sancho**, and **M. de Esteban**. 2018. *Annotated Indices on the State of the Spanish Education System*. Madrid: Centro de Estudios Ramón Areces.

Sanjurjo, J., **A. Fernández**, and **J. M. Arias**. 2017. 'Analysing students' content-learning in science in CLIL vs. non-CLIL programmes: empirical evidence from Spain,' *International Journal of Bilingualism and Bilingual Education* 22/6: 661-674.

Sanmartín, O. E. 2013, December 18. 'La educación bilingüe en inglés ralentiza el aprendizaje de Conocimiento del Medio' *El Mundo*. Retrieved from: https://www.elmundo.es/

Spanish Ministry of Education, Culture and Sport. 2017, May 6. 'Order ECD/393/2017, of 4 May, regulating the final assessment for Compulsory Secondary Education in the academic year 2016/2017,' *Boletín Oficial del Estado* 108. 36868-36897.

Torres, A. 2018a, December 17. 'El "boom" del bilingüismo llena las aulas de docentes que no dominan el inglés,' *El País*. Retrieved from: https://elpais.com/

Torres, A. 2018b, October 20. 'El problema de los colegios bilingües es que generan falsas expectativas,' *El País*. Retrieved from: https://elpais.com/

Tsuchiya, K. and **M. D. Pérez**. 2015. 'Comparing the language policies and the students' perceptions of CLIL in tertiary education in Spain and Japan,' *Latin American Content and Language Learning Journal* 8/1: 25-35.

Turner, M. 2012. 'CLIL in Australia: The importance of context,' *International Journal of Bilingual Education and Bilingualism* 16/4: 395-410.

Van Mensel, L., **P. Hiligsmann**, **L. Mettewie**, and **B. Galand**. 2019. 'CLIL, an elitist language learning approach? A background analysis of English and Dutch CLIL pupils in French-speaking Belgium,' *Language, Culture and Curriculum*.

Figure captions

Figure 1. Non-bilingual average total scores. SES levels

Figure 2. Bilingual average total scores. SES levels

Figure 3. Non-bilingual vs. bilingual L2 skills scores. SES levels

Figure 4. Non-bilingual vs. bilingual L1 skills scores. SES levels

Table 1. Sample size. No. of bilingual and non-bilingual students who sat the tests

	English L2 test	Spanish L1 test	History test
Bilingual students	1,018	1,033	1,043
Non-bilingual students	2,780	2,805	2,808

Table 2. Bilingual vs. non-bilingual L2 scores. Mean and SD

	Non-bilingual L2 scores		Bilingual L2 scores	
	Mean	SD	Mean	SD
SES 1	14.11	6.94	18.82	7.02
SES 2	16.15	6.66	19.40	6.39
SES 3	17.32	6.92	22.12	4.65
SES 4	21.52	5.75	21.07	6.10

Table 3. Bilingual vs. non-bilingual L1 scores. Mean and SD

	Non-bilingual L1 scores		Bilingual L1 scores	
	Mean	SD	Mean	SD
SES 1	9.66	4.22	12.29	3.68
SES 2	10.86	4.09	10.77	4.20
SES 3	11.29	4.34	12.76	3.72
SES 4	13.14	3.58	11.71	4.22

Table 4. Bilingual vs. non-bilingual history scores. Mean and SD

	Non-bilingual history scores		Bilingual history scores	
	Mean	SD	Mean	SD
SES 1	14.61	6.20	17.39	6.17
SES 2	17.16	6.52	17.91	6.24
SES 3	17.89	7.10	17.32	5.91
SES 4	21.05	6.37	17.66	6.37

Table 5. Differences between non-bilingual and bilingual groups

	Differences between non-bilingual and bilingual groups	
L2	Net scores difference	Percentage-wise difference
SES 1	4.71	33%
SES 2	3.25	20%
SES 3	4.8	28%
SES 4	-0.45	-2%
L1	Net scores difference	Percentage-wise difference
SES 1	2.63	27%
SES 2	-0.09	-1%
SES 3	1.47	13%
SES 4	-1.43	-11%
History	Net scores difference	Percentage-wise difference
SES 1	2.78	19%
SES 2	0.75	4%
SES 3	-0.57	-3%
SES 4	-3.39	-16%
Average	Net scores difference	Percentage-wise difference
SES 1	3.37	27%

