



ELSEVIER

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

When CLIL is for all: Improving learner motivation through peer-tutoring in Mathematics

Otilia Martí Arnándiz, Lidon Moliner^{*}, Francisco Alegre

Jaume I University, Av. Sos Baynat s/n, 12071, Castellon, Spain

ARTICLE INFO

Keywords:

CLIL motivation
Peer-tutoring
Secondary education
Gender
Mathematics

ABSTRACT

The obligatory provision of CLIL is raising concerns about the discouraging effects it may have. Yet little research so far has explored the impact of collaborative classroom methodologies on learners' motivation towards both content and language in integration. The present study sets out to examine the potential benefits of peer-tutoring in raising motivation towards CLIL in Mathematics for Compulsory Secondary Education. A pretest-posttest with control group design was adopted. Participants were 408 students from grades 7, 8 and 9 in a state-run secondary school based in the Valencian Community (Spain). Two factors were analyzed: intrinsic motivation (i.e. positive attitude and classroom enjoyment) and motivational force (i.e. confidence and anxiety). Eight focus groups provided qualitative data. Findings point to statistical significant increases for the experimental group in both factors but with moderate effect sizes. The peer-tutoring experience proves more positive for the younger participants (7 graders) and for enhancing males' intrinsic motivation and females' motivational force. The qualitative results obtained from the focus groups also suggest that these pedagogical practices may reduce students' anxiety, build their self-confidence and encourage more positive attitudes towards CLIL. Implementing peer-tutoring in CLIL, then, might improve students' motivation towards Mathematics through English.

1. Introduction

Motivation ranks high in the purported benefits of Content and Language Integrated Learning (CLIL) as understood by education policymakers in Europe. When stating that CLIL “increases learners’ motivation and confidence in both the language and the subject being taught” (European Commission, 2012), this supranational organism aligns with the perception of leading researchers in the field (Coyle, 2006; Coyle, Hood, & Marsh, 2010; Darn, 2006; Meyer, 2010). Such belief in the motivating power of CLIL provisions—held by scholars, policymakers and stakeholders alike (Hüttner, Dalton-Puffer, & Smit, 2013)—is so widespread that, in Lasagabaster’s words, “probably because motivation is taken for granted, there are not that many systematic studies on the interaction between CLIL and motivation” (2019, p. 349).

Actually, despite having been accounted for in a number of research agendas for CLIL (Coyle, 2007; Dalton-Puffer & Smit, 2013; Pérez-Cañado, 2012; Ruiz de Zarobe, 2015), investigations focused on students’ motivation in this learning environment do not abound. As the next section will show, there is a clear mismatch between expectations and empirical evidence. This is especially so when this matter is approached from the prism of how collaborative work like peer-tutoring may bolster learner motivation towards

^{*} Corresponding author.

E-mail address: mmoliner@uji.es (L. Moliner).

<https://doi.org/10.1016/j.system.2022.102773>

Received 8 April 2021; Received in revised form 24 February 2022; Accepted 24 February 2022

Available online 28 February 2022

0346-251X/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

CLIL, an area where studies are still scarce but much needed.

2. (De-)Motivation in CLIL

2.1. Language bias and methodological flaws

A closer look at the extant literature (Lasagabaster, 2019; Sylvén, 2017, for an overview) reveals some important insights. Drawing on second language motivation theories like Dörnyei's (2005, 2009), a first strand of studies conceives motivation as the enhancement of favorable attitudes and positive affect towards the CLIL vehicular language, which more often than not is English. Apart from this language bias, mixed results have been obtained. Language interest and motivation seem to be generally boosted (Li & Zhang, 2020; Massler, 2012) and CLIL students tend to be more motivated than non-CLIL cohorts (Arribas, 2016; Denman, Tanner, & Graaff, 2013; Hunt, 2011; Lasagabaster, 2011; Lasagabaster & Sierra, 2009; Lorenzo, Casal, & Moore, 2010). But the opposite is also true, with studies like Fernández Fontecha and CangaAlonso (2014) where the English as a Foreign Language (EFL) lessons appear to be more motivating than the CLIL ones. Moreover, some strong negative reactions against CLIL are gathered among Italian secondary students (Coonan, 2012) or Finnish upper-primary pupils, who show higher degrees of anxiety and lower self-esteem as language learners (Seikkula-Leino, 2007). Increasing anxiety becomes intellectual helplessness in the case of Polish primary children studying Mathematics and Science through English (Otwinowska & Foryś, 2017). A state of negative affect causes a reduction on pupils' intrinsic motivation. Moreover, it seriously hinders students' cognitive abilities needed for concentrating, understanding classroom input or problem-solving and, therefore, increases their frustration and resentment over CLIL.

Even when positive findings in terms of language motivation are reported, this line of enquiry presents some methodological flaws. First, the lack of pre-CLIL data on subjects' motivation prevents us from unequivocally attributing the increasing levels of motivation observed to the program *per se*. That is, without baseline data, what is being measured is motivation *for* instead of motivation *from* CLIL (Mearns, de Graaf, & Coyle, 2017). In elected or selective CLIL programs, in which students volunteer to participate or are admitted for their prior language ability and/or academic achievement, no cause-effect relationship between CLIL and motivation can be established since here learners should be regarded as self-motivated (Bruton, 2011a, 2011b, 2013; Broca, 2016; Rumlich, 2014, 2017; Sylvén & Thompson, 2015). Second, as shown by Doiz, Lasagabaster and Sierra's (2014a) research about the effect of students' age, gender and socio-cultural background on motivation and CLIL, caution should be taken whenever some relevant individual or contextual variables, other than the CLIL experience itself, are not controlled.

2.2. Learners' individual differences and type of content subjects

Increasing attention has been paid to gender, one of the previously neglected individual variables with a potential impact on motivation towards CLIL. The working hypothesis here is that CLIL could narrow the gender gap existing between foreign languages as a feminized domain, where girls are more intrinsically motivated and perform better (Carr & Pauwels, 2006; Jones, 2009; Kissau, 2007; Kissau & Wierzalis, 2008; Kobayashi, 2002; López Rúa, 2006; Norton & Pavlenko, 2004; Pavlenko & Piller, 2008; Sunderland, 2010), with respect to some content subjects like Physical Education, Science or Mathematics, which are presented as more male-oriented (He, Zhou, Salinitri, & Xu, 2019; Leyva, 2017; McCall, 2012).

Bearing this in mind, some research about the impact of gender on motivation towards CLIL has been recently undertaken. Findings from these studies, though, are neither definite nor unanimous. Some empirical evidence confirms that the gender gap is reduced (Fernández Fontecha & CangaAlonso, 2014; Heras & Lasagabaster, 2015; Martínez Agudo, 2021; Sylvén & Thompson, 2015) or even inverted (Mearns & de Graaff, 2018). Yet, other studies show no difference in motivation between males and females (Merisuo-Storm, 2007) or maintain the superior motivation of girls in both CLIL and non-CLIL groups (Gallardo-del-Puerto & Blanco-Suárez, 2021; Lasagabaster, 2008; Roquet, Llopis, & Pérez-Vidal, 2016). Besides, once again, most of this research on gendered-motivational patterns suffers from a clear language bias. As indicated by Ushioda and Dörnyei (2017) and Lasagabaster and Doiz (2020), most of the motivation studies in the last years were set in contexts where the focus was on learning English as the target language, thus providing evidence of a strong language bias in current theorizing and research in the CLIL motivation field. By and large, studies mainly concentrate on the extent to which male subjects improve their motivation towards the target language —exceptions include Fernández Barrionuevo and Baena Extremera (2018) who take into account gendered-perceptions towards both foreign language learning and Physical Education.

Starting with proficiency in the target language, it appears that good grades in the English subject, mainly based on conversational competence or Cummins' (1979) notion of Basic Interpersonal Communication Skills (BICS), are poor predictors of the degree of anxiety and intellectual helplessness that even participants in elitist CLIL programs may suffer (Apsel, 2012; Otwinowska & Foryś, 2017; Seikkula-Leino, 2007; Zydariß, 2012). As claimed by Zydariß (2012), a double language threshold would be required to successfully deal with demanding content learning through an additional language. BICS accounts for the lower language threshold, while Cummins' (1979) Cognitive Academic Language Proficiency (CALP) represents the upper one. Limited proficiency closely linked to CALP would be at the root of CLIL learner's most serious academic difficulties and related anxiety problems (Apsel, 2012; Zydariß, 2012).

As far as the type of CLIL content subject is concerned, Cenoz, Genesee and Gorter deplore the fact that "use of the language might actually diminish if the subject matter is novel and/or complex" (2014, p. 15). Content of the CLIL subject itself has been shown to have a major impact on upper-primary students' negative affectivity. In Otwinowska and Foryś' (2017) study, for example, previous grades in Mathematics and Science *do* predict intellectual helplessness. A report on stakeholders' views of CLIL in Catalonia attests that most

parents are against including “serious” subjects like Science in their children’s primary curriculum just to avoid loss of core curricular content (Pladevall-Ballester, 2015).

