



**HOW FAR FROM THE PREMISE;
A META-ANALYSIS AND CRITICAL APPRAISAL OF LITERATURE
ON CONTENT AND LANGUAGE INTEGRATED LEARNING**

Fatemeh Tabari^{1*},

Elena Garayzabal Heinze²

Adriana Sampaio³

¹PhD candidate, Department of linguistics,
Faculty of philosophy and literature,
University Autónoma de Madrid,
Madrid, Spain

²PhD, Associate Professor, Department of linguistics,
Faculty of philosophy and literature,
University Autónoma de Madrid, Madrid, Spain

³MD, PhD, Assistant Professor, Neuropsychophysiology Lab,
CIPsi, School of Psychology, University of Minho,
Braga, Portugal

Abstract

Background: Although there are numerous success records of Content and Language Integrated Learning (CLIL) implementation in educational settings and its principles have been found effective, the impact of the whole educational methodology may be overrated. **Method:** A comprehensive search conducted to retrieve articles published between August 2005 and July 2016 using the scholarly databases. **Results:** We identified 102 potentially relevant articles of which only 6 met the inclusion criteria. Positive effect sizes in 5 studies indicated that the experimental group outperformed the control group. These effect sizes were analyzed separately to provide an interpretative context for the main results. High heterogeneity was observed ($Q = 5, P < .001$). The Chi-squared significance test shows that the distribution of effect sizes has heterogeneity. Likewise, I-squared statistic quantifies the heterogeneity on the data. **Conclusion:** Despite the positive feedbacks reported by researchers on the efficiency of such methodology applied in primary and secondary schools, it is occasionally admitted that the results of the study are in doubt as the participants of CLIL and non-CLIL groups do not have equal exposure to the foreign language. To compare both groups in the same conditions, it is necessary for both to receive the same number of instruction hours in L2.

¹ Correspondence: email fatemeh.tabari@estudiante.uam.es / ella_tabari@yahoo.com

Keywords: content and language integrated learning, meta-analysis, critical appraisal, content, integrated learning

“Content and Language Integrated Learning (CLIL) is a rapidly developing phenomenon in global education. It raises important issues of ethics, it challenges the role of EL teachers and there is a concern that the implementation of CLIL in education systems is outpacing a measured debate about the appropriateness of using an L2 as the medium of instruction”.

Guardian Macmillan ELT event for IATEFL 2005 (Onestop Magazine)

Because of social transition and the international needs of working life, the learning environments of today are developing towards plurilinguality of a new kind (Common European Framework of Reference for Languages 2001; Hartiala 2000; Jäppinen 2002, 2005a, 2005b). French and German government’s efforts to foster reconciliation between the two countries after World War II, gave rise to bilingual education in 1960s. Accordingly, content and language integrated learning (CLIL) program was established in Germany in 1963 following the Franco-German agreement (Breidbach and Viebrock, 2012). During 1970s, this type of education was more common in the areas close to national borders or where more than two languages were spoken aimed to help students to acquire proficiency in the second language comparable to that of native peers. However, the development of the program was influenced by the emergence of successful “immersion” program in Canada which was strongly supported by educational authorities and learners’ parents (Eurydice, 2010). Although “immersion” program was not transferrable to European educational system, it could lend some principals and initiatives to innovate a new scope of language teaching. Hence, the continuing trend towards internationalization and European integration has opened up a new horizon in European education. It focused on a more integrated system by stressing on both the non-language subjects and the language in which the subjects are though under the contemporary banner of CLIL.

Despite the early belief that CLIL increased foreign language proficiency within the required curriculum with no additional time demands, the new system has been more demanding on teachers and educators than before. They have to devote more time on adopting teaching strategies and curriculum development as well as improving their language skills that could be tailored to meet the specific needs of CLIL instruction. Comparing with conventional language teaching, CLIL calls for a significantly greater extent of human resources (specialist teachers) and suitable teaching materials (Mehisto and Marsh, 2011). Moreover, without competent CLIL teachers, successful CLIL learning seems impossible (Jäppinen, 2006). CLIL’s dual learning goal calls for subject teachers’ knowledge about the theories behind, language and content learning and planning exercises, that support both learning targets and a curriculum development based on two sets of theories (Kukkonen, 2006). Learning is also more demanding for students in CLIL environments in the beginning (Jäppinen, 2005). Although the

delivery of the subject content using a second language may seem as killing two birds with one stone, but for some educationists, it is not promising as it may hamper students' knowledge on the subjects.

1.1 Definition of CLIL and Its Goals

It is believed that CLIL is not historically unique as it is originated from bilingual education including immersion program in Canada and content-based instruction (CBI) in the United States. At the first glance, it is obvious that CLIL and other form of bilingualism and immersion programs share some common features; however, they display more differences than similarities. According to Coyle, Hood and Marsh (2010: 6), *“what separates CLIL from some established approaches such as content-based language learning, or forms of bilingual education, is the planned pedagogic integration of contextualized content, cognition, communication, and culture into teaching and learning practice”*. The differences between the mentioned approaches cannot be discussed fully within the scope of this paper.

Although the relationship between CLIL and bilingualism is complicated, CLIL is currently a European label for bilingual education (Lorenzo, 2007). CLIL has also taken variety of names such as bilingual integration of language and curricular subjects, teaching content in a foreign language, content-based language teaching, theme-based language teaching, or content-enhanced teaching and many more names. CLIL can cover a wide range of educational practice (Marsh, 2008, p.236) and be used as an umbrella term covering a dozen or more educational approaches (Mehisto, Frigols, and Marsh, 2008). CLIL may encompass a wide range of potential models ranging from single or dual, semi or complete immersion, translanguaging, modular thematic blocks and language showers (Lorenzo, F., Casal, S. and Moore, P., 2010).

Since the content taught in CLIL depends on the context in which the subject is delivered, content in a CLIL setting could also be thematic, cross-curricular, and interdisciplinary or even focus on citizenship. According to Eurydice's research in 30 European countries, the approach has been taken different names in different contexts (Lasagabaster and Sierra, 2010) depends on whether the focus is language-oriented or content-oriented (see e.g. Met 1998; Brinton, Snow and Wesche 2003; Ellis 2003).

CLIL has also been known with different terms in different countries *“Bilingualer Sachfachunterricht”* in Germany, *“Fremdsprache als Arbeitssprache”*, (FsAA– Foreign Language as a working language) or even Teaching and Learning of Science and Mathematics in English (PPSMI), English as a medium of instruction (EMI) in Hong Kong or Integrating Content and Language in Higher Education (ICLHE) in Malaysia. Thus, CLIL is a broad concept and, at the same time, confusing (Coyle 2008 and Marsh 2008).

As long as the scope and core features of CLIL are not clearly identified, it would be difficult to develop it as a pedagogical coherent method (Cenoz et al., 2013). There are still controversies between the advocates and experts in describing the goal of the method in different settings. While Marsh (2002) states that the goal of the CLIL is a

much less advanced level of second language (L2) in comparison with immersion program which goal is native-like proficiency in L2, some other like Varkuti (2010: 68) points out that the aim of CLIL approach (for example in Hungarian schools) is an ideal balanced bilingualism.

According to general consensus among learning researchers, there are three principal objectives in education that should be achieved including life-long learning, depth of understanding, and knowledge creation and enabling (e.g. Bereiter 2002; von Krogh, Ichijo and Nonaka 2000; Nonaka and Teece 2001; Scardamalia 2001; Stehr 2001). The question is then whether or to what extent CLIL can fulfill this task.

1.2 CLIL: Unique or Similar?

While Coyle (2008: 97), one of the most inspiring scholars of CLIL, believes that CLIL is unique and different from bilingual or immersion education. Cenoz et al. (2013) did not find it pedagogically unique as it is defined, interpreted and understood in different ways by its advocates. Actually, what makes such method unique is its internal ambiguity in comparison with other approaches. Such ambiguity also questions the dual role of language and content and the Ting's (2010: 3) proposed theory of '50:50/Content: Language CLIL-equilibrium'. Moreover, some research conducted in CLIL classes reported many difficulties to achieve a strict balance of language and content (Dalton-Puffer, 2007; Mehisto, 2008; Pérez-Vidal and Juan-Garau, 2010). Immersion and CLIL are not also considered comparable regarding the fact that the first is language-focused and the latter is content-focused (Marsh, 2008: 235). This fact also violates the 50:50/ Content: language CLIL-equilibrium.

Methodologically, CLIL is not considered unique while it shares some features with other approaches. It ranges from a comprehensive program of instruction to isolated lessons or activities conducted in an additional language. It is also believed to be flexible with respect to its curricular design and timetable organization. According to Baetens Beardsmore (2002), CLIL ranges from early total, early partial, late immersion type programs to modular subject-determined slots. It is called flexibility or lack of cohesion around CLIL pedagogies, "*there is neither one CLIL approach nor one theory of CLIL*" (Coyle, 2008: 101).

CLIL might be unique just because it is the only approach that includes many variants and/or a wide range of different approaches. Using it as an umbrella term makes it difficult 'to pin down the exact limits of its reality (Hondris et al. 2007; Marsh 2008; Alejo and Piquer 2010: 220). Mehisto et al. (2008), introduces CLIL as an approach which include variety of approaches including language showers, 2 CLIL camps, student exchanges, local projects, international projects, family stays, modules, work study abroad, one or more subjects, partial immersion, total immersion, two-way immersion, and double immersion.

Another point of dilemma is when some advocates like Marsh (2008: 233) state that CLIL can be applied differently based on the educational level, environment and the adopted approach. Similarly, Coyle (2007) believes that there is no set formula and

methods for CLIL and neither one model which applies to all CLIL contexts and can integrate content and language teaching. According to Cenoz et al. (2013), insistence on uniqueness of CLIL could be potentially harmful as it isolates CLIL theoreticians and researchers from mainstream research on multilingual and L2 education.

Unique or similar, the result is still a mishmash of contradictions and borrowed elements of different approaches jammed together. While the CLIL practitioners have their own interpretation and prescriptions of the method applying in the educational setting, it is impossible to assess the effectiveness of the method as a uniform methodology. It is important to square away all educational elements before implementing an educational method. Since education is one of the most important issues in the world today and should be given careful attention, it is quite important to think carefully about all aspects of the desired methodology before putting it into practice.

1.3 Early vs Late CLIL

Another issue which rises is the importance of considering the differences between early, middle or late introduction of bilingual education (in our study specifically CLIL program) in the educational curriculum. So far, empirical and theoretical research has frequently compared the benefits of early or late bilingualism. While critical period hypothesis (Johnson and Newport, 1989) strongly emphasizes on the positive outcome of learning L2 earlier in life (i.e. the native-like proficiency), the Competition and Entrenchment model of Lorenzo et al. (2005) pays special attention to the process of learning L2.

While the proficiency acquired in the early bilingualism can be neurologically justified from neuroplasticity perspective, late bilinguals' proficiency can be attained by increased exposure, cognitive skills and meta-cognitive strategies. Accordingly, early and late bilinguals cannot be easily compared as the late bilingual learner has two separate neural networks for language acquisition (Kalia et al. 2014).

Majority of psycholinguistic and neurologically-oriented research confirmed the benefits of early bilingualism (Muñoz 2006, 2008; Nikolov and Mihaljevic' Djigunovic' 2006; Van de Craen et al. 2007). However, the results of the study by Lorenzo et al. (2010) showed that middle or late start of CLIL program can result in competences similar to those of early one. Similar benefits for late (Wesche, 2002) and low frequency programs (Marsh 2002) has also been reported, may be due to increasing cognitive and meta-cognitive abilities and more advanced L1 academic proficiency (e.g. in later primary or early secondary learners) which can offset the neuropsycholinguistic benefits of early introduction (Lorenzo et al., 2010).

Almost majority of CLIL theoreticians and practitioners believe that the approach can and does work. However, it is important to know if the program (if it is supposed to be launched as a regular, effective education approach) only works for a chosen few or for students of all ages and capacities (Björklund, 2006), because language acquisition naturally goes hand in hand with cognitive development.

In another study conducted by García Mayo et al. (2005) on Basque/Spanish CLIL students, older learners pass through developmental stages much faster than younger learners considering the same educational approach and the same number of class-time hours. The authors pointed out that the exposure is presented based on the learner's age range because instruction could be more metalinguistically oriented in older learners. They also mentioned cognitive development as a decisive factor in this regard (García Mayo et al., 2005). The older Catalan students of English in the study by Muñoz (2006) also demonstrated higher accuracy rates than younger ones with the same number of hours of exposure. It is also worth mentioning that CLIL program applies to Catalan and Basque students' L3 (third language) not L2 as they are already bilingual. Apart from age and exposure, cognitive development and instruction, the variable of the type of classroom input is another involved factor (Rothman and Guijarro-Fuentes, 2010).

Jäppinen (2005) compared thinking and content-learning processes in Finnish CLIL and non-CLIL students aged 7–15 who were taught mathematics and science through English, French, or Swedish. Despite the favorable conditions for thinking and content-learning processes observed in CLIL environments, no statistically significant differences were found between the two groups in this regard. On the other hand, younger CLIL learners, aged 7–9, reported facing difficulties especially in abstract scientific concepts. Accordingly, learning in CLIL settings in the beginning seems more demanding than in environments where L1 is the medium of learning the contents.

Another question which raises here is the importance of using first language (L1) in the acquisition of L2. In late bilingualism, not only the learners' knowledge of L1 (including intricate lexical, syntactic and semantic) make L2 more comprehensible, but the literacy developed in L1 can also be transferred into L2. While some educationists believe that learning challenging content can occur while the learners are learning the target language, Krashan (1982) emphasizes that comprehensible input is a necessary condition for successful language learning which results in the increase of underlying linguistic competence. In early CLIL, subjects like science or math are directly taught in L2 without the help of L1. Whether these contents are comprehensible or the learners are cognitively mature to receive the input in L2, is an interesting topic that should be discussed in details in another paper.

Actually, proficiency in L1 can directly affect L2 proficiency development and cognitive academic growth especially for highly demanding tasks in which the learner needs to use all knowledge of the L1 and available L1 linguistic resources (Pérez-Vidal 2002). Furthermore, the cognitive load resulted from the difficulty of the subject presented in L2 can significantly decrease the learners' motivation.