SES 2	1.30	8%
SES 3	1.90	13%
SES 4	-1.76	-10%

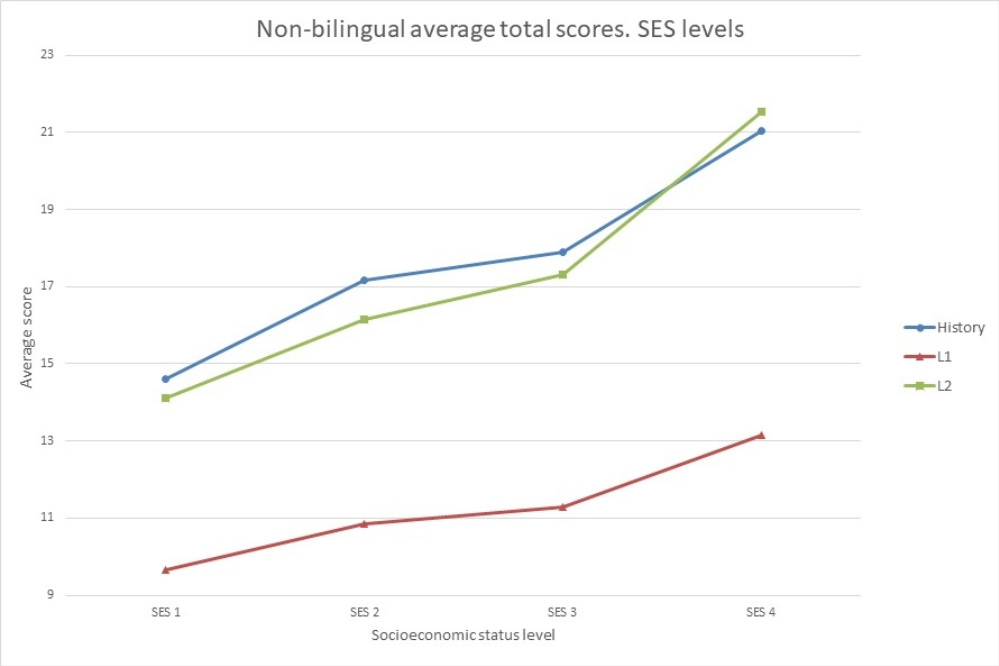

Figure 1. Non-bilingual average total scores. SES levels

235x157mm (96 x 96 DPI)

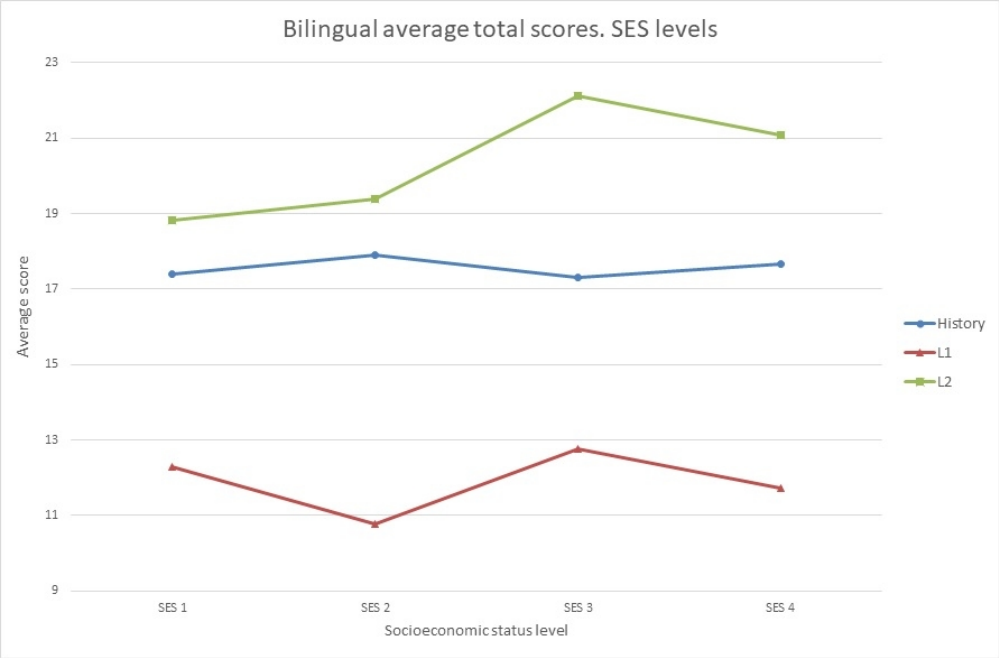

Figure 2. Bilingual average total scores. SES levels

236x155mm (96 x 96 DPI)