The real (or perceived) academic obstacles entailed in learning cognitively demanding content subjects do not only evoke widespread fears (Dalton-Puffer, 2011), but have brought into question the feasibility of implementing CLIL programs for younger learners. Following the example set by Finland (Moate, 2014), researchers like Lorenzo et al. (2010), Otwinowska and Forys (2017) or Zydatiņ (2012) posit the need to postpone CLIL to lower-secondary or even to upper-secondary education. Conversely, other voices argue in favor of an early introduction of CLIL programs in infant and primary education (Nikula, 2017; Pérez, Lorenzo, & Pavón, 2016; Siqueira, Landau, & Paraná, 2018).

2.3. Classroom methodologies: the importance of collaborative work

Taken together, the need to cope with limited language proficiency and complex content subject brings to the fore the third cause explaining CLIL students’ (de-)motivation, namely, classroom methodologies (Halbach, 2009). Deficient classroom management and poorly adapted materials (Otwinowska, 2013) or teacher-centered and monological lessons (Lim Falk, 2008, as reported in; Sylvén, 2017) are pinpointed among the factors responsible for creating an unsatisfactory learning environment. To Sylvén, a distinction should be made between the ideal *versus* the reality of CLIL, where the latter is “in many cases far from this ideal, motivational environment” (2017, p. 60). To Somers and Llinares (2018), though, these learning conditions have nothing to do with CLIL, real or otherwise. Rather, they reduce “CLIL from immersion to submersion, at which point what happens in the classroom can no longer be termed ‘CLIL’” (Somers & Llinares, 2021, p. 4).

Understanding what really happens in the classroom is paramount to identify and reproduce the pedagogical conditions required to ensure the success of “CLIL for all”, one the main tenets of this teaching approach (Genesee & Hamayan, 2016; Mehisto, Marsh, & Frigols, 2008). After all, in CLIL, “it would not be effective to teach the same content, the same way, with another language” (Pavón Vázquez & Ellison, 2013, p. 72). Attention to CLIL methodologies, then, is long overdue at least for two reasons. First, the absence of clear guidelines in its implementation (Dalton-Puffer, 2011), due to the “terminological and pedagogical vagueness of CLIL” (Pérez Cañado, 2016, p. 18), propitiates grossly misinterpretations of a global education approach whose pedagogical dimension has already been insisted upon (Baetens-Beardsmore, 2001; Coonan, 2007; Coyle, 2008; de Bot, 2002; San Isidro, 2018). Second, the fact that, at the beginning of the second decade of the 21st century, “democratising CLIL” (Escobar Urmeneta & Evnitskaya, 2013, p. 113) is not an aspiration but a reality makes the exploration of CLIL classroom methodologies an urgent task.

Therefore, the advent of CLIL for all students explains why the focus of some ongoing research like the ADiBE project (which brings together key figures with ample experience in the field of CLIL from Spain, Austria, Finland, Germany, Italy, and the UK) has geared towards how educational practices in this teaching environment can meet the demands of diverse learners (Madrid & Pérez-Cañado, 2018). Some methodological approaches in CLIL experiences included the use of cooperative learning. Outcomes of qualitative research on CLIL learning environments have attested that motivation increases whenever students participate in interactive activities where they explain processes to each other (Ballester-Roca & Spaliviero, 2021; Huang, 2011), contents are negotiated with the teacher (Banegas, 2012, 2013) or collaborative techniques are implemented (Coyle, 2013; Czura & Ankiewicz, 2018; Lasagabaster & Doiz, 2016).

However, existing research has tended to focus on group rather than pair work. This unbalance in favor of team work can be attributed to the interest raised by the implementation of project-based learning in CLIL (Cáceres & Unigarro, 2007; Lasagabaster & López Beloqui, 2015; Pastrana, 2019; Pistorio, 2010; del Pozo, 2019; Sánchez-García & Pavón-Vázquez, 2021; Sierra, 2011, 2016). This research gap of pair work and CLIL, then needs to be filled because when CLIL secondary students are asked about the importance they attach to peer support, they deplore that most teachers overlook the fact that “students among themselves always understand one another better than when the teacher explains it” (Bauer-Marschallinger et al., 2021, p. 10). In this sense, peer tutoring (a pair work methodology) seems mostly needed in a context like ours where CLIL is for everybody and for any content, including “harder” academic subjects like Mathematics. Besides, recent research conducted in Spain (Busse, 2017; Campillo, Sánchez, & Miralles, 2019; Nikula & Moore, 2019) and, more specifically, in the Valencian Community (Guillamón-Suesta & Renau, 2015), has raised serious concerns about students’ low motivation towards these obligatory CLIL programs. Alarming, negative attitudes towards CLIL persist after students’ high-school days and even after their college graduation (Merino & Lasagabaster, 2018). Hence, searching for classroom methodologies that ensure and sustain the success of CLIL is key, as collaborative classroom practices may not only produce deeper content and language learning (Meyer, Coyle, Imhoff & Connolly, 2018), but also reinforce students’ positive attitudes towards CLIL (Alejo & Piquer-Pfritz, 2016).

Precisely, the starting point of the current study is addressing the call for investigating the impact of group dynamics on learner motivation towards CLIL (Doiz, Lasagabaster, & Sierra, 2014b; Lasagabaster, 2019). CLIL motivation is understood here as the integrated construct, proposed by Somers and Llinares (2021), in which both content and language learning are taken into account in an interdependent way. As these authors note, “CLIL motivation does not consist of ‘L2 learning motivation + subject learning motivation’ because, in their own words, “neither construct is any longer recognized to exist in its own right, but has been assimilated into the construct of an integrated content and language learning motivation” (Somers & Llinares, 2021, p. 843). The teaching/learning method under analysis, peer-tutoring, has proven beneficial in foreign language learning (Hwang, Nguyen, & Pham, 2019; Duran, Flores, Oller & Ramírez, 2019) but remains underexplored in relation to CLIL motivation (Lialikhova, 2019). Peer-tutoring may be defined as the acquisition of academic contents or skills through active help and support between pairs of students (Topping, 2015). The student with higher knowledge at that moment serves as tutor and the other acts as tutee. This way, both help each other to learn and learn themselves by so doing (Moliner & Alegre, 2020).

3. Aims and research questions

The current study endeavors to measure the impact of classroom methodology (peer-tutoring *versus* teacher-fronted lessons) on CLIL motivation towards Mathematics learnt through English in Compulsory Secondary Education (CSE). Differences regarding motivational trends are also discussed in connection with participants' age and gender. With these aims in mind, two main factors (intrinsic motivation and motivational force) are analyzed, according to the following research questions:

RQ1: Are students who do peer-tutoring more motivated towards CLIL Mathematics than those in teacher-fronted lessons?

RQ2: How does age affect motivation towards CLIL Mathematics in peer-tutoring?

RQ3: How does gender affect motivation towards CLIL Mathematics in peer-tutoring?

On the basis of the previous literature reported above, it is expected that students who participate in peer-tutoring during Mathematics sessions through English (henceforth, the experimental group) will be more motivated towards CLIL than those in the control group, who only receive teacher-fronted classes. We also hypothesize, in response to RQ2, that statistical significant differences for increasing levels of intrinsic motivation and motivational force will be reported in favor of older subjects in the experimental group who, supposedly, are more linguistically and cognitively prepared for CLIL Mathematics. Finally, regarding participants' gender (RQ3), we assume that the peer-tutoring intervention might cause a transfer in the opposite motivational profiles exhibited by males and females in Mathematics and in English when these are taught separately. It is hoped that building a cooperative classroom environment will bring a reduction in both the boys' dislike for English and the girls' anxiety and lack of confidence in Mathematics.

4. The study

4.1. Context and participants

CLIL provision in Spain is not only a top-down enterprise but also a very rich and diverse one (Lasagabaster & Ruiz de Zarobe, 2010; Otto & San Isidro, 2019). Since 2012, CLIL has been part of the school curricula in the Valencian Community, a bilingual region in Spain where both Spanish and Catalan are official languages. A new plurilingual decree, developed by the Valencian educational government since 2015 (Castells & Mestre-Mestre, 2015) and passed by the regional parliament in April 2018, issues that by September 2020 all students, from infant to upper-secondary education, should learn from two to three content subjects through a foreign language. The choice of both the content subject and the language medium are in the hands of school principals.

The current research was conducted in a state-run secondary school based in a medium-sized Valencian city (50,000 inhabitants) in which Mathematics is learnt via CLIL. Mathematics became the CLIL content subject thanks to the pedagogical training of the Maths teachers and their Common European Framework of Reference (CEFR) C2 level in English.