1.4 CLIL for all?

According to its proposers, CLIL can increase linguistic proficiency, content knowledge, cognitive skills and creativity in learners of all capabilities and not just brilliant students (Marsh 2002, Baetens, Beardsmore 2008, Coyle et al. 2010). Besides, it is believed that CLIL learners normally outperform their non-CLIL peers without any disadvantage in

their levels of achievements in their first language or the content subjects (Serra, 2007, Dalton-Puffer 2007, Lasagabaster 2008, Alonso et al. 2008, Hood 2006, Swain and Lapkin 2005, Holmes et al. 2009).

The concept of “CLIL for all” (Coyle et al., 2010) can be a potential source of debate (Breibach and Viebrock, 2012). Despite the insistence of its advocates, CLIL, as a selective scenario, is not necessarily that beneficial (Bruton, 2011). Especially in Germany, CLIL has traditionally been applied in high-ranked schools and also offered to those perceived as the more able learners from wealthy or middle class non-migrant family. However, some German schools or vocational schools, nowadays, offer CLIL program to the children of middle and low family class to some extents but not in the form of mainstream CLIL education (comprehensive vs partial CLIL). In Germany, this program is still considered as a form of foreign language learning and bilingual education is tied with elitism (de Mejía, 2002). The fact that CLIL is often considered elitist, challenges the rationale behind the comparability of outcomes between mainstream and CLIL-stream learners.

Apart from exclusiveness aspect of this type of education, it seems quite important to know if the statement of “CLIL for All” is also applicable in all settings and age ranges. It seems necessary to neatly define its target audience. Does the word “All” covers populations like immigrants who could not be enrolled because they are in lower class of society? Is it possible to include the students with special needs or even those who cannot pass special tests to enter such program or they are deprived from participation in CLIL program due to their lower to moderate intelligence? Accordingly, it is natural if we dispute over the authenticity of the claims made by those who support the benefits of CLIL in the achievement of good students and those with less language learning aptitude (Wiesemes, 2009). Anthony Bruton (2015) criticizes “*a wholesale adoption of CLIL*” and argues that such educational method acts as a discriminator against socio-economically and educationally disadvantaged learner groups.

The results of a research in the Basque country in Spain reported that admission into CLIL classes is voluntary, but selective; therefore, CLIL learners have higher average ability and motivation in both FL and the other subjects (Ruiz de Zarobe and Lasagabaster, 2010). Apart from motivation, which is an important factor for students, socio-economic and educational status of parents plays an important role in the selection of a CLIL program at school. According to a consensus reported by CLIL teachers in Andalusian educational system, CLIL option has been chosen mostly by the students with higher socio-economic-status especially in privileged families (Lorenzo, 2008). In another study in Spain by Alonso et al. (2008), 65% of the CLIL students’ parents had higher education. It is not surprising that bilingual education is seen as a criterion for social selection and educational qualifications in the world today (Bernhardt, 2015).

According to some studies, such students are also very likely to take extra English classes outside school (Ruiz de Zarobe and Lasagabaster, 2010; Villarreal

Olaizola and García Mayo, 2009). Logistically speaking, majority of educationists believe that CLIL is open to all students; but implicitly its target group are consisted of socially advantageous students who opt for bilingual education (Alonso et al., 2008; or Ruiz de Zarobe and Lasagabaster, 2010). This claim is frequently confirmed by the teachers working in bilingual schools (Bernhardt, 2015). Since 2004-2005 in Spain an educational evolution has taken place in which both traditional bilingual regions, such as Catalonia and the Basque country and Castilian speaking areas, such as Andalucía and the Madrid area opted for CLIL. While in Spain, state schools are not allowed to stream students, optional CLIL stream is mostly selected by higher socio-economic-status parents who opted to put their children into such program (Villarreal Olaizola and García Mayo 2009; Lorenzo 2008).

The prevailing idea is that CLIL students are capable of achieving greater language proficiency as well as more extensive knowledge of the content presented, using concepts more accurately, and better perceiving their own perspectives. However, all these competences and capabilities could be developed before attending bilingual program under the effects of their parents (Stefanie Lamsfuß-Schenk, 2008; Bernhardt, 2015). Marsh et al. (2000) reported that many parents in Hong Kong were more interested in English-medium programs because they think that their children will have more opportunities in the future if they can speak English.

The emerging trend among the educationist and researchers in Europe in support of benefits of CLIL in improving foreign language education at all levels undermines the performance and leaning of non-CLIL students (Bruton, 2011). Despite the arguments of Marsh Lorenzo et al. (2010) that CLIL is egalitarian; all attentions are shifted from education for all towards an exclusive approach of all education for one group. Actually many potential pitfalls of CLIL are avoided by the educational systems by offering the program solely to academically motivated students (Bruton, 2011).

The idea of using CLIL for all is like the controversy between the ideas of one-size-fits-all models against one man's meat is other's poison. Furthermore, evidence suggests that some students are apparently prejudiced by CLIL especially in the state educational sector and institutional interests are taking precedence over students' interests (Bruton, 2011). However, in contrast to the notion of "*CLIL serves all students*", some research showed that CLIL programs are not available for all students (Mehisto, 2007; Lasagabaster and Sierra 2010; Bruton 2011). Such educational discrimination and class prejudice reached to the level where some researchers like Mehisto (2007: 63) and Bruton (2011: 524) points out that "*CLIL can attract a disproportionately large number of academically bright students*" or "*many of the potential pitfalls which CLIL might encounter are actually avoided by selecting for these programs students who will be academically motivated to succeed in the foreign language (FL), as in other subjects*". Unsurprisingly, among the participants of the program, the least disadvantaged have always been those who had higher levels of L2 to begin with (Zydati, 2012). Even the academically bright with low initial English are the most prejudiced comparing with those academically bright students with higher proficiency in FL (Bruton, 2011).

Considering immigrant students in Spain, Lasagabaster and Sierra (2010: 372) stated that CLIL may be even more elitist than immersion programs insofar as immigrant students are usually enrolled in immersion programs, whereas they seldom take part in CLIL programs. Another source of prejudice could be the teachers' motivation to teach CLIL groups because they are motivated and easier to manage.

Besides, a project which has been successful in one setting for a particular group of people could not be prescribed for different people in a different setting. Some cases have also been reported about failure of the CLIL implementation due to educational policy issues and lack of facilities. In a three-year research project in South Africa, on looking at issues arising from children changing from an African language to English as their medium of instruction in Year 5 of primary school, showed that such difficulties did not arise not only as a result of ineffective language teaching methods but also due to the inappropriate language policies which traumatizes both teachers and pupils. Inadequate time and materials were applied to prepare children for a change in the medium of instruction, although the curriculum was not carefully designed to lead children towards an understanding of abstract concepts (Macdonald and Moodie, 2006). By observing the result of CLIL program for over three years of implementation in a secondary school in Hong Kong, Marsh et al. (2000) concluded that the students were very disadvantaged by learning academic subjects like geography, history and science (to a lesser extent, mathematics) through the medium of a foreign language.

Accordingly, it seems crucial to investigate the real effect of CLIL on the development of the content learning and the methodology of the content classrooms, for the average student, those with less ability and those who are not selected to attend the program rather than selected and highly motivated ones with additional language exposure (Bruton, 2011). If the ultimate goal of CLIL program is preparing learners for future academic and working life, all groups of learners from different socio-economic status and different mental capacities should be taken into consideration. Whether successful or not, such discriminative attitudes which exist in the nature CLIL program can jeopardize the notion of comprehensive education.

1.5 Are CLIL student and non-CLIL students comparable?

Comparison of CLIL students with other groups with no CLIL streams seems invalid as CLIL groups and control groups do not share similar characteristics. CLIL students who are more academically gifted may have higher content subject scores or language proficiency scores due to the mentioned factors (Vázquez et al., 2013). Therefore, such comparison does not make sense without considering any pretest scores. It is naturally expected that CLIL students outperform their non-CLIL peers on post-test scores in the same or different schools due to specified factors. Such comparison is like track and field in which contestants start running from different lane with different starting points. Apparently, those whom you expect to arrive first are mostly highly motivated, well-trained contestants who occasionally began running from a point closer to the end point.

However, there are some counterexamples in the literature like the one mentioned by Nave's and Victori (2010) in Catalonia where CLIL groups started with a proficiency advantage which they maintained subsequently and did not increase. Many other cases have also reported that non-CLIL students outperformed their CLIL peers in some academic subjects such as maths (Seikkula-Leino, 2007). Gené-Gil et al. (2015) observed significant progress of non-CLIL students in lexical complexity compared with CLIL learners. Fontecha, et al. (2014) similarly found that non-CLIL learners are significantly more motivated than CLIL learners. Even, there are examples of CLIL students who started off with much higher motivational levels but after two years of participating in the program their average motivational levels generally decreased but in non-CLIL group such level increased and they closed the gap (Lasagabaster and Sierra, 2009).

More exposure to L2 by receiving more hours of instruction comparing with non-CLIL students is the most obvious factor reflecting unequal conditions of both groups which void the validity of the majority of the comparisons made. In spite of all the positive findings reported, the possibility of comparison of a CLIL student who receives more L2 education hours with the one who only study L2 in FL sessions seriously questions the efficacy of CLIL methodology. Besides, the occasional presence of language assistants as an authentic source of L2 is another undeniable parameter in this regard.

Bruton (2011a, 2011b) published a criticism on the report of the study conducted by Lorenzo et al. (2010) in the region of Andalusia which emphasizes on the significant gains in favor of CLIL and foreign language score differences between CLIL and non-CLIL groups on language competence. He argues that the mentioned study, like the other studies done on CLIL, has serious limitations like lack of pre-post average scores, valid comparison groups, comparable contexts, control over extra FL instruction outside school, and the support of additional teachers and coordination time (Bruton, 2011a, 2011b). He also addresses several confounding factors (e.g. socioeconomic status, extra language courses, motivation and higher proficiency in L2) in the context of CLIL research which definitely accounts for making critical difference in favor of CLIL students.

Furthermore, there is a criticism over the claim that significant differences exist in favor of CLIL students with respect to lexical transfer and outperforming non-CLIL peers on lexical availability task and a cloze tests (Agustín Llach, 2009; Celaya, 2008; Celaya and Ruiz de Zarobe, 2010, Jiménez and Ojeda, 2008a). This claim can be denied by the opponents arguing that such students obtain wider range of vocabulary in L2 due to higher number of hours of exposure in their educational curriculum (Celaya, 2006; Manzano, 2014).

In a study conducted by San Isidro (2010) empirical evidence exists on the success of CLIL approach to improve students' competence in FL. But the author admitted that possibly the degree of motivation and FL competence of CLIL participants could be possibly higher than non-CLIL learners and such difference

should be taken into consideration cautiously while the results are reported. Apart from motivation, there are several individual (age and sex) and contextual (socio-cultural) factors that should be taken into account in analyzing the results of such comparisons (Doiz et al., 2014). A similar study in Austria reported the benefits of CLIL on accuracy of verb forms; however, the authors also admitted that CLIL learners could opt for such program due to the factor of motivation (Hüttner and Rieder-Bünemann, 2007). Regarding the factor of age, CLIL learners at secondary schools have been reported more motivated than those at primary education because of their awareness towards the importance of leaning a FL (Lasagabaster, 2011; Murtagh, 2007; Seikkula- Leino, 2007).

Apart from small sample size and more exposure to the target language (Ruiz de Zarobe; 2007, Villarreal Olaizola and García Mayo; 2009), the lack of pretest scores has been the most important criticisms that have been raised against the optimistic attitudes towards the absolute efficacy of CLIL approach. According to Bruton (2011a, 2011b), establishing a benchmark is a fundamental step in conducting research into any form of development. Without any pretest scores, there is no point of departure to assess any form of changes and any improvement occurred and to make sure that the two groups begin with initial average score differences (Bruton, 2011a, 2011b).

Considering the factors like self-selection, higher motivated, more hours of exposure to the FL, taking extra English courses outside school, we cannot make sure whether the recruited groups are actually comparable. On the other hand, the comparison between CLIL and non-CLIL groups cannot be valid unless they are similar in many other respects including students' and teachers' proficiency levels and the type of methodology used (Alonso et al., 2008). The methods of data collection and interpretation are other important factors that should be taken into account while comparing both groups. The data, required for comparison, should be collected using both qualitative and quantitative methods. The data collected by only questionnaire on classroom activity, are unreliable and limited in scope and should be supported by at least some observational evidence (Lorenzo et al., 2010; Bruton, 2011).

1.6 CLIL applied to all topics?

By reviewing the literature, we found out that only selected topics have been taught in L2 in special places over a limited period of time (Krechel 1999). For example, mathematics has been frequently excluded from CLIL practices in Austria (Nadja Wilhelmer, 2010). Breidbach and Viebrock's (2012) experience in implementing CLIL in Germany suggest that structural selectivity of CLIL appears to have a greater impact on student achievement than CLIL itself has on student achievement.

Special subjects such as art, music, drama and physical education, biology, geography and history, and to a lesser extent social and political studies, have been covered by CLIL program. However, a subject like math is rarely offered through an additional language, especially in Germany, due to some factors: its abstract non-linguistic nature, difficulty of the subject that could be magnified if presented in L2 and

the fact that there are only few teachers who have the sufficient knowledge and competency to combine languages and math (Breidbach and Viebrock, 2012).

In Germany, sciences and math are mostly absent from CLIL programs because in secondary level, CLIL teachers are usually recruited from teachers who has competency and received training in both the foreign language and CLIL subject. Accordingly, some studies cast doubt on the suitability of combining individual subjects and languages in CLIL program (Mentz 2010; Rymarczik 2003, Witzigmann 2011). In another study conducted in Hong Kong, high school students were very disadvantaged consistently across the first three years of high school by studying subject matters like geography, history, science, and, to a lesser extent, mathematics through English as a FL (Marsh et al., 2000: 337).

According to the findings of Seikkula-Leino (2007) in a comprehensive school in Finland, where CLIL groups were all selected to attend a program in which 40–70% of the CLIL math classes were conducted in English, the participants reported low self-esteem on FL ability. On their math test, the distribution of students in terms of under-average-high-achievers according to their IQ potential, there were far fewer over-achievers in the CLIL group than in the normal group (30%–10%) indicating that, despite the pre-selection, the CLIL group might have been disadvantaged by studying math in English.