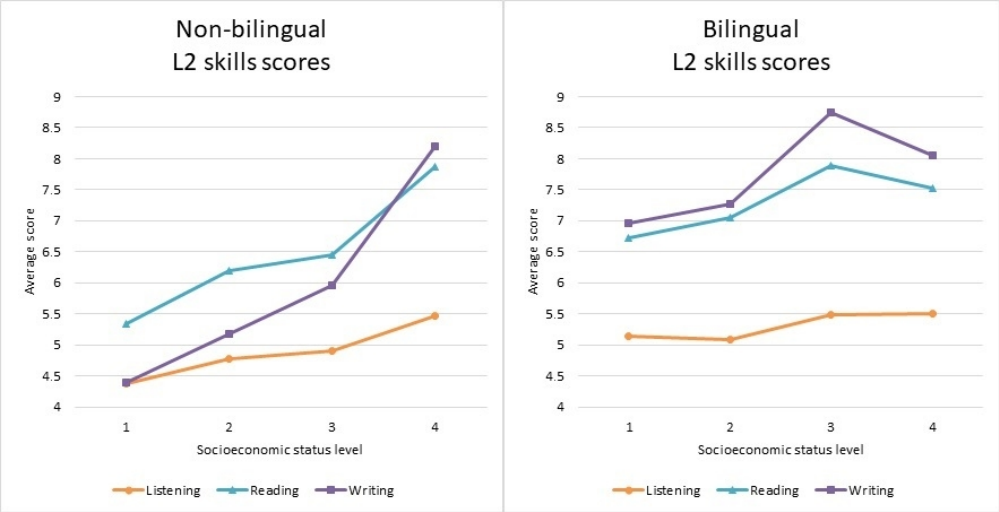

Figure 3. Non-bilingual vs. bilingual L2 skills scores. SES levels

233x120mm (96 x 96 DPI)

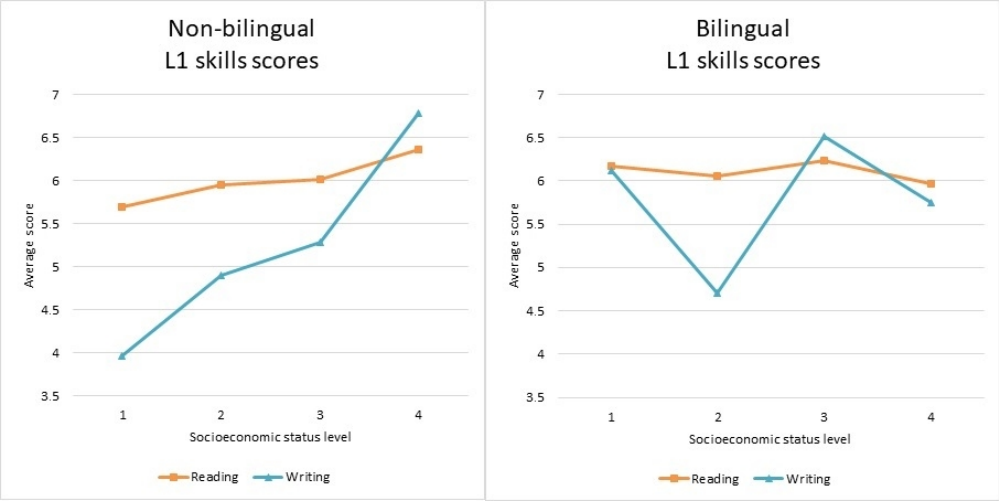

Figure 4. Non-bilingual vs. bilingual L1 skills scores. SES levels

239x120mm (96 x 96 DPI)