The study involved a total of 408 students enrolled in the first three years of CSE in the Spanish educational system: 136 from grade 7 (12–13 year olds), 136 from grade 8 (13–14 year olds) and 136 from grade 9 (14–15 year olds). At the beginning of the study, the average age was 13.96 years old with a standard deviation of 1.89 years. Unlike this equal division for age, the gender distribution was not balanced with 259 females (63.48%) *versus* 149 males (36.52%). As many as 171 pupils (41.91% of the sample) belong to migrant families coming from Romania, Morocco, China or Guinea, whose home languages are not Spanish or Catalan but Romanian (21.32%), Arabic (18.87%), Chinese (0.98%) or Portuguese (0.49%). Participants' socioeconomic profile was considered average in the light of, first, the National Institute data for the neighborhood where the secondary school is located and, second, the fact that 59.34% of students' parents have completed their university studies—a percentage close to the OECD (Organization for Economic Co-operation and Development) indicator of 60.20% of Spanish population with tertiary education. Students were randomly assigned to the experimental (doing peer-tutoring) or to the control group (receiving teacher-fronted lessons). Thus, the experimental group consisted of 68 subjects and the experimental group was made of 68 subjects in each grade. Since none of the pupils had ever experienced any CLIL-based instruction before, our study offers baseline data collected at the onset of a CLIL obligatory program. This is the case because, prior to the entry into force of the current plurilingual decree, instruction in a foreign language for non-linguistic subjects was only compulsory for certain schools classified as “plurilingual centers”. The highschool in which this intervention took place was allowed to teach Mathematics through English because it had been previously qualified as a “plurilingual center”. Consequently, while there are 3 h of English language lessons a week, CLIL Mathematics comprises 4 h of lessons a week in grades 7 and 9 and three weekly sessions in grade 8. As for the participants' exposure to the target language before entering secondary education, they had attended EFL lessons for 3 h a week since grade 1. Determining their proficiency level, though, is difficult because no official exams have to be taken to get their level of English certified before they finish their primary studies.

4.2. Research design

Given that experimental designs may act as moderators in peer-tutoring experiences influencing the academic or psychological outcomes (Alegre, Moliner, Maroto & Lorenzo-Valentin, 2019c; Leung, 2015), the present study opts for a pretest posttest control group design, so that effect sizes of the intervention are not overestimated.

4.2.1. Sample access and sample power

With respect to sample access, and considering the administrative, ethical and organizational difficulties acknowledged by Zirkel, Garcia, and Murphy (2015) or Etikan and Bala (2017), participants were selected through clustered sampling. In this procedure, the total population is divided into groups (secondary schools in our case). Although the groups are mutually homogenous, they are

internally heterogeneous (Celik, 2017). Once the state-run secondary school was selected, authorization to participate in the study was obtained from the Valencian Government of Education, the school principal and the students' families. Ethical standards given by the Ethics Committee of the Spanish National Research Council (CSIC) were followed.

StudySize 3.0 software by Creostat HB was performed to establish the sample power. A sample power of .89 was determined when using inferential statistics (Students' t-test and Analysis of Variance) with a significance level of 0.05 with 408 participants.

4.2.2. Peer-tutoring treatment and implementation

As far as the treatment is concerned, during the first trimester of the academic year, all participants in CLIL Mathematics received teacher-fronted lessons with students working and sitting individually. The second and third semesters saw the introduction of peer tutoring for the experimental group.

In order to achieve the higher possible psychological benefits during the implementation of peer-tutoring (Alegre, Moliner, Maroto, & Lorenzo-Valentin, 2019a), reciprocal and same age tutoring were chosen (Sytsma, Panahon, & Houlihan, 2019). First, reciprocal tutoring –in which students switch roles, going from tutor to tutee– was preferred over fixed peer-tutoring, where the maintenance of roles (i.e. tutors are always tutors and tutees are always tutees) may decrease tutees' self concept as the permanent recipients of the tutor's help (Alegre, Moliner, Maroto, & Lorenzo-Valentin, 2019b). Second, we opted for same age tutoring *versus* cross age tutoring because the latter might make the younger peers feel uncomfortable for having to work with unknown upper grade students (Alegre, Moliner, Maroto & Lorenzo-Valentin, 2019c).

Drawing on Zeneli and Tymms' (2015) and Leung's (2019) expertise on temporal and organizational issues, the peer-tutoring interactions lasted no more than 20 min in three to four weekly sessions held for approximately 6 months (two trimesters). Besides, the same teacher was in charge of the experimental and control groups in each grade. Content learning in terms of exercises and problems were also identical for both groups, save from the classroom practices implemented, namely, peer-tutoring in the experimental group *versus* teacher-fronted lessons in the control one.

Before the implementation of peer-tutoring, subjects in the experimental group were trained to follow Topping's (2015) protocols and principles, including "pause, prompt and praise" techniques or mutual respect and patience during tutoring sessions. Following Duran, Corcelles, Flores, and Miquel (2019), learners' interactions were closely monitored. After all, teachers play a vital role in ensuring that peer-tutoring encounters are rich in academic language and effective with students really helping each other.

In reciprocal peer-tutoring, the academic gap between participants must be minimized. The objective of pairing students with similarly achieving peers is guaranteeing their ease and comfort (De Backer, Van Keer, & Valcke, 2015). Accordingly, students' first trimester Mathematics grades were used to order pupils and form a list. The first two students in that list were the first pair, the next two, the second pair and, so on, until all pairs were formed.

A typical session in the experimental group started by investing the first 20 min in reviewing homework, correcting it on the blackboard and presenting new content. Afterwards, 10 min were given for learners to do from two to three exercises and one or two problems. Extra material was ready for fast finishers but no teacher help was provided at this stage. Then, peer-tutoring was implemented for 20 min, during which pupils had to work in pairs, share their results and solve together the activities they had not finished yet. If any pair of students could not complete an exercise or a problem by themselves, the teacher's help could be requested as many times as needed. The recommended language of student-student and student-teacher interaction was English. Finally, the teacher devoted the last five to 10 min to solve the assigned tasks on the blackboard.

4.2.3. Instruments and data analysis

In order to answer the three research questions guiding this study, we adopted a mixed-method approach (Guetterman, Babchuk, Howell Smith, & Stevens, 2019), which combines both quantitative and qualitative data.

Quantitative data were elicited by means of a written questionnaire drawn on Doiz, Lasagabaster, and Sierra (2014a). This instrument was selected because its psychometric properties, validity and reliability had been previously tested (see Lasagabaster, 2016; Somers & Llinares, 2021). Nevertheless, as the instrument was tailored to meet our aims, Gudmundsson's (2009) guidelines for translating and adapting psychological instruments were followed in order to retest its reliability. An Alpha Cronbach value of 0.88 for the intrinsic motivation factor and one of 0.89 for the motivational force factor were obtained. These Alpha Cronbach coefficients, similar to the values above 0.72 in all cases reported by Doiz et al. (2014a, p. 216), confirm the reliability of the two scales. Given that the purpose of our research was to explore the impact of CLIL methodologies on learner motivation, only two out of the six factors or subscales considered in the original survey were incorporated. These subscales were intrinsic motivation (4 items) and motivational force (9 items), two of the factors most closely linked to the CLIL learning experience. Although other factors such as instrumental orientation and interest in foreign languages/cultures could have been tested, we decided to focus on those which, in our view, might help us to remove the language bias we wanted to avoid. Besides, in line with Csizér and Dörnyei (2005) we believe that our subjects were too young to be influenced by job prospective, an "incentive motive" (Dörnyei, 2000) that might be blurred or even fade away when what is being measured is not the usefulness of learning English but that of learning Mathematics through English.

Accordingly, all items in the survey were rewritten to incorporate the CLIL motivation construct proposed by Somers and Llinares (2021). The factor of intrinsic motivation, which has to do with classroom enjoyment, was measured by statements such as: "I like to learn Mathematics in English" (item 1); or, "I enjoy Mathematics in English classes" (item 2). In turn, motivational force, a factor combining students' anxiety and motivational strength (Doiz et al., 2014a, p. 217), includes statements like: "I never feel confident when I have to speak in English during Mathematics in English class" (item 1) or "I feel so nervous during Mathematics in English classes that I forget even what I know" (item 5). Reversely coded items were also included (e.g. "Many times I feel bored during Mathematics in English class").

The items, presented on a five-point Likert scale ranging from 1 (“absolutely disagree”) to 5 (“absolutely agree”), were translated into Catalan, the main vehicular language at the school and many participants’ first language. The questionnaire was administered in class during the weekly tutorial and under the researchers’ supervision. Each subject took up to 10 min on average to complete it. In order to avoid any Hawthorne effect (Torgerson & Torgerson, 2007), students were not informed of the purpose of the study. Furthermore, for them to be honest, the researchers made clear that the teacher would not see their responses, which would be treated in the strictest confidence.