In case of teaching math in CLIL, the most important prerequisite is the ability to transfer the content to pupils using a comprehensible input. Special attention should be paid to the use of the language along with a more rigorous and synthetic use of mathematical language and language accuracy for teaching (Breidbach and Viebrock, 2012). In math, vocabulary may be confusing because different words convey different meaning based on its usage in mathematics and non-mathematics contexts (e.g. two different words sound the same such as whole in whole number which can be mistakenly understood as hole or because more than one word is used to describe the same concept). Similarity of symbols representing mathematical objects and different representations for describing the same process all can be confusing. Graphic representations can be another source of misperception because of formatting variations or because the graphics are not consistently read in the same direction (Kenney et al., 2005).

Moreover, Kenney et al. (2005), points out that the greatest difficulty in learning the language of mathematics is a double decoding process (i.e. decoding spoken mathematics words in the initial context of normal parlance, and then translating it to the different context of mathematics usage). While double decoding seems such a difficult process for most students, it could be quite complex for L2 learners.

Furthermore, problem-solving and reasoning which are fundamental skills required in mathematics, include four predominant actions including modeling and formulating, transforming and manipulating, inferring and communicating (Schwartz and Kenney, 1995). For more complex problems some extra proficiency like creating a mathematical model and generalizing and extending the results of a mathematical

action seems necessary. While too much mental efforts have to be made to learn mathematical abstract concepts and solve the problems, changing the language of instruction not only makes a rod on learners' back but also reduces their efficiency by involving some psychological factors such as lack of motivation and stress.

1.7 CLIL for all teachers?

Three groups of teachers are involved in CLIL program with different areas of expertise, teaching styles and even different types of activities and materials. Language teachers and content teachers provide semi-immersion system while the first ones work on sentence-level grammar and the latter focus at the textual level. Language assistants, oppositely, represent a full-immersion system by fostering conversational style language. Accordingly, native-speaker assistants use L2 most frequently; while, the results of several studies suggest that state FL teachers use significant amount of L1 (e.g. Carless 2004; Dalton-Puffer 2007; Orafi and Borg 2009). Actually, language assistants are considered as a compensation for limited L2 proficiency and random counterproductive code-switching of content teachers (Lorenzo, 2008).

According to its stakeholders, CLIL teaching contributes to raising students' motivation and confidence in the target language, decreasing their anxiety (Pérez Cañado, 2012), enabling students to develop better speaking skills (Dalton-Puffer, 2008), improving receptive and productive vocabulary (Dalton-Puffer, 2011) and decreasing L1 transfer (Agustín Llach, 2009). To achieve such goals new roles and responsibilities are demanded of both language and subject teachers (Simões et al., 2013). Non-native subject teachers are required to develop their own language skills and overcome linguistic insecurity (Dale and Tanner, 2012). Furthermore, for being involved in CLIL program both language or subject teachers are required to hold special training certificates for example in Europe, it is necessary to obtain the certificate of Common European Framework of Reference for Languages and a certificate of advanced foreign language proficiency (Gobierno de La Rioja, 2004, 2005, 2009).

Apart from proficiency and knowledge, the most problematic issues are CLIL teachers' lack of sufficient strategies to present academic content through another language with an understandable and sound output. They do not either realize that the presented subject is also a place for language development and practice as much as content acquisition (Lorenzo, 2008, Mehisto, 2008, Lyster 2007, Gajo 2007).

Moreover, teaching language and content are two separate entities and should be taught using two separate systems and different strategies. While content teachers lack enough qualities in teaching subjects through languages, language teachers are not competent in teaching contents (especially math and science) and language assistants are qualified neither in language nor content pedagogy, the result of such educational program seems not to be promising and the theory of balanced pedagogic integration of content and language in CLIL.

The last chance for solving the issue can be the collaboration of both language and content teachers by interchanging their knowledge and ideas in designing

curriculum and materials as well as teaching techniques. However, the fact indicating that the number of language teachers who are involved in this program are less than content teachers and majority of CLIL teachers are subject specialists without formal qualifications in foreign language and/or general language pedagogy (Dalton-Puffer et al., 2009; Evnitskaya and Morton , 2011) imbalances the equilibrium of collaboration between content and language teachers.

1.8 Addressing the issues

There are a lot of “ifs” and “buts” involved in CLIL program that cannot be ignored. Like any other educational models, it suffers from limitations that are rarely recognized and shortcomings that are not fully addressed (Banegas, 2011). Assimilation of CLIL to a “linguistic bath” which can equip the learners with all the knowledge and skills required for real life communication (Dalton-Puffer 2007, Lasagabaster and Sierra 2009) is considered as an overestimation of linguistic potential of the approach that failure to fulfill its expected outcomes and can result in an early and unfair disappointment with the results (Harrop, 2012).

Reviewing the literature shows that the success stories about CLIL approach are not often supported with substantial empirical evidence and the authors have frequently concluded that the outperformance of CLIL groups over mainstream peers has been in FL, (but not content). Therefore, more caution should be exercised while reporting the strengths and shortcomings of the approach and its effectiveness in diverse contexts as well as interpreting its definition (Cenoz et al., 2013). Despite the positive results reported in favor of CLIL regarding their higher levels of proficiency and higher communicative competence than their non-CLIL peers, the differences have not always been substantial (Dalton-Puffer 2008, Ruiz de Zarobe et al. (ed.) 2009, Alonso et al. 2008, Admiraal 2006, Airey 2009).

Even though the development of CLIL has stimulated research on content and language integrated learning, there are important empirical gaps in our understanding of its effectiveness. Bruton (2011), for example, points out that although the rationale for integrating content with language teaching includes the assumption that this will increase motivation and, thus arguably, use of the target language, it could have the opposite effect. More specifically, students’ motivation might be reduced because of loss of self-esteem when they are required to use a language that they do not know. On the other hand, the use of language might actually diminish if the subject matter is novel and/or complex resulting in reduced language acquisition. Similar concerns have been identified by immersion researchers (Lyster 2007) and by Lin and Man (2009).

CLIL instruction usually entails more contact hours with the target language during the school day. Hence, this extended exposure to the target language is considered as a crucial variable (Tedick and Cammarata 2012). Perhaps the same number of hours of direct language instruction would be as effective or more effective without a CLIL approach (Bruton, 2011).

The need for more research in general has been noted by some CLIL experts (Dalton-Puffer et al. 2010a). Diversity of CLIL program formats and the lack of a standardized CLIL blueprint pose several challenges to carrying out research on it (Van de Craen et al. 2007). The first prerequisite of carrying out a research on CLIL is to provide a clear description of the implemented program so that others can understand the limits of generalizability of the results. As long as there is no specified and refined definition of CLIL that can cover all diversity of formats, the results of the studies cannot be interpreted with confidence and positive findings cannot necessarily be assigned to the method applied.

Rather than insistence on the uniqueness of this approach, scholars have to provide a critical empirical examination of strengths and weaknesses of CLIL in its diverse forms and different learning contexts (Cenoz et al., 2013). Although literature has recorded many success stories of CLIL and its influence on practice; some studies have reported its negative outcomes and some others reported no significant difference between CLIL students and their non-CLIL peers regarding academic achievement or other linguistic skills after running the program for a considerable period of time. Such findings are in consistence with the result of a test on a traditional group and CLIL group by Ruiz de Zarobe (2007) which showed no overall significant differences between CLIL groups and their peers in the traditional group with respect to oral proficiency, despite the fact that the CLIL group had more exposure to L2 (105 hours per year more English) than the other group prior to the test. However, finding no difference between the groups is acceptable but implementing a program which causes lack of achievement, cognitive skills and self-esteem among its participant needs an urgent consideration of educationists.

Some of the most significant weaknesses of CLIL which are summarized in a mini-analysis by Bruton (2011) are the variety of interpretation of research results and bias due to researchers' interest as well as limitations including pretesting, sampling, data and less proficiency and motivation levels of mainstream groups compared with their CLIL peers. Another reason that can discard the possibility of comparison between CLIL and non-CLIL students is that the participants of the control groups are the remnants from the (selected) CLIL groups in the same schools. No study has ever reported the comparison between CLIL groups with non-CLIL groups of the same initial proficiency in other non-CLIL schools (Casal and Moore; 2009).

According to Harrop (2012), there is lack of focus on form in CLIL classes due to its content-led nature which can result on an early fossilization of errors. Additionally, the majority of error corrections are lexical and little attention is paid to grammatical errors in CLIL practice due to lack of systematic and constructive method of error correction which focuses on form (Harrop, 2012).

Drop outs are other elements of CLIL that are conveniently ignored (Apsel, 2012). Initial reports mention a 10% drop-out rate due to inability to cope with the demands of the program (Hidalgo, 2010). Netten and Germain (2009) also reported cases of 20% drop-out rates in immersion contexts. The problem seems to be getting

worse due to the lack of linguistic proficiency, which causes insufficient level of understanding and learning (Lightbown and Spada 2006), insufficient teacher proficiency or a limited range of teaching strategies to support linguistic development (Harrop, 2012).

Due to the intrinsically challenging nature of CLIL, the educational system is responsible to ensure adequate standards in the L1-medium education for all students before establishing such program as exposing the learner to plenty of language above their current level of competence. Besides, there is less risk in presenting the program to elder learners as preserving the learners' self-esteem in the initial stages of CLIL is necessary while they are adjusting to the new challenge (Hood, 2006).

So far, educationists and CLIL proposers have not been good at working out the cost of implementing CLIL that was supposed to be a genuinely promising model, and it may be due to unrealistic expectation and ignorance of shortcomings. Therefore, it has been changed into a costly model in terms of financial and human resources. Vulnerability of weak learners to impaired linguistic development and lack of cultural awareness in CLIL models are among other defects that need to be reconsidered (Harrop, 2012).

2. Research Question

Although there are numerous success records of CLIL implementation and its principles in educational settings, the impact of the whole educational methodology may be overrated. There is no question that CLIL learners outperformed non-CLIL learners in some aspects but the results need to be interpreted with caution. Personal experiences and intuition have no place in adopting or rejecting an assumption. It is not logical to state that an educational methodology is effective unless the statistics shows the exact effect size and scientifically justify the obtained results. In this critical review and meta-analysis, we revisit the question of whether CLIL might have enhancing effects on learner's competence, achievement and motivation. The results of the present study can help the education policymakers to make sound decision in their future policy formulation.

3. Methodology

This study was planned and conducted in adherence to PRISMA standards of quality for reporting meta-analysis. PRISMA 2009 checklist was followed in reporting each section, such as introduction, methods, results, and discussion.

3.1 Study selection

We conducted a literature search on the studies published between August 2005 and July 2016. A comprehensive search was executed articles completed up to July 2015 using the scholarly databases such as Google Scholar, Research Gate, Educational

Resources Information Center (ERIC), Web of Science, CrossRef and Scopus. The literature search was limited to articles published in English. We extracted only references that included the words “CLIL” or “Content and Language Integrated Learning/ Teaching” in the title or abstract.

3.2 Criteria for considering studies for this review

The researchers reviewed the full texts of the eligible articles for possible inclusion based on the inclusion criteria. The authors applied the following inclusion criteria to the full texts: 1) Online published in peer-reviewed journals 2) English-language publication 3) Year of publication sufficiently recent (between August 2005 and July 2016) 4) Available data comparing groups on main characteristics 5) Experimental or quasi-experimental studies focused on CLIL 6) Quantitative studies with sufficient statistical data to calculate effect size (e.g. the number of participants, mean and standard deviation). Studies with insufficient data were excluded. 7) The participants were reported to be CLIL students (studying at least one subject matter in L2 apart from EFL classes) or mainstream students (studying L2 only in EFL classes). 8) Assignment of the groups to experimental (CLIL) and control groups (non-CLIL) 9) Studies performed on primary and secondary school students 10) Measured outcomes were clearly reported. 11) They were publicly available online.

Of the potentially relevant 137 articles, screening of the title and abstracts resulted in 102 relevant studies (35 were excluded as they were published in other languages than English). A total of 6 articles were retained for the second screening based on inclusion criteria after excluding those with inadequate statistical data (Figure 1 – Flow diagram).

3.3 Data Coding

All studies were coded and organized into 9 major categories in our selected database spreadsheet tool (Excel). These include (a) study identification, (b) measured outcomes, (c) aim and research questions, (d) study design, (e) date of publication, (f) sampling strategy, (g) data collection, (h) data analysis, and (i) results and conclusion. Unfortunately, most of the studies in the present literature lack a long-term perspective and comparable data and research. The manual included information regarding effect size calculations and the characteristics of the study and the report.

3.4 Statistical Analysis

We aimed to synthesize the results of the studies using meta-analysis to integrate the results of the empirical research and make a conclusion about the effectiveness of CLIL in educational settings. The software Comprehensive Meta-Analysis version 2 (Biostat, Englewood, New Jersey) was used to conduct the data analysis. Effect size estimates were adjusted for sample size (Hedges’s *g*), and 95 % confidence intervals were calculated to assess the statistical significance of average effect sizes.

Fixed effects models assume that the primary studies have a common effect size. In contrast, random effects models attempt to estimate the distribution of the mean effect size, assuming that each primary study has a different population. Accordingly, a test for heterogeneity of the intervention effects was performed using the Q statistic (Table 1). As the results of the test for heterogeneity was statistically significant, we used the random effects models to accommodate this heterogeneity for the main effect

3.5 Computing Effect Sizes

The importance of research results is often assessed by statistical significance, usually that the p-value is less than 0.05. P-value is a statistical measure, but the effect size is the estimate which tells us about the practical significance. Effect size can be determined by calculating the value of Cohen's d and the effect-size correlation, r_{YI} , using the means and standard deviations of two groups or using the t-test value for a between subjects t-test and the degree of freedom. To correct the bias which may occur due to inflated effect size especially for small sample sizes, Cohen's d was converted to Hedges's g, an unbiased estimate of the standardized mean difference effect size (Hedge and Olkin, 1985).

Each effect size was first multiplied by the inverse of its variance to yield the weighted effect size. Then the sum of all the weighted effect sizes was divided by to derive an overall weighted mean estimate of the effect of the treatment. The weighted mean effect size was estimated. The standard error of Hedges'g unbiased estimate of the mean effect size was then computed.

4. Results

We identified 102 potentially relevant articles using the search strategy described above, of which 6 met the inclusion criteria. The characteristics of the 6 studies included in this meta-analysis are listed in Table 1. Positive effect sizes in 5 studies indicated that the experimental group outperformed the control group. Besides, in two studies that measured two different outcomes, the authors reported the final effect size as the mean of the two effect sizes because the outcomes, conceptually, can be merged and we can use the average. As we do not need to analyze the impact of CLIL per outcome, we combined the effect sizes in two of the studies.