The statistical analysis was performed in SPSS 25. In order to gauge whether data were normally distributed in the experimental and control groups (Banerjee & Pradhan, 2018), we ran Kolmogorov Smirnov tests. Means, standard deviations, and student’s *t*-test (95% confidence level) were calculated to find out inter and intra-group differences. Analyses of variance (ANOVAs) were also conducted to compare results per year (i.e. 7, 8 and 9 grades). In case statistical significant differences were found in the ANOVA, an Scheffe post-hoc test was carried out (Daniel, 2019). Multiple linear regression analyses were also performed taking gains (difference between posttest and pretest scores) as the independent variable and experimental condition (0 = control, 1 = experimental), gender (0 = male, 1 = female), and grades (grade 7 = -1, grade 8 = 0 and grade 9 = 1) as the dependent variables (Kaplan & Garner, 2020). This statistic indicates the percentage of the variance in the dependent variable that the independent variables explain collectively to Effect sizes were calculated and Hedge’s *g* was used as a measure of effect size (Simpson, 2019).

Qualitative information was gathered through focus groups (Anwaruddin, 2016) with a total of 24 participants (8 groups of 3 students each) who were randomly selected. The main purpose of these focus groups was to supplement the information obtained from the quantitative analysis. As Bryman (2017) holds, qualitative information *does* play a vital role when combined with quantitative research because it strengthens the verification of the quantitative results. In a similar vein, Clark (2019) reasons that some unique insights and a better understanding of the issues at hand can be gained when integrating quantitative and qualitative methods. So, to gain a deeper knowledge on the motivational processes that took place during the peer-tutoring sessions (Watzek, Anselmann, & Mulder, 2019), the researchers formulated open questions such as: “Why do you think that you feel more or less comfortable during Mathematics in English classes?” or “How did you interact with your peers during the tutoring sessions?” Same as with the questionnaire, the focus groups, lasting 20 min, were conducted in Catalan and held during tutorials but in private spaces. A content analysis was carried out (Cabrera & Reiner, 2018) by means of ATLAS.ti 8. Students’ contributions were coded as follows: Number of focus group and grade. For example, FG3_7 refers to focus group number 3 with participants from grade 7.

5. Results and discussion

5.1. RQ1: peer-tutoring versus teacher-fronted CLIL lessons

Descriptive results of the two factors under consideration are displayed in Table 1, for intrinsic motivation and, in Table 2, for motivational force.

Since, according to the Kolmogorov Smirnov test we ran, our data were normally distributed ($p = .87$), inferential statistics by means of *t*-tests were chosen. An asterisk marks the tests in which statistical significant differences were reported.

With respect to the first research question posed (i.e. Are students who do peer-tutoring more motivated towards CLIL Mathematics than those in teacher-fronted lessons?), Table 3 presents the comparison between the experimental and the control groups in relation to intrinsic motivation. As shown in Table 3, no inter-group statistical significant differences were observed in the pretest for any of the grades (tests 1 to 3). Yet, overall statistically significant improvements *did* appear for the experimental group between the pretest and the posttest stage. These improvements were also reported for 7 graders, but not for 8 or 9 graders in the experimental group (tests 4 to 7). Analogous results were reported when analyzing the gains (difference between the posttest and the pretest scores) in the comparison between experimental and control group (tests 8 to 11) No statistically significant improvements were found between the pretest and the posttest for the control group, either overall or per grade (tests 12 to 15). Linear regression analysis resulted in a R-squared value of 0.51 including experimental condition as the only dependent variable ($p < .05$) and 0.54 including experimental

Table 1
Descriptive results for intrinsic motivation towards CLIL.

	Experimental group			Control group		
	Mean	SD	<i>N</i>	Mean	SD	<i>n</i>
7th grade pretest	3.36	1.04	68	3.49	1.11	68
7th grade posttest	3.74	0.78	68	3.32	1.13	68
8th grade pretest	3.20	1.03	68	3.29	1.22	68
8th grade posttest	3.42	0.77	68	3.43	0.98	68
9th grade pretest	3.46	1.20	68	3.31	0.88	68
9th grade posttest	3.78	1.05	68	3.41	1.00	68
Male pretest	3.23	1.17	76	3.42	1.05	73
Male posttest	3.69	0.88	76	3.43	0.96	73
Female pretest	3.44	1.04	128	3.18	0.98	131
Female posttest	3.47	0.98	128	3.37	0.89	131
All grades pretest	3.34	1.10	204	3.36	1.08	204
All grades posttest	3.64	0.88	204	3.39	1.04	204

Table 2
Descriptive results for motivational force towards CLIL.

	Experimental group			Control group		
	Mean	SD	N	Mean	SD	n
7th grade pretest	2.85	0.89	68	2.97	0.87	68
7th grade posttest	3.16	0.76	68	2.94	0.73	68
8th grade pretest	3.05	0.83	68	3.00	0.81	68
8th grade posttest	3.14	0.85	68	3.12	0.86	68
9th grade pretest	3.03	0.70	68	2.96	0.80	68
9th grade posttest	3.13	0.72	68	2.78	0.81	68
Male pretest	3.02	0.75	76	2.72	0.82	73
Male posttest	3.05	0.82	76	2.81	0.81	73
Female pretest	3.01	1.04	128	3.22	0.78	131
Female posttest	3.29	0.98	128	3.09	0.76	131
All grades pretest	2.98	0.81	204	2.98	0.83	204
All grades posttest	3.14	0.78	204	2.95	0.81	204

Table 3
Student's *t*-tests between groups for intrinsic motivation towards CLIL.

Test	Group A	Group B	<i>t</i> (sig.)
1	7th grade experimental group pretest	7th grade control group pretest	0.70 (p = .48)
2	8th grade experimental group pretest	8th grade control group pretest	0.46 (p = .64)
3	9th grade experimental group pretest	9th grade control group pretest	0.83 (p = .41)
4	7th grade experimental group posttest	7th grade experimental group pretest	2.14 (p = .02)*
5	8th grade experimental group posttest	8th grade experimental group pretest	1.41 (p = .16)
6	9th grade experimental group posttest	9th grade experimental group pretest	1.65 (p = .19)
7	7th, 8th, & 9th grade experimental groups posttest	7th, 8th, & 9th grade experimental groups pretest	3.04 (p < .01)*
8	7th grade experimental group gains	7th grade control group gains	2.02 (p < .05)*
9	8th grade experimental group gains	8th grade control group gains	0.51 (p = .62)
10	9th grade experimental group gains	9th grade control group gains	0.62 (p = .54)
11	7th, 8th, & 9th grade control group gains	7th, 8th, & 9th grade control groups gains	2.15 (p = .03)*
12	7th grade control group posttest	7th grade control group pretest	-0.88 (p = .38)
13	8th grade control group posttest	8th grade control group pretest	0.74 (p = .76)
14	9th grade control group posttest	9th grade control group pretest	1.18 (p = .24)
15	7th, 8th, & 9th grade control groups posttest	7th, 8th, & 9th grade control groups pretest	0.29 (p = .78)

condition and grade as dependents variables (p < .01). A Hedge's *g* effect size of 0.26 for intrinsic motivation was reported for the experience.

Table 4 summarizes outcomes from student's *t*-tests regarding motivational force towards CLIL. Results in Table 4 also show no inter-group statistical significant differences in the pretest for any of the grades (tests 1 to 3). Overall statistically significant improvements were reported between the pretest and the posttest for the experimental group. Besides, as in Table 3, statistical significant improvements were only exhibited by the youngest participants (7 graders) in the experimental group (tests 4 to 7) and overall for the

Table 4
Student's *t*-tests between groups for motivational force towards CLIL.

Test	Group A	Group B	<i>t</i> (sig.)
1	7th grade experimental group pretest	7th grade control group pretest	0.80 (p = .43)
2	8th grade experimental group pretest	8th grade control group pretest	0.36 (p = .72)
3	9th grade experimental group pretest	9th grade control group pretest	0.54 (p = .59)
4	7th grade experimental group posttest	7th grade experimental group pretest	2.18 (p = .03)*
5	8th grade experimental group posttest	8th grade experimental group pretest	0.62 (p = .53)
6	9th grade experimental group posttest	9th grade experimental group pretest	0.82 (p = .41)
7	7th, 8th, & 9th grade experimental groups posttest	7th, 8th, & 9th grade experimental groups pretest	2.03 (p = .04)*
8	7th grade experimental group gains	7th grade control group gains	2.40 (p = .02)*
9	8th grade experimental group gains	8th grade control group gains	0.15 (p = .88)
10	9th grade experimental group gains	9th grade control group gains	1.28 (p = .20)
11	7th, 8th, & 9th grade control group gains	7th, 8th, & 9th grade control groups gains	2.14 (p = .03)*
12	7th grade control group posttest	7th grade control group pretest	-0.22 (p = .83)
13	8th grade control group posttest	8th grade control group pretest	0.84 (p = .40)
14	9th grade control group posttest	9th grade control group pretest	-1.30 (p = .19)
15	7th, 8th, & 9th grade control groups posttest	7th, 8th, & 9th grade control groups pretest	-0.37 (p = .71)

experimental group. Analogous results were found when analyzing the gains for the experimental and the control group (tests 8 to 11). Although overall statistical significant differences were reported for the experimental group, only significant improvements were reported for 7 graders. No statistically significant improvements were found between the pretest and the posttest for the control group, either overall or per grade (tests 12 to 15). Linear regression analysis resulted in a R-squared value of 0.54 including experimental condition as the only dependent variable ($p < .05$) and 0.59 including experimental condition and grade as dependents variables ($p < .05$). A Hedge's g effect size of 0.24 for motivational force was reported for the experience.

Qualitative data gathered in the focus group sessions are highly consistent with the quantitative results in [Tables 3 and 4](#) above. Concerning intrinsic motivation, that is, the attitudinal dimension of motivation towards peer-tutoring in CLIL Mathematics, participants' attitudes are told to have improved substantially. This positive change transpires in the following answers:

- "If at least we can help each other is not that bad" (FG2_7).
- "No one likes Mathematics in English classes. If at least I have someone with me, it's not that bad" (FG6_9).

Actually, it seems that working with a peer provides a familiar routine that facilitates pupils' interactions:

- "Working with a classmate is way better than doing it alone" (FG4_8).
- "Me and Josh (pseudonym) always want to work together. It was great to have that chance in Mathematics" (FG1_7).

Besides, some students express their willingness to continue with the experience in the future:

- "We wouldn't mind to take another year of Mathematics in English. It wasn't that bad eventually" (FG4_8).
- "If we can work together is less hard. It is really hard to listen to all the explanation in English. Yes, you can always ask the teacher, but you can't ask all the time" (FG5_8).

Moving to the second factor, namely, motivational force (i.e. anxiety and motivational strength), participants report to feel more confident and less nervous after the peer tutoring intervention.

- "I feel more comfortable sharing results with my peer during Mathematics in English classes" (FG2_7).

They also think that sharing doubts with a peer generates less anxiety than asking the teacher:

- "I used to feel more nervous when we didn't work in pairs. If I didn't understand something in English I couldn't do the exercises most of the times. It's good to have someone that can help you or that you can ask something" (FG1_7).
- "When the teacher said that we had to help our classmates in English I thought I wasn't going to be able to do it, but now I think that I can and my mate will understand me" (FG8_9).

RQ1 was put forward to verify whether participants' motivation towards CLIL would increase after peer-tutoring. Both quantitative and qualitative data indicate that there are indeed significant differences between the control and the experimental group. A Hedge's g effect size of 0.26 for intrinsic motivation and 0.24 for motivational force show that the inter-group differences are significant but moderate ([Cheung & Slavin, 2016](#)). These outcomes, then, are consonant with those that support the adoption of classroom methodologies enhancing active learning in CLIL ([Banegas, 2012, 2013](#); [Coyle, 2013](#); [Czura & Anklewicz, 2018](#); [Huang, 2011](#); [Lasagabaster & Doiz, 2016](#); [Pistorio, 2010](#); [Sánchez-García & Pavón-Vázquez, 2021](#); [Sierra, 2011, 2016](#)). Furthermore, in comparison with the findings of studies where the impact of methodologies like project work ([Lasagabaster & López Beloqui, 2015](#)) or peer-tutoring ([Lialikhova, 2019](#)) did not prove to be as positive as expected—especially for lower achievers—, our results appear to substantiate the benefits of same age reciprocal peer-tutoring in driving CLIL motivation. Whether this collaborative methodology is more effective in the early grades, as suggested by [Lasagabaster and Doiz \(2016\)](#), is an issue we will discuss next.

5.2. RQ2: the age factor

Moving to the relationship between age and motivation towards CLIL Mathematics after the peer-tutoring intervention (RQ2), results pointed to statistical significant improvements in intrinsic motivation and motivational force only for the younger participants in the experimental group (see tests 4 to 7 in [Tables 3 and 4](#)). The ANOVA reported statistical significant differences among 7, 8 and 9 grade scores in both intrinsic motivation 3.45 ($p = .04$) and motivational force towards CLIL 3.65 ($p = .03$). Moreover, a Scheffe test showed that these differences existed between the first year students *versus* the second and the third year ones, with 7 graders outscoring their elder peers. Hence, our second hypothesis was disconfirmed.

The fact that 7 graders in the experimental group were more motivated towards CLIL Mathematics after the peer-tutoring experience contradicts those scholars who hold that with CLIL, the later, the better ([Lorenzo et al., 2010](#); [Moate, 2014](#); [Otwinowska & Forys, 2017](#); or, [Zydatiþ, 2012](#)). Even though the first year in CSE may be demanding for pupils coming from primary education, or maybe for that very reason, the experience of peer-tutoring appears to be more beneficial for those students who elsewhere have described their first encounters with CLIL Mathematics as "very intimidating" ([Czura & Anklewicz, 2018, p. 55](#)). Our outcomes in that respect seem to align with prior research suggesting that the later students embrace CLIL, the more difficult it will be for them to accept

its benefits and challenges (Nikula, 2017; Pérez et al., 2016; Siqueira et al., 2018). They also tally with longitudinal studies such as Lasagabaster and Doiz (2016) underscoring that the early grades' preference for cooperative methodologies involving group work and active participating declines over time.

5.3. RQ3: males versus females

With regard to our third research question (i.e. How does gender affect motivation towards CLIL Mathematics in peer-tutoring?), statistical significant differences between males and females were reported in the results from student's t-tests displayed in Table 5. In terms of intrinsic motivation, male subjects in the experimental group outscored both females in the experimental group and their male peers in the control group. This fact is reported when analyzing both, the posttest and gains differences between males and females in the experimental group. Linear regression analysis resulted in a R-squared value of .59 including experimental condition and gender as dependents variables ($p < .01$). Readers must bear in mind that comparisons in this subsection are limited to the gender variable. That is why no additional information about grades is included below.

Table 6 presents the results of student's t-tests obtained in relation to motivational force towards CLIL Mathematics. Statistical significant differences between males and females are also found here. Yet, it is female learners in the experimental group those who outscored both male subjects in the experimental group and their female peers in the control group. Linear regression analysis resulted in a R-squared value of .57 including experimental condition and gender as dependents variables ($p < .05$).

To put it simply, then, it seems that the peer-tutoring experience boosts males' intrinsic motivation and reduces females' anxiety. This counterbalances the tendency noted in the literature on foreign language learning where boys' positive attitudes and classroom enjoyment are very low (Carr & Pauwels, 2006; Jones, 2009; Kissau, 2007; Kissau & Wierzalis, 2008; Kobayashi, 2002), in contrast with research on Mathematics, a highly masculinized discipline where girls are shown to get more anxious and feel less confident (Dowker, Sarkar, & Looi, 2016; He et al., 2019; Holm, Hannula, & Björn, 2017; Leyva, 2017; Passolunghi, Caviola, De Agostini, Perin, & Mammarella, 2016; Stoet, Bailey, Moore, & Geary, 2016).

Hence, our third hypothesis, predicting a transfer in the motivational profiles of males and females, appears to be confirmed. Such transfer mirrors previous peer-tutoring studies in which females are reported to achieve higher benefits in confidence or self-concept thanks to this methodology (Knight, Dipper, & Cruice, 2018; Rickard & Mills, 2018), whilst their male peers may experience greater gains in positive attitude and intrinsic motivation towards the content subject (Ayvazo & Aljadef-Abergel, 2019; Healy, Block, & Kelly, 2019).

6. Concluding remarks

This study adds to the scant literature on the effects of collaborative learning methodologies like peer-tutoring on CLIL learner motivation. As opposed to previous research, the novelty of our contribution lies in (i) having addressed motivation towards CLIL as a whole (Somers & Llinares, 2021) and in (ii) having paid attention to the mediating effects of individual variables like age and gender.

On the whole, our results reveal that the implementation of peer-tutoring in CLIL Mathematics reduces anxiety, builds up learners' self-confidence and encourages students' positive attitudes. Significant but moderate improvements have occurred in both students' intrinsic motivation (i.e. positive attitudes and classroom enjoyment) and motivational force (i.e. confidence and anxiety issues) towards CLIL. A lesson to be drawn from the higher motivation displayed by the younger participants (7 graders) is that early implementation of CLIL programs should not be feared. Besides, a change in previous gendered motivational profiles took place with females experiencing higher gains in motivational force and males benefitting from an attitudinal perspective.