4.1 Overall analysis

We found the effect sizes derived from 6 publications which met the inclusion criteria (Lasagabaster, 2011; Fontecha AF, and Alonso AC, 2012; Vázquez BM, 2014; Lorenzo, F., Casal, S., Moore, P., 2010; Heras and Lasagabaster, 2015). The studies were combined in the meta-analysis. These effect sizes were analyzed separately to provide an interpretative context for the main results. High heterogeneity was observed ($Q=5, P<.001$) (Table 1). A positive effect size indicates that CLIL participants fared better than the control group. The studies were found to have a statistically significant,

combined effect size of $g=0.642$ with 95 % confidence intervals of 0.98–1.209 (Table 1, 2) (Fig. 2).

4.2 Heterogeneity Analysis

Different values for heterogeneity analysis are summarized in Table 2. The Chi-squared significance test shows that the distribution of effect sizes has heterogeneity. Likewise, I-squared statistic quantifies the heterogeneity on the data.

4.3 Forest plot

A forest plot summarizes overall effect with a pooled result and shows the amount of variation among studies. Figure 2 shows the forest plot with Hedge's g , which is a corrected standardized mean difference estimate for the effect size. Individual squares represent each study's effect size estimate. The area of each square corresponds to the weight that the individual study contributed to the meta-analysis. Larger squares also indicate the studies of larger samples and larger sample size and precision mean the more weight assigned for each study. The lines extending from the squares show the 95% confidence interval for the estimate and the diamond represents the overall estimate from the meta-analysis and its confidence interval.

4.4 Effect Sizes by Individual Studies

Considering numerous success records of CLIL implementation in educational settings and efficacy of its principles in different educational settings, it seems necessary to analyze the magnitude of efficacy of our final studies with different outcome measures, instrument and the content area. Although, there are many plausible reasons which void the validity of comparison between CLIL and non-CLIL groups which mentioned before, they can also be investigated in every individual study. Unsurprisingly, we observed that the majority of studies yield a higher effect size for CLIL group that may not be related to true program effects and achievement differences. Publication bias could also be another reason which gives an unfair advantage to CLIL group.

There is a wide range of variability in programs, grades, sample sizes and outcome measures. Table 1 lists the summary of the coded studies and associated effect sizes. We see that the average effect size is highly influenced by number of the participants included in the experiment and control group. Apart from two studies of which one yields a negative effect size, the rest have unequal number of participants recruited in each group. Table 3 depicts the statistics for each study along with their sample size.

Furthermore, the magnitude of an effect size reflects the between-group difference in units of the standard deviation of the control group. As it is shown in fig. 1, the fourth study with the biggest sample size has the biggest effect size which can also cause publication bias. For example, the results showed the size of the sample for non-CLIL group to be about 59% of the size of the CLIL group. The mean effect size was

calculated as 1.260. This indicates that the experiment group scored about 60% of a standard deviation higher than the case group

A negative effect size indicates that the comparison group fared better. In the study conducted by Fontecha and Alonso (2012), non-CLIL learners are more motivated than CLIL learners. Oppositely, Heras and Lasagabaster (2015) in the sixth study reported some positive effect on particular aspects of students' affective factors (motivation and self-esteem). However, there are some points about such results that should be considered cautiously. Apart from confounding factors like small sample size and low-middle intensity of CLIL program, no significant difference were observed between CLIL and non-CLIL groups with respect to affective factors unless gender has been added as a factor. The factor of motivation has been the focus of the first, second and the sixth studies; however, motivation may not be a valid criterion for comparison between educational groups because it normally changes over time due to environmental factors and learners' characteristics.

In Fontecha and Alonso's study, the population was selected from language learners of the Basque Country which is already a bilingual community where both Basque and Spanish are official languages and are taught at school from the outset. Therefore, English is considered as the third language there. In the same vein, the factor of bilingualism could be also another confounding factor in generalizing the results of the studies done in such settings for other communities, especially those who learn English as their second language inside the Spain (like Andalusian or those from the community of Madrid) or other European countries. In case of the first, fourth and fifth studies, the recruited population are already bilingual students in Basque Country and the positive effect sizes of such studies should be considered cautiously. Learning academic content through the second or third language (bilingual vs trilingual) could be similar but not equal as the development of second and third language competence do not occur equally.

4.5 Testing for publication Bias

To determine whether the reported results can be addressed as valid, it is necessary to examine the potential impact of publication bias which is a threat to the validity of meta-analyses because statistically significant results are more likely to be published and accessible for inclusion in meta analyses. These forms of biases tend to have more effect on small studies and contribute to the phenomenon of "small study-effects" (Sterne et al., 2000). Since 5 out of the 6 analyzed studies had relatively small sample sizes, they tended to have larger and more favorable effects compared to the study with larger sample size.

A common method to detect publication bias is using a graphic plot of treatment effect against standard error (sometimes based on sample size or precision) for each of the initial studies in a meta-analysis. Accordingly, we generated a funnel plot to examine the distribution of effect sizes in relation to the studies' sample sizes. Figure 3 shows the funnel plot constructed upon random effect model by considering each study

in the sample of studies as unit of analysis. The presence of bias led to an asymmetrical appearance of the funnel plot (figure 2). In this situation the effect calculated in a meta-analysis tend to overestimate the intervention effect. The more pronounced the asymmetry, the more likely it is that the amount of bias will be substantial.

Firstly, any conclusions about publication bias should be drawn cautiously because of the small number of the studies. As is shown in figure 3, there are more studies with significant positive effects published in the statistics literature than studies with negative effects indicating the possibility of publication bias because of the strong asymmetry and heterogeneity of the scatter points around the mean of the effect sizes. Furthermore, this funnel plot shows that studies with larger samples (particularly the one on the top right corner) have larger effects. We noted that Lorenzo et al. (2010) was potential outliers for the analysis.

Secondly, asymmetric plot confirm the existence of correlation between the treatment effect estimate and the studies' size suggesting the possibility of publication bias and systematic difference between the smaller and larger studies. As it is shown in figure 3, large studies appear toward the top of the graph, and tend to cluster near the mean effect size. Smaller studies appear toward the bottom of the graph since there is more random variation in the small studies. Moreover, if there are more studies on the right than on the left, the concern is that studies may be missing from the left. The 'trim and fill' method is a non-parametric approach that makes strong assumptions about funnel plot asymmetry (Duval and Tweedie 2000a, 2000b). It imputes the missing studies, adds them to the analysis, and then re-computes the summary effect size. The study right to the mean (big effect size) causes publication bias. So the trim and fill imputes left to the mean to make the funnel plot symmetrical (table 3, fig.3). According to trim and fill only 1 study was imputed to guarantee the funnel's plot symmetry.

Likewise, if there is a publication bias, generally a skewed and asymmetrical spread is expected on the funnel plots. The asymmetry of our funnel plot also signifies that positive effect sizes are overrepresented. In this situation, the overall effect estimated in meta-analysis overestimates the treatment's effect by resulting in an effect size of 0.642, which would be expected to be 0.514, as calculated by trim and fill method, if there would be no bias (table 4).

5. Discussion

This review aimed to investigate the efficacy of CLIL approach on learner's competence, achievement and motivation. Although database search provided more than 137 results, only 6 articles met the inclusion criteria. Firstly, the results of the meta-analysis show clearly that CLIL group outperformed non-CLIL group on different outcomes including linguistic competence, achievement and affective factors. Using meta-analysis, we computed the summary effect of all 6 studies, resulting in an average (weighted) effect size estimate of Hedges'g = 0.642 (SE = 0.295; confidence interval of Hedges'g = 0.0642 to 1.219, p=0.029): a significant effect.

However, we acknowledge that a majority of the included studies had small sample sizes and we cannot make any strong causal claims about the obtained results. The Q-statistic indicates that there was a significant heterogeneity among the effect sizes ($Q = 67.373$; $df = 5$; $p = 0.000$), which means that it is unlikely that all interventions shared the same true effect size. On the other hand, narrow confidence interval reflects the existence of small variance for the mean effect size while wider prediction interval results from high heterogeneity of the data. Our findings also yields a modestly biased effect size estimate of 0.642, which was corrected as 0.514 indicating that the effect size was overestimated and corrected to a lower level by trim and fill method. That is, the bias might result from some missing studies, or even may not exist at all. In the following sections, we discuss some potential implications of our findings. In general, we cannot make any strong causal claims about the efficacy of CLIL due to the publication bias, insignificant weighted effect size and small sample size.

Another important point to be underlined about this analysis is that among the all studies searched in the literature, the all 6 (which left after screening based on inclusion criteria) are performed in Spain which is an important fact that should be taken in to account in discussing the results. Due to pervasive feeling of dissatisfaction with the state of English language teaching in Spain during the last few years (Vez, 2007), many attempts have been done to improve this challenging situation. Although the project of teaching English at school from early ages (as early as age of four) has been implemented in most of Spanish school, the results were not satisfactory. It ended with language skills far below the desired level at the end of high school and not enough competence to prepare the students to take subjects taught through English at university level (Lasagabaster 2009). The reason could be assigned to the malfunction of educational system and methodology applied for FL teaching. To compensate such failure, CLIL programs are becoming very fashionable in Spain (Lasagabaster 2009). Therefore, our results suggest that there is an obsession with implementation of CLIL technique in Spanish educational system. Strong tendency of Spanish educationists towards the use of CLIL may sacrifice academic achievement of the learners.

On the other hand, comparing the results of the CLIL studies conducted in the communities with already two official languages (e.g. Basque Country, Catalonia and Galicia in Spain) with those of one should be done cautiously. Discussing the advantage of trilingualism over bilingualism is beyond the scope of this paper, but it is worth mentioning that trilinguals develop a larger cognitive supply and experience a higher level of cognitive demands than bilingualism (e.g. a trilingual has to remember even more words and has to inhibit even more languages than a bilingual) (Schroeder and Marian, 2016). Furthermore, trilinguals may show larger gains than bilingual older adults in cognitive reserve (Chertkow et al., 2010; Perquin et al., 2013).

5.1 A fundamental CLIL argument: two-for-the-price-of-one theory

The ambitious endeavor to deliver school-subject content and FL development simultaneously and interdependently fostered the crude theory of 2 for 1 in CLIL

programs (Dalton-Puffer et al., 2010 and Zydari, 2012). The promised expectant outcomes proposed by CLIL proponent were “producing more of everything at low anxiety level” and “developing capacity in the other language, while saving on separate other language classes, with no detriment to their progress in non-language subject content” (Dalton-Puffer, 2007:276; De Graaff et al., 2007, Mehisto et al., 2008 and Van de Craen et al., 2007). This sounds perfect on paper, but in practice many issues come up.

The assumption that the content focus gives more purpose to the FL learning in CLIL courses than in general FL learning (Lorenzo et al., 2010) is evidently violated by frequent drop-outs reported by participants of the program because their FL proficiency cannot cope (Apsel, 2012). The same story was recorded in immersion settings by approximately 20% drop-outs before grade 5 because immersion did not respond well to the needs of all students especially those with learning challenges (Netten and Germain, 2009). Even, there would seem to be no particularly logical reason behind the idea of studying subject content through the medium of a FL or to study the FL with the content without considering whether the students like foreign languages or the content. If one of the objective of implementing such program is achieving proficiency in oral communication, learning FL through the academic subjects cannot be motivating and helpful as in more technical topics, less interaction occurs (Smit, 2010). As a student in Makropoulos's (2010) Canadian French immersion study mentioned: *“I’m not going to be speaking French to somebody about science or something like that, ...”* (p.9).

Lucietto (2008) summarizes CLIL initiatives in 3 options: 1) Learn the FL separately, in order to learn the content through the FL; 2) Learn the FL through the content, which has already been learnt in the L1; 3) Learn the FL and the content together. However, the third option is simply violated by the statement made by Mehisto (2008) as *“no CLIL teachers stated both language and content goals”*. Accordingly, the theory of *“content and FL learning go hand-in-hand”* seems waste of followers’ effort by perusing a wrong path. The idea of using L2 as the medium of instruction is also an absurd effort *“to make unnatural natural”* (Smith, 2005).

Similarly Mehisto et al. (2008) believe that *“common sense seems to say that students studying in a second language cannot possibly learn the same amount of content as students studying in their first language”* (p.20). Furthermore, novel subjects in academic contents cannot stimulate the same amount of oral interaction in L2 as manifested by learners in their L1 (Bruton, 2011; Dalton-Puffer, 2007). Nonetheless, the learners in such situations are on a hiding to nothing as conceptually difficult contents which are above the students’ competence can complicate FL development and FL medium will make content learning more difficult to assimilate and the lack of language can be a serious hindrance for content development (Seikkula-Leino, 2007; Bruton, 2011 and Apsel, 2012). Likewise, acquiring complicated or unfamiliar new concepts might hinder rather than benefit language development and language processing (Coonan, 2007 and Tan, 2011).

Since everything has a price, resorting to their L1 especially in peer work is the first strategy adopted by learners while facing difficult content presented in L2

(Coonan, 2007). The problem could be even more complicated leading to a breakdown due to CLIL's complexity as a task (Mehisto, 2008). Therefore, achieving a threshold in the L2 seems necessary for students and content teachers to be able to cope and to develop adequately in the content (Gierlinger, 2007, Marsh et al., 2000; Hoare 2010 and Várkuti, 2010); otherwise both the language and the content can possibly become problematic. Even the parents cannot help the students at home with the content not well understood in a foreign language medium at school especially when the same content is also available in their L1 (Apsel, 2012).

On the other hand, there are many stories in the literature stating how the development of content ability was negatively affected by exclusiveness of instruction in L2 and how certain types of student suffered from it. For example, the authorities in Turkey stopped the instruction of math and science through the *medium of English* due to the students' complaints about their failure in university entrance exams (Kirkgoz, 2007). According to Hellekjaer (2010), delivering university lectures through L2 appeared to be problematic for 42% of the sampled students.

Regarding the success stories reported from CLIL defenders, it is enough to mention that if any improvement is seen, it is the matter of bilingualism not CLIL. As Blakemore and Frith (2005) declare learning in a CLIL setting results in discrete brain activity by affecting cognitive aspects and brain plasticity in young learners. Hence, such effects could not be the results of CLIL itself. Since the importance of brain functioning has not been recognized efficiently in language pedagogy, all improvements in language learning have been assigned to the methodology rather than the most important factors such as neural activities. Accordingly, not only methodology, educational system and materials should be taken into account but also some decisive factors such as physiological, psychological and neurological aspects of every individual learner should be studied in depth.

The reason behind implementing an increasing number of CLIL programs and an enormous interest in improving students' foreign language command among some European countries like Hungary, Portugal, Italy, the UK and Ireland, and especially Spanish population is their lower level in the grasp of FL (Eurydice, 2006b). But such a haste to embrace global market of multilingualism does not justify taking wrong steps, diverting much of attention to CLIL population and ignoring non-CLIL foreign language teaching, mainstream school FL teaching and non-CLIL students: "It is very possible that deficit FL teaching might become even more deficient, especially for the less academically able, the less linguistically proficient, or the less economically privileged" (Bruton, 2013, p. 595).

5.2 Psychological Implications

Another fact that should be carefully considered in this regards is the impossibility of generalizing the results of the studies that compare CLIL-learners with their EFL counterparts for all general populations. Apart from some environmental and administrative factors such as composition of the class group, degree/type of support

from the administration and diversity of educational settings, other decisive factors like personality of the teacher and especially characteristics of the learner should be taken into account (Lasagabaster, 2011). Learners' identity which has always been taken for granted in CLIL studies, can be the master key to open the door of success to this young methodology. All learners are idiosyncratic in their approach to learning as "*identical like cancer cells in a Petri dish*" (Ting, 2010, page 4). Therefore, the same approach that works for one group of learners cannot be prescribed for the other ones. There is no methodology that works for all population or for all purposes.

No one questions that today most of learning and teaching challenges are solved by contribution of psychology in education. However, psychological factors that have implications for successful teaching and learning are the lost pieces of learning puzzle in CLIL setting. The most important factors of this type include readiness, interest, intelligence, motivation, attitude, feeling, frustration, aptitude, mental health, individual differences, orientation and fear of failure. Apart from motivation that has been studied in several investigations, the rest of factors are rarely or never mentioned in CLIL studies.

On the other hand, effective learning happens when the characteristics of the learners and learning styles are considered and supported in all adopted educational approaches. If the aim of education is to help the learners to develop competency and requisite skills toward autonomy, achieving such goal is not feasible unless the required guidance is provided for them based on their abilities, differences and needs. Accordingly, adopting an educational approach in a large scale project requires setting a plan while catering individual differences and learning styles and considering all psychological factors affecting learning process. Likewise, not only the amount of exposure to L2 but also learners' psychological differences and status should be taken into account before making any comparison between the groups receiving two different educational methodologies.

6. Conclusion and Recommendations

In sum, this meta-analytic study has shed some light on the present studies done on the effectiveness of various aspects of CLIL method. Despite the positive feedbacks reported by researchers on the efficiency of such methodology applied in primary and secondary schools, they occasionally admit that the results of the study are in doubt as the participants of CLIL and non-CLIL groups do not have equal exposure to the foreign language. To compare both groups in the same conditions it is necessary for both to receive the same number of instruction hours in L2.

On the other hand, small sample size and the lack of pretest scores have been the most important criticisms that have been raised against the optimistic attitudes towards the absolute efficacy of CLIL approach. Besides, it is necessary for all educational systems to review regularly the objectives of the CLIL provision set by Eurydice to

check whether all socio-economic, sociocultural, linguistic and educational objectives have been met.

From the insights of this meta-analysis, the researchers recommend further investigations on the effectiveness of CLIL program and conducting longitudinal studies on students' performance to capture and trace the impact of CLIL on learners' academic achievement and competence. Likewise, the findings of large-scale studies with larger sample sizes can be better interpreted, compared and synthesized by providing a more comprehensive picture.

To the best of our knowledge, this meta-analysis gives more evidence about the validity of the studies which investigated the effectiveness of CLIL program on different educational and psychological factors. The results of the present study may have implications for not only educationists and psychologists but also can help the education policymakers to make sound decision in their future educational policy formulation.

6.1 Limitations of the study

One of the limitations of our study is that the authors relied on literature written in a language they know (i.e. English); this excludes literature written in other languages. Likewise, our results might be affected by dissemination bias because we did not search any grey literature. To avoid the threat of publication bias, researchers suggest including gray literature and unpublished studies (e.g., dissertations) for further investigations, which will either counter this threat or at least allow us to evaluate the magnitude of this bias.

Table 1: Summary of the coded studies and associated effect sizes

Study	Country	Number of CLIL participants	Number of non-CLIL participants	Outcome measures	Hedges' g
English achievement and student motivation in CLIL and EFL settings [®] (Lasagabaster, 2011)	Spain (Basque Country)	164	27	Achievement and Motivation	1.182
A preliminary study on motivation and gender in CLIL and non- CLIL types of instruction 46. (Fontecha AF, and Alonso AC, 2012)	Spain (La Rioja)	31	31	Motivation and Gender	-0.753
Lexical transfer in the written production of a CLIL group and a non-CLIL group (Vázquez BM, 2014).	Spain (Andalusia)	18	18	Proficiency	0.741
The effects of content and language integrated learning in European education: key findings from the Andalusian sections evaluation project (Lorenzo, F., Casal, S., Moore, P., 2010).	Spain (Andalusia)	754	448	Proficiency	1.260
Foreign Language competence in CLIL courses (Lasagabaster, D., 2008)	Spain (Basque Country)	170	28	Language Competence	1.185
The impact of CLIL on affective factors and vocabulary learning (Heras and Lasagabaster, 2015)	Spain (Basque Country)	25	21	Affective Factors	0.352

Table 2: Weighed effect size and heterogeneity

Model	Effect size and 95% confidence interval						Test of null (2-Tail)		Heterogeneity			
	Number Studies	Point estimate	Standard error	Variance	Lower limit	Upper limit	Z-value	P-value	Q-value	df (Q)	P-value	I-squared
Fixed	6	1,099	0,056	0,003	0,989	1,209	19,560	0,000	67,373	5	0,000	92,579
Random	6	0,642	0,295	0,087	0,064	1,219	2,177	0,029				

Table 3: Statistics for individual studies

Model	Study name	Outcome	Statistics for each study							Sample size	
			Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Group-A	Group-B
	1,000	Combined	1,182	0,216	0,047	0,759	1,605	5,482	0,000	164	27
	2,000	motivation-	-0,753	0,260	0,067	-1,262	-0,244	-2,898	0,004	31	31
	3,000	lexical	0,471	0,331	0,109	-0,177	1,119	1,425	0,154	18	18
	4,000	linguistic	1,260	0,065	0,004	1,133	1,387	19,409	0,000	754	448
	5,000	language	1,185	0,212	0,045	0,771	1,600	5,599	0,000	170	28
	6,000	affective	0,352	0,293	0,086	-0,222	0,927	1,202	0,230	25	21
Fixed			1,099	0,056	0,003	0,989	1,209	19,560	0,000		
Random			0,642	0,295	0,087	0,064	1,219	2,177	0,029		

Table 4: Duval and Tweedi's trim and fill

	Fixed Effects				Random Effects	
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit
Observed values		1,09916	0,98902	1,20929	0,64176	0,06401
Adjusted values	1	0,51493	0,43166	0,59821	0,49561	-0,22699

Figure 1: Flow of study analysis through different phases of the meta-analysis

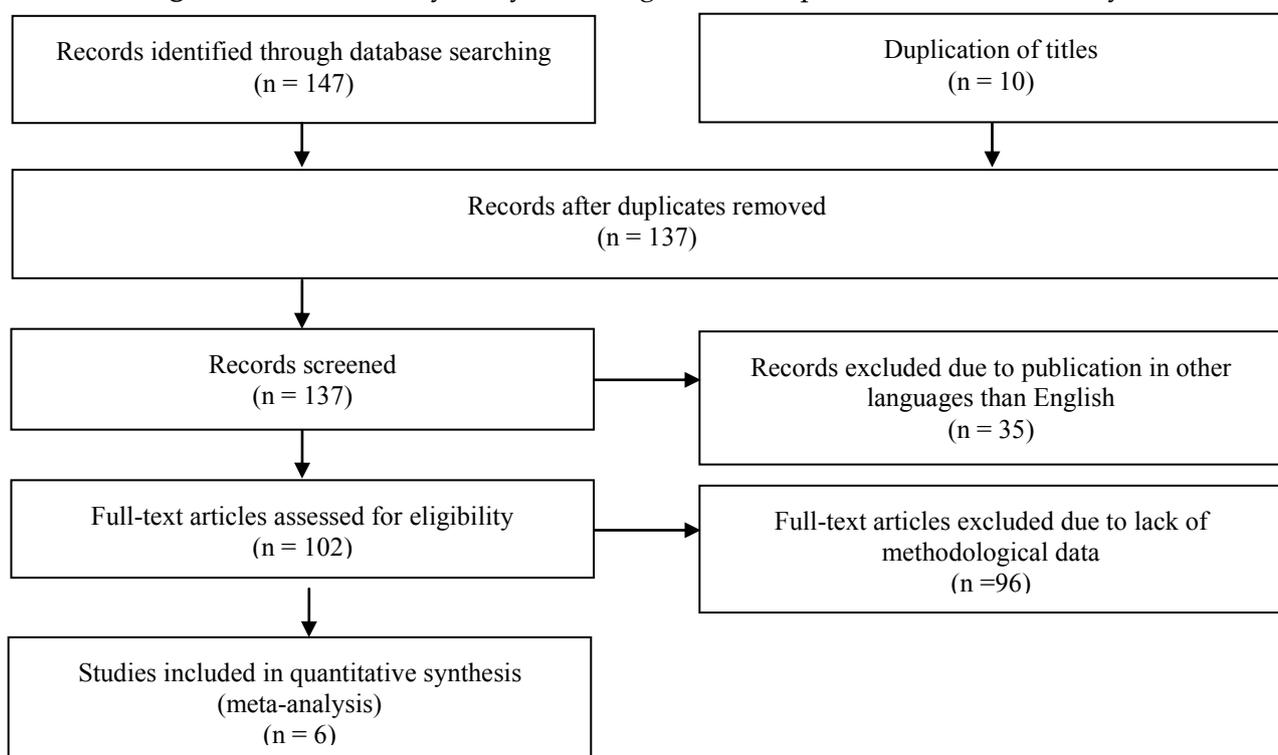


Figure 2: Forest plot for primary studies

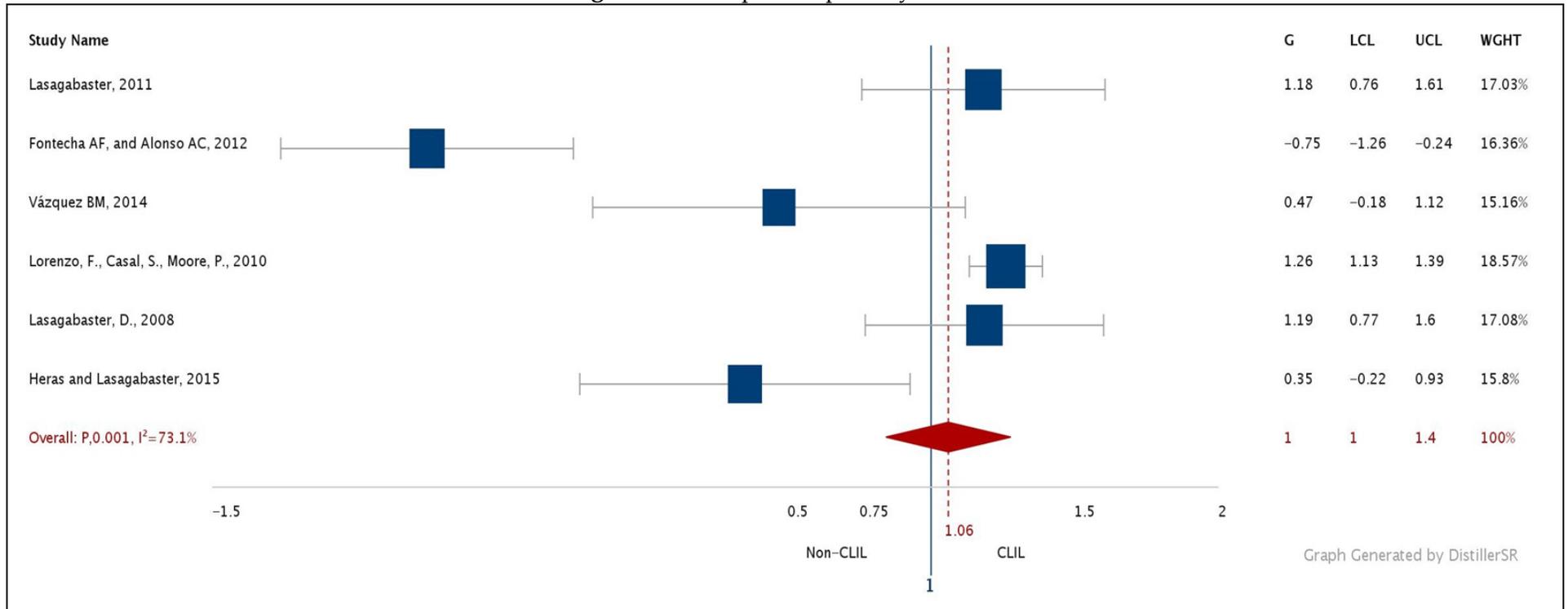
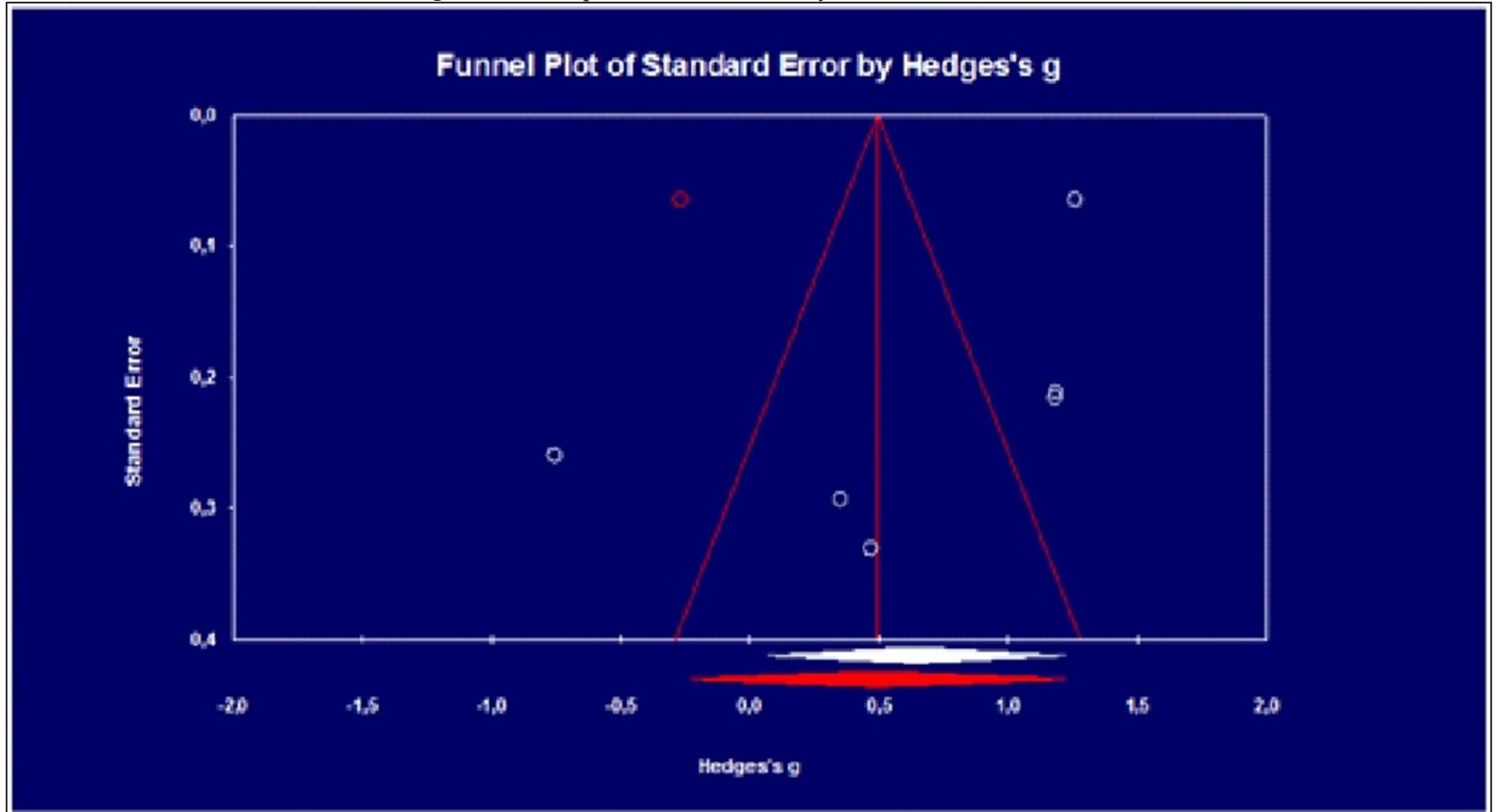