Yet, the current study is a preliminary effort subject to several limitations that should be noted. The first limitation has to do with the sample size, which is constrained to a single secondary school. Although randomized control sampling is rarely available in educational research and the normal distribution of the obtained data was tested, the fact that the sample was obtained by means of cluster must be acknowledged as a limitation (Haq, 2017). The second limitation results from the unbalanced distribution of participants in terms of gender with 63.48% females versus 36.52% males. Third, ceiling effects (Resch & Isenberg, 2018) must be considered as 26 out of 204 students in the experimental group (12.74%) already reached the maximum score in the pretest for one of the two analyzed factors and 4 of them (1.96%) also achieved the maximum score in the two analyzed factors in the pretest. Moreover, in the

Table 5
Student's t-tests by gender for intrinsic motivation.

Test	Group A	Group B	t (sig.)
1	Male experimental group pretest	Male control group pretest	1.04 (p = .30)
2	Female experimental group pretest	Female control group pretest	1.88 (p = .06)
3	Male experimental group posttest	Male experimental group pretest	2.74 (p < .01)*
4	Female experimental group posttest	Male experimental group pretest	0.24 (p = .81)
5	Male control group posttest	Male control group pretest	0.06 (p = .95)
6	Female control group posttest	Female control group pretest	-1.64 (p = .10)
7	Male experimental group posttest	Female experimental group posttest	2.33 (p = .02)*
8	Male experimental group gains	Female experimental group gains	2.64 (p < .01)*

Table 6
Student's *t*-tests by gender for motivational force.

Test	Group A	Group B	<i>t</i> (sig.)
1	Male experimental group pretest	Male control group pretest	0.44 (p = .81)
2	Female experimental group pretest	Female control group pretest	1.84 (p = .07)
3	Male experimental group posttest	Male experimental group pretest	1.93(p = .06)
4	Female experimental group posttest	Male experimental group pretest	2.22 (p = .03)*
5	Male control group posttest	Male control group pretest	0.06 (p = .95)
6	Female control group posttest	Female control group pretest	-1.64 (p = .10)
7	Male experimental group posttest	Female experimental group posttest	-2.22 (p = .03)*
8	Male experimental group gains	Female experimental group gains	2.05 (p < .05)*

regression analysis, although the majority of the variance in the outcome variable (gains) was explained by the regression models, a strong positive linear relationship (R-squared values greater than 0.6) via a firm linear rule (Gelman, Goodrich, Gabry, & Vehtari, 2019) was not reported for any of them. Finally, increasing motivation was measured according to self-reported values without comparing learners' perceived gains with their actual class performance and/or academic achievement.

To address some of these shortcomings, future research might benefit from a wider sample size, from a more symmetrical distribution of participants in terms of gender or from data triangulation. Clearly, more work is also needed to tease out the relationship between CLIL related pedagogies like peer-tutoring and type of content matter. Ultimately, in comparison with more humanistic subjects, Mathematics forms part of those scientific disciplines whose technical language may facilitate the implementation of CLIL (Jäppinen, 2005; Prochazkova, 2013; Salvador-García, Chiva-Bartoll, & Capella-Peris, 2019). Last but not least, a placement test should have been implemented in order to determine participants' entry level in General English. Even though we concur with Zydati (2012) that the main linguistic problems students face in CLIL lessons usually derive from their limited proficiency in academic language or CALP, having more information about participants' competence in conversational English or BICS would have been highly desirable. In conclusion, despite all the aforementioned limitations, we hope that this paper might be seen as a step forward in the pedagogical concretion of those CLIL classroom practices that may guarantee the quality of this teaching approach. A pedagogical concretion mostly needed when CLIL is for all.

Funding details

This research was not supported by any funding

Data availability statement

Authors make the data underlying this study open. Data set associated with the paper can be accessed at https://osf.io/qjnrs/?view_only=646b8e71f8214f81a4ea1741e0a1c9a0.

Author statement

Otilia Martí: Visualization, Writing- Reviewing and Editing. Lidon Moliner: Conceptualization, Methodology, Supervision. Francisco Alegre: Software, Data curation, Writing- Original draft preparation.

Declaration of competing interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

Funding for open access charge: CRUE-Universitat Jaume I.

References

- Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin, G. (2019a). Peer tutoring in algebra: A study in middle school. *The Journal of Educational Research*, 112(6), 693–699. <https://doi.org/10.1080/00220671.2019.1693947>
- Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin, G. (2019b). Peer tutoring in mathematics in primary education: A systematic review. *Educational Review*, 71(6), 767–791. <https://doi.org/10.1080/00131911.2018.1474176>
- Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin, G. (2019c). Peer tutoring and mathematics in secondary education: Literature review, effect sizes, moderators, and implications for practice. *Heliyon*, 5(9), Article e02491. <https://doi.org/10.1016/j.heliyon.2019.e02491>
- Alejo, R., & Piquer-Piriz, A. (2016). Urban vs. rural CLIL: An analysis of input-related variables, motivation and Language attainment. *Language Culture and Curriculum*, 29(3), 245–262. <https://doi.org/10.1080/07908318.2016.1154068>
- Anwaruddin, S. M. (2016). Language teachers' responses to educational research: Addressing the 'crisis' of representation. *International Journal of Research and Method in Education*, 39(3), 314–328. <https://doi.org/10.1080/1743727X.2016.1166485>