Figure 3: Funnel plot of standard error by effect size for all studies



Conflict of interest

We the authors hereby declare that there are no financial or non-financial conflicts of interest to disclose. We attest to the fact that all Authors listed on the title page have contributed significantly to the work, have read the manuscript, attest to the validity and legitimacy of the data and its interpretation.

Acknowledgement

The authors hereby would like to thank Dr. Helga Braga for reviewing the methodology and assistance in statistical analysis of the manuscript.

Corresponding Author

Fatemeh Tabari, Department of linguistics, Faculty of philosophy and literature, University Autónoma de Madrid 28049- Madrid, Spain.

Email address: fatemeh.tabari@estudiante.uam.es / ella_tabari@yahoo.com

References

- Admiraal, W., Westhoff, G. and de Bot, K. 2006, "Evaluation of bilingual secondary education in The Netherlands: students' language proficiency in English", *Educational Research and Evaluation*, 12:1, 75-93. doi:10.1080/13803610500392160
- Agustín Llach, M.P. (2009a). The role of gender in lexical transfer in EFL written compositions across grades: a preliminary study. Paper presented at the 19th EUROSLA Conference, University College Cork, Ireland, September, 2-5, 2009.
- Agustín Llach, M.P. (2009b). The role of Spanish L1 in the vocabulary use of content and non-content EFL learners. In Y. Ruiz de Zarobe, and R.M. Jiménez Catalán R.M. (Eds.), *Content and Language Integrated Learning: Evidence from Research in Europe*. (pp.112-129). Bristol: Multilingual Matters. doi:10.1111/j.1540-4781.2011.01188.
- Airey, J. 2009, "Estimating undergraduate bilingual scientific Literacy in Sweden". *International CLIL Research Journal*, 1:2, 26-35. <http://www.icrj.eu/12/article3.html>
- Alejo, R. and A. Piquer. (2010). 'CLIL teacher training in Extremadura: A needs analysis perspective' in D. Lasagabaster and Y. Ruiz de Zarobe (eds): *CLIL in Spain. Implementation, Results and Teacher Training*. Cambridge Scholars Publishing, pp. 219-42.
- Alonso, E., Grisalena, J. and campo, A. (2008). "Plurilingual education in Secondary Schools: analysis of results", *International CLIL Research Journal*, 1; 1, 36-49. doi:10.1093/elt/ccr056
- Apsel, C. (2012). Coping with CLIL: dropouts from CLIL streams in Germany. *International CLIL Research Journal* 1, 4. <http://www.icrj.eu/14/article5.html>
- Baetens Beardsmore, H. (2002). 'The significance of CLIL/EMILE' in D. Marsh (ed.): *CLIL/EMILE the European Dimension*. University of Jyvaskyla pp. 24-6.

- Baetens Beardsmore, H. (2008) Multilingualism, Cognition and Creativity. *International CLIL Research Journal*, 1:1, 4- 19. <http://www.icrj.eu/11/article1.html>
- Banegas, D. (2011). Content and language integrated learning in Argentina 2008-2011. *Latin American Journal of Content and Language Integrated Learning*, 4(2), 32-48. [doi:10.5294/laclil.2011.4.2.4](https://doi.org/10.5294/laclil.2011.4.2.4)
- Bereiter, C. (2002). *Education and Mind in the Knowledge Age*. New Jersey: Lawrence Erlbaum Associates, Inc. [doi:10.4324/9781410612182](https://doi.org/10.4324/9781410612182)
- Bernhardt, Markus (2011). Bilingualität und historisches Lernen. Förderung von historischen Kompetenzen oder soziales Differenzkriterium? In: Hodel, Jan / Ziegler, Béatrice (Hg.): *Forschungswerkstatt Geschichtsdidaktik 09. Beiträge zur Tagung «geschichtsdidaktik empirisch 09»*. Bern 2011, pp. 214-223.
- Bernhardt, Markus (2015). "Bilingual History classes – no, thanks!" In: *Public History Weekly* 3 22, [doi: dx.doi.org/10.1515/phw-2015-4268](https://doi.org/10.1515/phw-2015-4268).
- Blakemore, S.J., and Frith, U. (2005). The learning brain: Lessons for education: a précis. *Developmental Science*, 8 (6), 459–465.
- Breidbach, Stephan and Viebrock, Britta (2012) CLIL in Germany: Results from Recent Research in a Contested Field of Education. *International CLIL Research Journal*. Volume 1 (4): 6-16. [doi:10.1111/j.1467-7687.2005.00434](https://doi.org/10.1111/j.1467-7687.2005.00434).
- Brinton, D.M., Snow, M.A., and Wesche, M. (2003). *Content-based second language instruction*. Ann Arbor: The University of Michigan Press. [doi:10.1093/oxfordhb/9780195384253.013.0019](https://doi.org/10.1093/oxfordhb/9780195384253.013.0019)
- Bruton, A. (2011a). Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore (2010). *Applied Linguistics*, 32(2), 236-241. [doi:10.1093/applin/amr007](https://doi.org/10.1093/applin/amr007)
- Bruton, A. (2011b). Is CLIL so beneficial, or just selective? *Re-evaluating some of the research*. *System*, 39(4), 1-10. [doi:10.1016/j.system.2011.08.002](https://doi.org/10.1016/j.system.2011.08.002)
- Bruton, A. (2013). CLIL: Some of the reasons why... and why not. *System*, 41: 587-597. [doi:10.1016/j.system.2013.07.001](https://doi.org/10.1016/j.system.2013.07.001)
- Bruton, A. (2015). CLIL: Detail matters in the whole picture. More than a reply to J. Hüttner and U. Smit (2014). *System*, 53, 119-128. [doi:10.1016/j.system.2015.07.005](https://doi.org/10.1016/j.system.2015.07.005)
- Cabezas Cabello, J. M. 2010. A SWOT analysis of the Andalusian Plurilingualism Promotion Plan (APPP). Proceedings of the 23rd GRETA Convention, 83-91. M. L. Pérez Cañado (Ed.). Jaén: Joxman
- Carless, D. R. (2003). Factors in the implementation of task-based teaching in primary schools. *System*, 31, 485–500. [doi:10.1016/j.system.2003.03.002](https://doi.org/10.1016/j.system.2003.03.002)
- Casal, S., Moore, P., (2009). The Andalusian bilingual sections scheme: evaluation and consultancy. *International CLIL Research Journal* 1 (2), 36-46. <http://www.icrj.eu/12/article4.html>
- Celaya, M. L. (2006). La producción de estructuras sintácticas en Inglés lengua extranjera en contexto AICLE y regular: estudio descriptivo. Paper presented at

- XXIV AESLA Conference. Universidad Nacional de Educación a Distancia (UNED), Madrid, March, 30 - April, 1.
- Celaya, M.L. (2008). "I study natus in English": lexical transfer in CLIL and regular learners". In R. Monroy and A. Sánchez (eds.). *25 Years of Applied Linguistics in Spain: Milestones and Challenges*. Murcia: Servicio de Publicaciones de la Universidad de Murcia.
- Celaya, M. Luz and Ruiz de Zarobe, Yolanda (2010) First Language and Age in CLIL and NonCLIL Contexts. *International CLIL Research Journal*, 1(3), 60-66. <http://www.icrj.eu/13/article6.html>
- Cenoz, J., Genesee, F., and Gorter, D. (2013). *Critical Analysis of CLIL: taking stock and looking forward*. Applied Linguistics. Oxford University Press, 1– 21.
- Council of Europe (2001). Common European Framework of Reference for Languages: Learning, Teaching and Assessment. Cambridge: Cambridge University Press. http://culture2.coe.int/portfolio/documents_intro/common_framework.html
[doi:10.1093/applin/amt011](https://doi.org/10.1093/applin/amt011)
- Chertkow, H., Whitehead, V., Phillips, N., Wolfson, C., Atherton, J., & Bergman, H. (2010). Multilingualism (but not always bilingualism) delays the onset of Alzheimer disease: evidence from a bilingual community. *Alzheimer Disease and Associated Disorders*, 24(2), 118–125. [doi:10.1097/wad.0b013e3181ca1221](https://doi.org/10.1097/wad.0b013e3181ca1221)
- Coyle, D.; Hood, P.; Marsh, D. (2010). *CLIL: Content and Language Integrated Learning*. Cambridge University Press.
- Coyle, D. (2008). CLIL—A pedagogical approach from the European perspective. In N. Van Deusen-Scholl and N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 4: *Second and Foreign Language Education*, pp. 97-111). New York, NY: Springer. [doi:10.1007/978-0-387-30424-3_92](https://doi.org/10.1007/978-0-387-30424-3_92)
- Coyle, D. (2007a). 'Towards a connected research agenda for CLIL pedagogies,' *The International Journal of Bilingual Education and Bilingualism*, 10: 543–62.
- Coyle, D. (2007b). 'The CLIL quality change' in D. Marsh and D. Wolff (eds): *Diverse Contexts-Converging Goals: CLIL in Europe*. Peter Lang, pp. 47–58.
- Coyle, D. (2010). Foreword. In D. Lasagabaster and Y. Ruiz de Zarobe (Eds.) *CLIL in Spain: Implementation, Results and Teacher Training* (pp. vii-viii). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Dale, L. and Tanner, R. (2012) "*CLIL Activities*" Cambridge: Cambridge University Press.
- Dalton-Puffer, C., (2007). Discourse in Content and Language Integrated (CLIL) Classrooms. *Language Learning & Language Teaching*. [doi:10.1075/llt.20](https://doi.org/10.1075/llt.20)
- John Benjamins, Amsterdam De Graaff, R. et al. (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.
- Dalton-Puffer, C. (2007), Discourse in CLIL classrooms. *Language Learning & Language Teaching*. [doi:10.1075/llt.20](https://doi.org/10.1075/llt.20)

- Dalton-Puffer, C. (2008), "Outcomes and processes in CLIL: current research from Europe", in Delanoy, W. and Volkman, L. (eds), *Future Perspectives for English Language Teaching*, Heidelberg, Carl Winter.
- Dalton-Puffer, C., Huttner, J. Schindelegger, V. and Smit, U. (2009). "Technology-Geeks speak out: what students think about vocational CLIL". *International CLIL Research Journal*, 1; 2, 18-25. <http://www.icrj.eu/12/article2.html>
- Dalton-Puffer, C., Nikula, T. and Smit, U.: (2010). Charting policies, premises and research on Content and Language Integrated Learning, in C. Dalton-Puffer, T. Nikula and U. Smit (eds.), *Language Use and Language Learning in CLIL Classrooms*, (1-19). John Benjamins, Amsterdam. [doi:10.1075/aals.7.01dal](https://doi.org/10.1075/aals.7.01dal)
- Anne-Marie de Meija (2002). *Power, Prestige and Bilingualism: International Perspectives on Elite Bilingualism (= Bilingual Education & Bilingualism)*. Clevedon: Multilingual Matters. 322 pp., ISBN Pb 185359590X. [doi:10.1558/sols.v5i1.151](https://doi.org/10.1558/sols.v5i1.151)
- Doiz, A., Lasagabaster, D., and Sierra, J. M. (2014). CLIL and motivation: The effect of individual and contextual variables. *The Language Learning Journal*, 42(2), 209-224. [doi:10.1080/09571736.2014.889508](https://doi.org/10.1080/09571736.2014.889508)
- Duval S, Tweedie R. (2000a). Trim and fill: a simple funnel plot based method of adjusting for publication bias in meta-analysis. *Biometrics*.56: 455-463. [doi:10.1111/j.0006-341x.2000.00455.x](https://doi.org/10.1111/j.0006-341x.2000.00455.x)
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634. [doi:10.1136/bmj.315.7109.629](https://doi.org/10.1136/bmj.315.7109.629)
- Ellis, E. (2003). Bilingualism among teachers of ESL: A study of second language learning experience as a contributor to the professional knowledge and Elizabeth Ellis 192 beliefs of teachers of ESL to adults. Unpublished PhD thesis, Griffith University, Brisbane.
- Eurydice (2006b) National summary sheets on education systems in Europe and ongoing reforms: Finland. Directorate-General for Education and Culture, European Commission, July 2006. [doi:10.7788/ijbe.2007.3334.1.227](https://doi.org/10.7788/ijbe.2007.3334.1.227)
- Eurydice (2010), National system overviews on education systems in Europe and ongoing reforms, 2010 Edition, Eurydice. [doi:10.7788/ijbe.2007.3334.1.227](https://doi.org/10.7788/ijbe.2007.3334.1.227)
- Fernández-Fontecha, A., (2014). Motivation and gender effect in receptive vocabulary learning: An exploratory analysis in CLIL primary education. *Latin American Journal of Content and Language Integrated Learning*, 7(2), 27-49. doi:10.5294/laclil.2014.7.2.2 eISSN 2322-9721
- Gajo, L. 2007, "Linguistic knowledge and Subject knowledge: how does bilingualism contribute to subject development?" *International Journal of Bilingual Education and Bilingualism*, 10:5,563-581. <http://www.tandfonline.com/doi/abs/10.2167/beb460.0>
- Fontecha AF, and Alonso AC (2012). A preliminary study on motivation and gender in CLIL and non- CLIL types of instruction. *IJES*, vol. 14 (1), 2014, pp. 21-36. [doi:10.6018/ijes/14/1/156681](https://doi.org/10.6018/ijes/14/1/156681)

- García Mayo, M.P. and A. Lázaro Ibarrola. 2005. "Pronouns revisited: Evidence from the morphosyntactic development of English as an L3". In F. Garrudo Carabias and J. Comesaña (eds.) *Estudios de Filología Inglesa en honor de Antonio Garnica Silva* (pp. 217-228). Sevilla: Universidad de Sevilla.
- Gené-Gil, M., Juan-Garau, M. y Salazar-Noguera, J. (2015). Writing development under CLIL provision. in *Content-based language learning in multilingual educational environments*, eds. M. Juan-Garau y J. Salazar-Noguera, 139-162. Berlin: Springer. [doi:10.1007/978-3-319-11496-5_9](https://doi.org/10.1007/978-3-319-11496-5_9)
- Gierlinger, E. M. (2007). The three pillar of modular CLIL: Findings from an Austrian research project. In D. Marsh and D. Wolff (Eds.), *diverse contexts – converging goals. CLIL in Europe* (pp. 211-226). Frankfurt: Peter Lang.
- Gobierno de La Rioja. (2004). Convocatoria de Proyectos de Innovación Lingüística en Centros. Curso 2004/2004. Consejería de Educación, Cultura y Deporte. 19 May 2004.
- Hellekjaer, G. O. 2004. Unprepared for English-medium instruction: *A critical look at beginner students. Integrating content and language: Meeting the challenge of a multilingual higher education*, 147-171. R. Wilkinson (Ed.). Maastricht: University of Maastricht. [doi:10.1515/cercles-2013-0009](https://doi.org/10.1515/cercles-2013-0009)
- Harrop, E. (2012). Content and Language Integrated Learning (CLIL): Limitations and possibilities. *Encuentro*, 21, 57-70. <http://w.encuentrojournal.org/textos/7.%20Harrop.pdf>
- Heras, A and Lasagabaster D (2015). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*. Vol. 19(1) 70–88 [doi:10.1177/1362168814541736](https://doi.org/10.1177/1362168814541736)
- Hoare, P. (2000). Hoare' Content-based language teaching in China: contextual influence on implementation'. *Multilingual Multicultural Dev.*, 31/1 (2010), pp. 69–86 [doi:10.1080/01434630903367207](https://doi.org/10.1080/01434630903367207)
- Hondris, G. I., S. Vlahavas, and Demetriadis.2007. 'Negotiation of meaning and digital textbooks in the CLIL classroom' in D. Marsh and D. Wolff (eds): *Diverse Contexts-Converging Goals: CLIL in Europe*. Peter Lang, pp. 319–330.
- Hood, P. (2006). Unpublished data from research interviews with students at Tile Wood Hill Language College, Coventry, UK, in Coyle et al 2010, *CLIL*, Cambridge, CUP.
- Hood, P. and Tobbutt, K. (2009). *Modern languages in the primary school*, London, SAGE. [doi:10.4135/9781446269466](https://doi.org/10.4135/9781446269466)
- Hüttner, J.and Rieder-Bünemann, A., (2007). The effects of CLIL instruction on children's narrative competence. *Views* 16 (3), 20e27. http://anglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/weitere_Uploads/Views/views1903CLILspecial.pdf
- Lasagabaster, D., 2008. Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal* 1, 30e41 [doi:10.2174/1874913500801010030](https://doi.org/10.2174/1874913500801010030)

- Jäppinen, A.-K. (2002) Ajattelu ja Sisältöjen Oppiminen Vieraskielisessä Opetuksessa. Raportti 1/3 [Thinking and Content Learning in Teaching through a Foreign Language. Report 1/3]. University of Jyväskylä. Jyväskylä: Jyväskylän yliopistopaino.
- Jäppinen, A.-K. (2005b). Cognitive Development of Mathematics and Science in the Finnish Mainstream Education in Content and Language Integrated Learning (CLIL) – Teaching through a Foreign Language. *Language and Education* 19:2.
- Jäppinen, A. (2005). Development and Structure of the Finnish Education System. Paper presented in the international conference Finland in PISA studies – Reasons behind the results. Helsinki.
- Jäppinen, A. K. (2006). CLIL and future learning. In S. Björklund, K. MårdMiettinen, M. Bergström and M. Södergård (eds.), Exploring dual-focussed education. *Integrating language and content for individual and societal needs* (September 1st, 2013).
- Jiménez, R. M. and Ojeda, J. (2008a). CLIL instruction and EFL learners' productivity in an availability task. Paper presented at the International CLIL Conference: CLIL fusion multilingual mindsets in a multicultural world. Building quality learning communities. Tallin, Estonia, October, 24– 25.
- Johnson, Jacqueline S. and Newport, Elissa L. (1989). Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language. *Cognitive Psychology*, 21, 60-99. [doi:10.1016/0010-0285\(89\)90003-0](https://doi.org/10.1016/0010-0285(89)90003-0)
- Kaliaa V., Wilbournb M. P. and Ghio K. (2014) Better early or late? Examining the influence of age of exposure and language proficiency on executive function in early and late bilinguals. *Journal of Cognitive Psychology*, 26 (7). DOI: 10.1080/20445911.2014.956748
- Kenney, J. M., Hancewicz, E., Heuer, L., Metsisto, D., and Tuttle, C. L. (2005). *Literacy strategies for improving mathematics instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Kirkgoz, Y. (2007). English language teaching in Turkey: Policy changes and their implementation. *Regional Language Centre Journal*, 38: 216-228. [doi:10.1177/0033688207079696](https://doi.org/10.1177/0033688207079696)
- Krashen, S. (1982) Newmark's "Ignorance Hypothesis" and current second language acquisition theory. Unpublished manuscript.
- Krechel, L.: (1999). Sprach- und Textarbeit im Rahmen von flexiblen bilingualen Modulen. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*. [online] <http://zif.spz.tu-darmstadt.de/jg-04-2/beitrag/krechel1.htm>
- Kukkonen P. (2006). Toward a working theory for physics CLIL classroom. *Exploring Dual-Focussed Education*, 147-157.
- Lamsfuss-Schenk, S., (2008). *Fremdverstehen im bilingualen Geschichtsunterricht: Eine Fallstudie*, Peter Lang, Frankfurt am Main

- Lasagabaster, D. (2008). Foreign Language competence in CLIL courses. *The Open Applied Linguistics Journal*, 1, 31-42. http://laslab.org/upload/foreign_language_competence_and_language_integrated_courses.pdf
- Lasagabaster, D. and J. M. Sierra. 2009. 'Language attitudes in CLIL and traditional EFL classes'. *International CLIL Research Journal* 2/1: 4-17. <http://www.icrj.eu/12/article1.html>
- Lasagabaster, D., and J.M. Sierra. (2010). Immersion and CLIL in English: More differences than similarities. *ELT Journal* 64, no. 4: 367-75. [doi:10.1093/elt/ccp082](https://doi.org/10.1093/elt/ccp082)
- Lasagabaster, D. (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in Language Learning and Teaching*, 5, 3-18. [doi:10.1080/17501229.2010.519030](https://doi.org/10.1080/17501229.2010.519030)
- Lightbrown, P. and Spada, N. (2006) *How Languages are Learned*, Oxford, OUP.
- Lorenzo, F. (2008) Instructional discourse in bilingual settings: an empirical study of linguistic adjustments in CLIL? *Language Learning Journal*, 36;1,21-23 [doi:10.1080/09571730801988470](https://doi.org/10.1080/09571730801988470)
- Lin, A. and Man E (2009). *Bilingual Education: Southeast Asian Perspectives*, Hong Kong: Hong Kong University Press.
- Lorenzo, F.; Hengst, H.; Hernández, H.; Pavón, V. (2005). Borrador para la elaboración del Currículo Integrado, Plan to Promote Multilingualism, Junta de Andalucía. Retrieved November 22, 2012, from the World Wide Web: www.juntadeandalucia.es/averroes/plurilinguismo/curriculo/borradorcil.pdf
- Lorenzo, F.; Casal, S.; de Alba y Moore P. (eds.) (2007) Models and Practice in Content and Language Integrated Learning. *RESLA. Revista Española de Lingüística Aplicada*, vol. 20. Número monográfico.
- Lorenzo, F. (2007). An analytical framework of language integration in L2 content-based courses. *Language and Education* 21: 502-513. [doi:10.2167/le708.0](https://doi.org/10.2167/le708.0)
- Lorenzo, F. (2008). Instructional discourse in bilingual settings: an empirical study of linguistic adjustments in CLIL. *Language Learning Journal*, 36;1,21-23 [doi:10.1080/09571730801988470](https://doi.org/10.1080/09571730801988470)
- Lorenzo, F., (2010). CLIL in Andalusia. In: Lasagabaster, D., Ruiz de Zarobe, Y. (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training*. Cambridge Scholars Publishers, Newcastle, pp. 2e11. [doi:10.1093/elt/ccr056](https://doi.org/10.1093/elt/ccr056)
- Lorenzo, F., Casal, S., Moore, P., (2010). The effects of content and language integrated learning in European education: key findings from the Andalusian sections evaluation project. *Applied Linguistics*, 31, 418e442. [doi:10.1093/applin/amp041](https://doi.org/10.1093/applin/amp041)
- Lucietto, S., 2008. A model for quality CLIL provision. *International CLIL Research Journal* 1 (1), 83e92. <http://www.icrj.eu/11/article7.html>
- Lyster, R. (2007). *Content and Language integrated teaching: a counterbalanced approach*, John Benjamin, Amsterdam. [doi:10.1111/j.1540-4781.2009.00870.6](https://doi.org/10.1111/j.1540-4781.2009.00870.6)

- Macdonald C. and Moodie P. (2006). Language, textbooks and science content learning in South African schools – the case of the Science for All Project. *Exploring Dual-Focussed Education*, 158-175.
- Makropoulos. (2010). Students' attitudes to the secondary French immersion curriculum in a Canadian context. *Language, Culture and Curriculum*, 23(1), 1-13. [doi:10.1080/07908310903494525](https://doi.org/10.1080/07908310903494525)
- Manzano, B. (2014). Lexical Transfer in the Written Production of a CLIL Group and a Non-CLIL Group. *International Journal of English Studies*, 14(2), 57-76. [doi:10.6018/j.166251](https://doi.org/10.6018/j.166251)
- Marsh D. and Langé, G. (Eds.). (2000). *Using languages to learn and learning to use languages*. Finland: University of Jyväskylä.
- Marsh, David (2000): An Introduction to CLIL for Parents and Young People in Ten Languages. In David Marsh and Gisella Langé, eds. *Using Languages to Learn and Learning to Use languages*. Jyväskylä: u of Jyväskylä.
- Marsh, K.T. Hau, C.K. Kong' (2000) Late immersion and language of instruction in Hong Kong high schools: achievement growth in language and non-language subjects' *Harv. Educ. Rev.*, 70 (2000), pp. 302–346 [doi:10.17763/haer.70.3.gm047588386655k5](https://doi.org/10.17763/haer.70.3.gm047588386655k5)
- Marsh, D (Ed) (2002) CLIL/EMILE- The European Dimension: Actions, Trends and Foresight Potential Public Services Contract DG EAC: European Commission
- Marsh, D. (2002). CLIL/EMILE: *The European dimension; Actions, trends and foresight potential*. Jyväskylä (Finland): UniCOM, Continuing Education Centre, University of Jyväskylä.
- Marsh, D. (2002). CLIL/EMILE – The European Dimension. Actions, Trends and Foresight Potential, Strasbourg, European Commission.
- Marsh, David (2008). Language Awareness and CLIL. J. Cenoz, N.H. Hornberger (Eds.), *Encyclopedia of Language and Education. Knowledge about Language* (pp.233-246). II edition, Volume 6, Springer Science and Business Media LLC, New York [doi:10.1007/springerreference_60061](https://doi.org/10.1007/springerreference_60061)
- Marsh, David (2008). *CLIL in primary East Asia Contexts: Primary Innovations in EastAsia: Indonesia, Malaysia, Thailand, Philippines*. British Council, East Asia. Bangkok: British Council. <https://www.teachingenglish.org.uk/sites/teacheng/files/download-accessenglish-publications-proceedings-bangkok-2008.pdf>
- Mehisto, P. (2007). What a school needs to consider before launching a CLIL programme: the Estonian Experience. In D. Marsh and D. Wolff (eds.) *Diverse Contexts, Converging Goals: CLIL in Europe*. Frankfurt: Peter Lang. 61–78.
- Mehisto, P., (2008). CLIL counterweights: recognizing and decreasing disjuncture in CLIL. *International CLIL Research Journal* 1 (1), 93e119. <http://www.icrj.eu/11/article8.html>

- Mehisto, Peeter, Marsh, David, Frigols Maria Jesús (2008): *Uncovering CLIL: Content and Language Integrated Learning in Bilingual and Multilingual Education*. Oxford: Macmillan.
- Mehisto, P. and Marsh, D. (2011). "Approaching the Economic, Cognitive and Health benefits of Bilingualism: Fuel for CLIL", in Y. Ruiz de Zarobe, J. Sierra and F. Gallardo del Puerto (eds.). *Content and Foreign Language Integrated Learning*. Bern: Peter Lang, 21-48.
- Mentz, O. (2010). Alle Sprachen eignen sich – oder doch nicht? Überlegungen zu einem bilingualen Fächerkanon. In: Doff, S. (ed.) *Bilingualer Unterricht in der Sekundarstufe*. Eine Einführung, (29-43). Narr, Tübingen.
- Met. M. (1998). Curriculum decision-making in content-based language teaching. In J. Cenoz and F. Genesee (Eds). *Beyond Bilingualism: Multilingualism and Multilingual Education* (pp. 35-63). Clevedon, Eng.: Multilingual Matters. [doi:10.2307/417241](https://doi.org/10.2307/417241)
- Muñoz, C. (2006) Influence translinguistique et changement de code dans la production d'une L4. *Acquisition et Interaction en Langue Étrangère (AILE)*, 24, 75-99.
- Muñoz C. (2006a). The effects of age on foreign language learning: The BAF Project. In C. Muñoz (ed.), *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters, 1-40.
- Muñoz C. (ed.) (2006b). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.
- Muñoz C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics* 24/4: 578-596. [doi:10.1093/applin/amm056](https://doi.org/10.1093/applin/amm056)
- Murtagh, L. (2007). Out-of-school use of Irish, motivation to learn the language and proficiency in immersion and subject-only post-Primary programmes. *International Journal of Bilingualism and Bilingual Education*, 10(4), 428–453. [doi:10.2167/beb453.0](https://doi.org/10.2167/beb453.0)
- Naves, T., and Victori, M. (2010). CLIL in Catalonia: An overview of research studies. In D. Lasagabaster and Y. Ruiz de Zarobe (eds.), *CLIL in Spain, Implementation, Results and Teacher Training* (p. 30-54). Cambridge: Cambridge Scholars. [doi:10.1093/elt/ccr056](https://doi.org/10.1093/elt/ccr056)
- Netten, J., Germain, C., 2009. The future of intensive French. *Canadian Modern Language Review* 65, 757e786. [doi:10.3138/cmlr.65.5.757](https://doi.org/10.3138/cmlr.65.5.757)
- Nikolov, M. and Mihaljević Djigunović, J. (2006) Recent research on age, second language acquisition and early foreign language learning. *ARAL* 26, 234–260 [doi:10.1017/s0267190506000122](https://doi.org/10.1017/s0267190506000122)
- Nonaka, I. and D. Teece (Eds) (2001). *Managing Industrial Knowledge. Creation, Transfer and Utilization*. London: SAGE Publications Ltd. <http://dx.doi.org/10.4135/9781446217573>
- Orafi, S. and Borg, S. (2009) Intentions and realities in implementing communicative curriculum reform. *System*, 37, 243-253 [doi:10.1016/j.system.2008.11.004](https://doi.org/10.1016/j.system.2008.11.004)

- Pavón Vázquez, V. and Gaustad, M. (2013) Designing bilingual programmes for higher education in Spain: Organisational, curricular and methodological decisions. *International CLIL Research Journal*, 2 (1), 82-94
<http://www.icrj.eu/21/article7.html>
- Pavón Vázquez, V. and Ellison, M. (2013) Examining teacher roles and competences in content and language integrated learning (CLIL). *LINGVARVMARENA*, 4, 65-78.
- Pérez-Vidal, C. (2002). "Spain", in M. Grenfell (ed.) *Modern Languages across the Curriculum*. London and New York: Routledge/ Falmer, 114-130.
- Pérez Cañado, M. L. 2012. "CLIL research in Europe: Past, present, and future". *International Journal of Bilingual Education and Bilingualism* 15(3): 315-341.
[doi:10.1080/13670050.2011.630064](https://doi.org/10.1080/13670050.2011.630064)
- Pérez-Vidal, C. and Juan-Garau, M. (2010). To CLIL or not to CLIL? From bilingualism to multilingualism in Catalan/Spanish communities in Spain. In Y. Ruiz de Zarobe and D. Lasagabaster (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training* (pp. 115-138). Newcastle Upon-Tyne, United
- Scardamalia, M. (2001). Getting Real About 21st Century Education. *The Journal of Educational Change* 2, 171-176.
- Schroeder, S.R., & Marian, V. (2016). Cognitive consequences of trilingualism. *International Journal of Bilingualism* June 25, 2016 doi: 10.1177/1367006916637288
[doi:10.1177/1367006916637288](https://doi.org/10.1177/1367006916637288)
- Stehr, N. (2001). Modern Societies as Knowledge Societies. In: *Handbook of Social Theory*, 494-508. Eds G. Ritzer and B. Smart. Lontoo: Sage Publications.
[doi:10.4135/9781446215548.n2](https://doi.org/10.4135/9781446215548.n2)
- Ting, Yen-Ling Teresa (2010): CLIL appeals to how the brain likes its information: examples from CLIL-(Neuro) Science, In: *International CLIL Research Journal* 1/3, 1-18. URL: <http://www.icrj.eu/13-73> (Retrieved July 17, 2011)
- Kingdom: Cambridge Scholars Publishing. [doi:10.1093/elt/ccr056](https://doi.org/10.1093/elt/ccr056)
- Rothman, J. and Guijarro-Furtes. (2010). Input Quality Matters: Some Comments on Input Type and Age-Effects in Adult SLA. *Applied Linguistics* 31, (2). pp. 301-306. [doi:10.1093/applin/amq004](https://doi.org/10.1093/applin/amq004)
- Ruiz de Zarobe, Y. (2007). CLIL in a bilingual community: similarities and differences with learning English as a foreign language. *View[z]* 16 (3), 47e52.
- Ruiz de Zarobe, Y. and Jimenez Catalan, R.M. (eds.) (2009) *CLIL: evidence from research in Europe*, Bristol, Multilingual Matters.
- Ruiz de Zarobe, Y., Lasagabaster, D., (2010). CLIL in a bilingual community: the basque autonomous community. In: Lasagabaster, D., Ruiz de Zarobe, Y. (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training*. Cambridge Scholars Publishing, Newcastle, pp. 12e29. [doi:10.1093/elt/ccr056](https://doi.org/10.1093/elt/ccr056)
- Rymarczik, J.: (2003). *Kunst auf Englisch. Ein Plädoyer für die Erweiterung des bilingualen Sachfachkanons*. Langenscheidt-Longman, München.

- San Isidro, X., 2010. An insight into Galician CLIL: provision and results. In: Lasagabaster, D., Ruiz de Zarobe, Y. (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training*. Cambridge Scholars Publishing, Newcastle, pp. 55e78. [doi:10.1093/elt/ccr056](https://doi.org/10.1093/elt/ccr056)
- Schwartz, J. L., and Kenney, J. M. (1995). *Assessing mathematical understanding and skills effectively*. Interim report of the Balanced Assessment Program, Harvard Graduate School of Education, Cambridge, Massachusetts.
- Serra, C. (2007), "Assessing CLIL at Primary School: a longitudinal study", *International Journal of Bilingual Education and Bilingualism*, 10:5,582-602.
- Seikkula-Leino, J. (2007). CLIL learning: Achievement levels and affective factors. *Language and Education* 21: 328-341. [doi:10.2167/le635.0](https://doi.org/10.2167/le635.0)
- Smit, U. (2010). CLIL in an English as a lingua franca classroom: on explaining terms and expressions interactively. In C. Dalton-Puffer, T. Nikula, and U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 259-278). Amsterdam: John Benjamins. [doi:10.1075/aals.7.13smi](https://doi.org/10.1075/aals.7.13smi)
- Smith, K. (2005), «Is this the end of the language class? », *Guardian Weekly*, January 21, 2005. Available at: <http://education.guardian.co.uk/tefl/teaching/story/0,15085,1394830,00.html>.
- Simões, A.R., Pinho, A.S., Costa, A.M. and Costa, A.R. (2013). The Project English Plus: a CLIL approach in a Portuguese school. *Indagatio Didactica*, 5 (4), 30-51
- Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessments, and adjustments* (pp. 75–98). West Sussex, England: John Wiley and Sons, Ltd.
- Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessments, and adjustments* (pp 99–110). West Sussex, England: John Wiley and Sons, Ltd. [doi:10.1002/0470870168.ch6](https://doi.org/10.1002/0470870168.ch6)
- Sterne JA, Gavaghan D, Egger M: Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol*. (2000), 53: 1119-1129. 10.1016/S0895-4356(00)00242-0.
- Swain, M. and Lapkin, S. (2005). "The evolving socio-political context of immersion education in Canada: some implications for programme development", *International Journal of Applied Linguistics*, 15:2, 169-186. [doi:10.1111/j.1473-4192.2005.00086.x](https://doi.org/10.1111/j.1473-4192.2005.00086.x)
- Tedick, D. J. and L. Cammarata (2012). Content and Language Integration in K–12 Contexts: Student Outcomes Teacher Practices, and Stakeholder Perspectives. *Foreign Language Annals* 45(S1), S28-S53. [doi:10.1111/j.1944-9720.2012.01178.x](https://doi.org/10.1111/j.1944-9720.2012.01178.x)
- Thornton A, Lee P: Publication bias in meta-analysis: its causes and consequences. *J Clin Epidemiol*. (2000). 53: 207-216. 10.1016/S0895-4356(99)00161-4. [doi:10.1016/s0895-4356\(99\)00161-4](https://doi.org/10.1016/s0895-4356(99)00161-4)

- Van de Craen, P.; Lochtman K.; Ceuleers, E.; Mondt, K.; Allain, L. (2007a). "An interdisciplinary approach to CLIL learning in primary schools in Brussels". In DaltonPuffer, C.; Smit, U. (eds.). *Empirical Perspectives on CLIL Classroom Discourse*. Frankfurt am Main: Peter Lang, 253-274. [doi:10.3726/978-3-653-01829-5/11](https://doi.org/10.3726/978-3-653-01829-5/11)
- Van de Craen, P.; Ceuleers, E.; Mondt, K.; (2007b). "Cognitive development and bilingualism in primary schools: teaching maths in a CLIL environment". In Marsh, D.; Wolff, D. (eds.). *Diverse Contexts – Converging Goals. CLIL in Europe*. Frankfurt am Main: Peter Lang, 185-200.
- Várkuti' A. (2010). Linguistic benefits of the CLIL approach: measuring linguistic competences'Int. CLIL Res. J., 1/3 (2010), pp. 67–79 <http://www.icrj.eu/13/article7.html>
- Vázquez BM (2014). Lexical transfer in the written production of a CLIL group and a non-CLIL group. *IJES*, vol. 14 (2), 2014, pp. 57-76 [doi:10.6018/j.166251](https://doi.org/10.6018/j.166251)
- Villarreal Olaizola, I. and García Mayo, M.P. (2009) Tense and agreement morphology in the interlanguage of Basque/Spanish Bilinguals: CLIL vs. non-CLIL. In Y. Ruiz de Zarobe and R. M. Jiménez Catalan (eds.), *Content and Language Integrated Learning: Evidence from Research in Europe* (pp. 157-175) [doi:10.1111/j.1540-4781.2011.01188.x](https://doi.org/10.1111/j.1540-4781.2011.01188.x)
- Krogh, G., K. Ichijo and I. Nonaka (2000). *Enabling Knowledge Creation. How to Unlock the Mystery of Tacit Knowledge and Release the Power of Innovation*. Oxford: University Press
- Wesche, Marjorie Bingham. (2002). Early French Immersion: How has the original Canadian model stood the test of time? In: Petra Burmeister, Thorsten Piske and Andreas Rohde (eds.) 2002. *An integrated view of language development*. Papers in honour of Henning Wode. Trier: Wissenschaftlicher Verlag Trier, 357-379. [doi:10.1017/s0272263105360207](https://doi.org/10.1017/s0272263105360207)
- Wiesemes, R. (2009). Developing theories and practices in CLIL: CLIL as post-method pedagogies. In *Content and Language Integrated Learning: Evidence from research in Europe*, eds. Y.R. De Zarobe, and R.M.J. Catalán, 41-62. Bristol: Multilingual Matters.
- Wilhelmer, Nadja. (2010). Content and language integrated learning (CLIL) in the mathematical setting. *VIEWS* 19 (3): 97-10. http://anglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/weitere_Uploads/Views/views1903CLILspecial.pdf
- Witzigmann, S.: (2011). *Bildende Kunst in der Zielsprache Französisch als Einstieg ins bilinguale Lernen und Lehren. Explorative Studie einer fünften Realschulklasse*. Peter Lang, Frankfurt am Main.
- Vez, J.M. (2007). De políticas (y politiquillas) del EEES y competencias idiomáticas en el grado de maestro. *Lenguaje y Textos* 25: 13-42.

Zydati, W. (2012). Linguistic thresholds in the CLIL classroom? the threshold hypothesis revisited. *Int. CLIL Res. J.*, 1/4 Retrieved: <http://www.icrj.eu/14/article2.html>

Creative Commons licensing terms

Author(s) will retain the copyright of their published articles agreeing that a Creative Commons Attribution 4.0 International License (CC BY 4.0) terms will be applied to their work. Under the terms of this license, no permission is required from the author(s) or publisher for members of the community to copy, distribute, transmit or adapt the article content, providing a proper, prominent and unambiguous attribution to the authors in a manner that makes clear that the materials are being reused under permission of a Creative Commons License. Views, opinions and conclusions expressed in this research article are views, opinions and conclusions of the author(s). Open Access Publishing Group and European Journal of Foreign Language Teaching shall not be responsible or answerable for any loss, damage or liability caused in relation to/arising out of conflicts of interest, copyright violations and inappropriate or inaccurate use of any kind content related or integrated into the research work. All the published works are meeting the Open Access Publishing requirements and can be freely accessed, shared, modified, distributed and used in educational, commercial and non-commercial purposes under a [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).