- Apfel, C. (2012). Dropouts from CLIL streams in Germany. *International CLIL Research Journal*, 1(4), 47–56. Retrieved from <http://www.icrij.eu/14/article5.html>, 30th March 2020.
- Arribas, M. (2016). Analysing a whole CLIL school: Students' attitudes, motivation, and receptive vocabulary outcomes. *Latin American Journal of Content and Language Integrated Learning*, 9(2), 267–292. <https://doi.org/10.5294/lacil.2016.9.2.2>
- Ayvazo, S., & Aljadeff-Abergel, E. (2019). Classwide peer tutoring in a martial arts alternative education program: Enhancing social and psychomotor skills. *Preventing School Failure: Alternative Education for Children and Youth*, 63(4), 359–368. <https://doi.org/10.1080/1045988X.2019.1619509>
- Baetens-Beardmore, H. (2001). Foreword: The past decade and the next millennium. In D. Marsh, A. Maljers, & A. Hartiala (Eds.), *Profiling European CLIL classrooms* (pp. 10–11). UNICOM: University of Jyväskylä & European Platform for Dutch Education.
- Ballester-Roca, J., & Spaliviero, C. (2021). CLIL and literary education: Teaching foreign languages and literature from an intercultural perspective: The results of a case study. In M. L. Carrió-Pastor, & B. Bellés Fortuño (Eds.), *Teaching language and content in multicultural and multilingual classrooms: CLIL and EMI approaches* (pp. 225–251). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-56615-9_9.
- Banegas, D. L. (2012). Integrating content and language in English language teaching in secondary education: Models, benefits and challenges. *Studies in Second Language Learning and Teaching*, 2(1), 111–136.
- Banegas, D. L. (2013). The integration of content and language as a driving force in the EFL lesson. In E. Ushioda (Ed.), *International perspectives on motivation: Language learning and professional challenges* (pp. 82–97). Basingstoke: Palgrave Macmillan.
- Banerjee, B., & Pradhan, B. (2018). Kolmogorov–Smirnov test for life test data with hybrid censoring. *Communications in Statistics - Theory and Methods*, 47(11), 2590–2604. <https://doi.org/10.1080/03610926.2016.1205616>
- Bauer-Marschallinger, S., Dalton-Puffer, C., Heaney, H., Katzinger, L., & Smit, U. (2021). CLIL for all? An exploratory study of reported pedagogical practices in Austrian secondary schools. *International Journal of Bilingual Education and Bilingualism*. <https://doi.org/10.1080/13670050.2021.1996533>
- de Bot, K. (2002). Relevance of CLIL to the European commission's language learning objectives. In D. Marsh (Ed.), *CLIL/EMILE-The European dimension: Actions, trends and foresight potential*. Jyväskylä: University of Jyväskylä.
- Broca, A. (2016). CLIL and non-CLIL: Differences from the outset. *ELT Journal*, 70, 320–331. <https://doi.org/10.1093/elt/ccw011>
- Bruno, A. (2011a). Are the differences between CLIL and non-CLIL groups in andalusia due to CLIL? A reply to Lorenzo, Casal and Moore. *Applied Linguistics*, 32(2), 236–241. <https://doi.org/10.1093/applin/amr007>, 2010.
- Bruno, A. (2011b). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, 39(4), 523–532. <https://doi.org/10.1016/j.system.2011.08.002>
- Bruno, A. (2013). CLIL: Some of the reasons why...and why not. *System*, 41(3), 587–597. <https://doi.org/10.1016/j.system.2013.07.001>
- Bryman, A. (2017). Quantitative and qualitative research: Further reflections on their integration. In *Mixing methods: Qualitative and quantitative research* (pp. 57–78). Routledge. <https://doi.org/10.4324/9781315248813>.
- Busse, V. (2017). Plurilingualism in Europe: Exploring attitudes toward English and other European languages among adolescents in Bulgaria, Germany, The Netherlands, and Spain. *The Modern Language Journal*, 101(3), 566–582. <https://doi.org/10.1111/modl.12415>
- Cabrera, L. Y., & Reiner, P. B. (2018). A novel sequential mixed-method technique for contrastive analysis of unscripted qualitative data: Contrastive quantized content analysis. *Sociological Methods & Research*, 47(3), 532–548. <https://doi.org/10.1177/0049124116661575>
- Cáceres, A., & Unigarro, D. (2007). Analysis of children's English language evidences when engaging in project work. *Profil*, 8, 9–23.
- Campillo, J. M., Sánchez, R., & Miralles, P. (2019). Primary teachers' perceptions of CLIL implementation in Spain. *English Language Teaching*, 12(4), 149–156. <https://doi.org/10.5539/elt.v12n4p149>
- Carr, J., & Pauwels, A. (2006). *Boys and foreign language learning: Real boys don't do languages*. Houndmills, UK: Palgrave Macmillan.
- Castells, & Mestre-Mestre, E. M. (2015). Latest language policy proposals in education in the Valencian Country. *Procedia-Social and Behavioral Sciences*, 178(10), 151–156. <https://doi.org/10.1016/j.sbspro.2015.03.172>
- Celik, H. C. (2017). Mathematical modelling research in Turkey: A content analysis study. *Educational Research and Reviews*, 12(1), 19–27. <https://doi.org/10.5897/ERR2016.3077>
- Cenoz, J., Genesee, F., & Gorter, D. (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics*, 35(3), 356–369. <https://doi.org/10.1093/applin/amt011>
- Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292. <https://doi.org/10.3102/0013189X16656615>
- Clark, V. L. P. (2019). Meaningful integration within mixed methods studies: Identifying why, what, when, and how. *Contemporary Educational Psychology*, 57, 106–111. <https://doi.org/10.1016/j.cedpsych.2019.01.007>
- Coonan, C. M. (2007). Insider views of the CLIL class through teacher self-observation-introspection. *International Journal of Bilingual Education and Bilingualism*, 10(5), 625–646. <https://doi.org/10.2167/beb463.0>
- Coonan, C. M. (2012). Affect and motivation in CLIL. In D. Marsh, & O. Meyer (Eds.), *Quality interfaces: Examining evidence and exploring solutions in CLIL* (pp. 53–66). Eichstaett: Eichstaett Academic Press.
- Coyle, D. (2006). Content and language integrated learning: Motivating learners and teachers. *The Scottish Language Review*, 13, 1–18.
- Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism*, 10(5), 543–562. <https://doi.org/10.2167/beb459.0>
- Coyle, D. (2008). CLIL: A pedagogical approach from the European perspective. In N. Van Dusen-Scholl, & N. H. Hornberger (Eds.) (2 ed., vol. 4. *Encyclopedia of language and education*) (pp. 97–112). Berlin: Springer. <https://doi.org/10.1007/978-0-387-30424-3>.
- Coyle, D. (2013). Listening to learners: An investigation into 'successful learning' across CLIL contexts. *International Journal of Bilingual Education and Bilingualism*, 16(3), 244–266. <https://doi.org/10.1080/13670050.2013.777384>
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge: Cambridge University Press.
- Csizér, K., & Dörnyei, Z. (2005). The internal structure of language learning motivation and its relationship with language choice and learning effort. *The Modern Language Journal*, 89(1), 19–36. <https://doi.org/10.1111/j.0026-7902.2005.00263.x>
- Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, 19, 121–129.
- Czura, A., & Anklewicz, A. (2018). Pupils' and teachers' perceptions of CLIL in primary school: A case study. *Linguodidactica*, 22, 47–63. <https://doi.org/10.15290/lingdid.2018.22.03>
- Dalton-Puffer, C. (2011). Content-and-Language Integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, 31, 182–204. <https://doi.org/10.1017/S0267190511000092>
- Dalton-Puffer, C., & Smit, U. (2013). Content and language integrated learning: A research agenda. *Language Teaching*, 46(4), 545–559. <https://doi.org/10.1017/S0261444813000256>
- Daniel, B. K. (2019). Big data and data science: A critical review of issues for educational research. *British Journal of Educational Technology*, 50(1), 101–113. <https://doi.org/10.1111/bjet.12595>
- Darn, S. (2006). Content and language integrated learning (CLIL): A European overview. *INGED News Online 2*. Retrieved from <https://eric.ed.gov/?id=ED490775>, 30th March 2020.
- De Backer, L., Van Keer, H., & Valcke, M. (2015). Exploring evolutions in reciprocal peer tutoring groups' socially shared metacognitive regulation and identifying its metacognitive correlates. *Learning and Instruction*, 38(1), 63–78. <https://doi.org/10.1016/j.learninstruc.2015.04.001>
- Denman, J., Tanner, R., & Graaff, R. de (2013). CLIL in junior vocational secondary education: Challenges and opportunities for teaching and learning. *International Journal of Bilingual Education and Bilingualism*, 16(3), 285–300. <https://doi.org/10.1080/13670050.2013.777386>
- Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014a). CLIL and motivation: The effect of individual and contextual variables. *Language Learning Journal*, 42(2), 209–224. <https://doi.org/10.1080/09571736.2014.889508>

- Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014b). Giving voice to the students: What (de)motivates them in CLIL classes? In D. Lasagabaster, A. Doiz, & J. M. Sierra (Eds.), *Motivation and foreign language learning: From theory to practice* (pp. 117–138). Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.40>
- Dörnyei, Z. (2000). Motivation in action: Toward a process-oriented conceptualization of student motivation. *British Journal of Educational Psychology*, 70(4), 519–538. <https://doi.org/10.1348/000709900158281>
- Dörnyei, Z. (2005). *The psychology of the language learner*. Mahway, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2009). The L2 motivational self system. In Z. Dörnyei, & E. Ushioda (Eds.), *Motivation, language identity and the L2 self* (pp. 92–142). Bristol: Multilingual Matters.
- Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, 7, 508. <https://doi.org/10.3389/fpsyg.2016.00508>
- Duran, D., Corcelles, M., Flores, M., & Miquel, E. (2019). Changes in attitudes and willingness to use co-teaching through pre-service teacher training experiences. *Professional Development in Education*, 1–10. <https://doi.org/10.1080/19415257.2019.1634631>
- Duran, D., Flores, M., Oller, M., & Ramírez, M. (2019). Reading in Pairs, description and results of a peer tutoring program for English as a foreign language. *Innovation in Language Learning and Teaching*, 13(4), 303–317. <https://doi.org/10.1080/17501229.2018.1462370>
- Escobar Urmeneta, C., & Evnitskaya, N. (2013). Affording students opportunities for the integrated learning of content and language. A contrastive study on classroom interactional strategies deployed by two CLIL teachers. In J. Arnau (Ed.), *Reviving Catalan at school: Challenges and instructional approaches* (pp. 158–182). Bristol: Multilingual Matters.
- Etikan, I., & Bala, K. (2017). Sampling and sampling methods. *Biometrics & Biostatistics International Journal*, 5(6), Article 00149. <https://doi.org/10.15406/bbij.2017.05.00149>
- European Commission. (2012). Content and language integrated learning. *European Commission for Languages*. Retrieved from http://ec.europa.eu/languages/language-teaching/content-and-language-integrated-learning_en.htm, 30th March 2020.
- Fernández Barrionuevo, E., & Baena Extremera, A. (2018). Motivation in physical education and foreign language learning in CLIL teaching: Gender differences and implications for future studies. *Porta Linguarum*, 30, 207–220.
- Fernández Fontecha, A., & Canga Alonso, A. (2014). A preliminary study on motivation and gender in CLIL and non-CLIL types of instruction. *International Journal of English Studies*, 14(1), 21–36. <https://doi.org/10.6018/ijes/14/1/156681>
- Gallardo-del-Puerto, F., & Blanco-Suárez, Z. (2021). Foreign language motivation in primary education students: The effects of additional CLIL and gender. *Journal of Immersion and Content-Based Language Education*, 9(1), 58–84. <https://doi.org/10.1075/jicb.19023.gal>
- Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for Bayesian regression models. *The American Statistician*. <https://doi.org/10.1080/00031305.2018.1549100>
- Genesee, F., & Hamayan, E. (2016). *CLIL in context. Practical guidance for educators*. Cambridge: Cambridge University Press.
- Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, 61(2), 29–45. <https://doi.org/10.1027/1901-2276.61.2.29>
- Guetterman, T. C., Babchuk, W. A., Howell Smith, M. C., & Stevens, J. (2019). Contemporary approaches to mixed methods–grounded theory research: A field-based analysis. *Journal of Mixed Methods Research*, 13(2), 179–195. <https://doi.org/10.1177/1558689817710877>
- Guillamón-Suesta, F., & Renau, M. L. (2015). A critical vision of the CLIL approach in secondary education: A study in the Valencian community in Spain. *Latin American Journal of Content & Language Integrated Learning*, 8(1), 1–12. <https://doi.org/10.5294/5037>
- Halbach, A. (2009). The primary school teacher and the challenges of bilingual education. In E. Dafouz (Ed.), *CLIL across educational levels: Experiences from primary, secondary and tertiary contexts* (pp. 19–26). London/Madrid: Richmond Santillana.
- Haq, A. (2017). Two-stage cluster sampling with hybrid ranked set sampling in the secondary sampling frame. *Communications in Statistics - Theory and Methods*, 46(17), 8450–8467. <https://doi.org/10.1080/03610926.2016.1183783>
- Healy, S., Block, M., & Kelly, L. (2019). The impact of online professional development on physical educators' knowledge and implementation of peer tutoring. *International Journal of Disability, Development and Education*, 1–13. <https://doi.org/10.1080/1034912X.2019.1599099>
- Heras, A., & Lasagabaster, D. (2015). The impact of CLIL on affective factors and vocabulary learning. *Language Teaching Research*, 19(1), 70–88. <https://doi.org/10.1177/1362168814541736>
- He, L., Zhou, G., Salinitri, G., & Xu, L. (2019). Female underrepresentation in STEM subjects: An exploratory study of female high school students in China. *Eurasia Journal of Mathematics, Science and Technology Education*, 16(1). <https://doi.org/10.29333/ejmste/109657>
- Holm, M. E., Hannula, M. S., & Björn, P. M. (2017). Mathematics-related emotions among Finnish adolescents across different performance levels. *Educational Psychology*, 37(2), 205–218. <https://doi.org/10.1080/01443410.2016.1152354>
- Huang, K. (2011). Motivating lessons: A classroom-oriented investigation of the effects of content-based instruction on EFL young learners' motivated behaviours and classroom verbal interaction. *System*, 39(2), 186–201. <https://doi.org/10.1016/j.system.2011.02.002>
- Hunt, M. (2011). Learners' perceptions of their experiences of learning subject content through a foreign language. *Educational Review*, 63(3), 365–378. <https://doi.org/10.1080/00131911.2011.571765>
- Hüttner, J., Dalton-Puffer, C., & Smit, U. (2013). The power of beliefs: Lay theories and their influence on the implementation of CLIL programmes. *International Journal of Bilingual Education and Bilingualism*, 16(3), 267–284. <https://doi.org/10.1080/13670050.2013.777385>
- Hwang, W. Y., Nguyen, T. H., & Pham, X. L. (2019). Peer tutoring to facilitate cognitive diffusion of English as a foreign language learning: Using speech translation and shadowing in familiar authentic contexts. *Journal of Educational Computing Research*, Article 0735633118776209. <https://doi.org/10.1177/0735633118776209>
- Jäppinen, A. K. (2005). Thinking and content learning of mathematics and science as cognitional development in content and language integrated learning (CLIL): Teaching through a foreign language in Finland. *Language and Education*, 19(2), 147–168. <https://doi.org/10.1080/09500780508668671>
- Jones, C. (2009). Parental support and the attitudes of boys and girls to modern foreign languages. *Language Learning Journal*, 37(1), 85–97. <https://doi.org/10.1080/09571730902717349>
- Kaplan, A., & Garner, J. K. (2020). Steps for applying the complex dynamical systems approach in educational research: A guide for the perplexed scholar. *The Journal of Experimental Education*, 88(3), 486–502. <https://doi.org/10.1080/00220973.2020.1745738>
- Kissau, S. (2007). Is what's good for the goose good for the gander? The case of male and female encouragement to study French. *Foreign Language Annals*, 40(3), 419–432. <https://doi.org/10.1111/j.1944-9720.2007.tb02867.x>
- Kissau, S., & Wierzaliz, E. (2008). Gender identity and homophobia: The impact on adolescent males studying French. *The Modern Language Journal*, 92(3), 402–413. <https://doi.org/10.1111/j.1540-4781.2008.00754.x>
- Knight, R. A., Dipper, L., & Cruice, M. (2018). Viva survivors—the effect of peer-mentoring on pre-viva anxiety in early-years students. *Studies in Higher Education*, 43(1), 190–199. <https://doi.org/10.1080/03075079.2016.1161018>
- Kobayashi, Y. (2002). The role of gender in Foreign Language learning attitudes: Japanese female students' attitudes towards English learning. *Gender and Education*, 14(2), 181–197. <https://doi.org/10.1080/09540250220133021>
- Lasagabaster, D. (2008). foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, 1(1), 30–41. <https://doi.org/10.2174/1874913500801010030>
- Lasagabaster, D. (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in Language Learning and Teaching*, 5(1), 3–18. <https://doi.org/10.1080/17501229.2010.519030>
- Lasagabaster, D. (2016). The relationship between motivation, gender, L1 and possible selves in English-medium instruction. *International Journal of Multilingualism*, 13(3), 315–332. <https://doi.org/10.1080/14790718.2015.1105806>
- Lasagabaster, D. (2019). Motivation in content and language integrated learning (CLIL) research. In T. Lamb, K. Csizer, A. Henry, & S. Ryan (Eds.), *The palgrave handbook of motivation for language learning* (pp. 347–366). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28380-3_17

- Lasagabaster, D., & Doiz, A. (2016). CLIL students' perceptions of their language learning process: Delving into self-perceived improvement and instructional preferences. *Language Awareness*, 25(1–2), 110–126. <https://doi.org/10.1080/09658416.2015.1122019>
- Lasagabaster, D., & Doiz, A. (2020). Epilogue: Multilingualism in northern European universities—reflections from the south. In *Language perceptions and practices in multilingual universities* (pp. 379–394). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-38755-6_15.
- Lasagabaster, D., & López Belouqui, R. (2015). The impact of type of approach (CLIL versus EFL) and methodology (Book-Based versus project work) on motivation. *Porta Linguarum*, 23, 41–57. <https://doi.org/10.30827/Digibug.53737>
- Lasagabaster, D., & Ruiz de Zarobe, Y. (Eds.). (2010). *CLIL in Spain: Implementation, results and teacher training*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Lasagabaster, D., & Sierra, J. M. (2009). Language attitudes in CLIL and traditional EFL classes. *International CLIL Research Journal*, 1(2), 4–17. Retrieved from <http://www.icrj.eu/index.php?vol=12&page=73>, 30th March, 2020.
- Leung, K. C. (2015). Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *Journal of Educational Psychology*, 107(2), 558–579. <https://doi.org/10.1037/a0037698>
- Leung, K. C. (2019). Compare the moderator for pre-test-post-test design in peer tutoring with treatment-control/comparison design. *European Journal of Psychology of Education*, 34(4), 685–703. <https://doi.org/10.1007/s10212-018-00412-6>
- Leyva, L. A. (2017). Unpacking the male superiority myth and masculinization of mathematics at the intersections: A review of research on gender and mathematics education. *Journal for Research in Mathematics Education*, 48(4), 397–433. <https://doi.org/10.5951/jresmetheduc.48.4.0397>
- Lialikhova, D. (2019). We can do it together!—but can they? How Norwegian ninth graders co-constructed content and language knowledge through peer interaction in CLIL. *Linguistics and Education*, 54, Article 100764. <https://doi.org/10.1016/j.linged.2019.100764>
- Lim Falk, M. (2008). *Svenska i engelskspråkig skolmiljö: Ämnesrelaterat språkbruk i två gymnasieklasser*. Stockholm University. Unpublished doctoral dissertation.
- Li, D., & Zhang, L. (2020). Exploring teacher scaffolding in a CLIL-framed EFL intensive reading class: A classroom discourse analysis approach. *Language Teaching Research*, Article 1362168820903340. <https://doi.org/10.1177/1362168820903340>
- López Rúa, P. (2006). The sex variable in Foreign Language learning: An integrative approach. *Porta Linguarum*, 6, 99–114.
- Lorenzo, F., Casal, S., & Moore, P. (2010). The effects of content and language integrated learning in European education: Key findings from the Andalusian bilingual sections evaluation project. *Applied Linguistics*, 31, 418–442. <https://doi.org/10.1093/applin/amp041>
- Madrid, D., & Pérez-Cañado, M. L. (2018). Innovations and challenges in attending to diversity through CLIL. *Theory Into Practice*, 57(3), 241–249. <https://doi.org/10.1080/00405841.2018.1492237>
- Martínez Agudo, J. (2021). Do CLIL programmes help to balance out gender differences in content and language achievement? *Language Culture and Curriculum*. <https://doi.org/10.1080/07908318.2021.1942033>
- Massler, U. (2012). Primary CLIL and its stakeholders: What children, parents and teachers think of the potential merits and pitfalls of CLIL modules in primary teaching. *International CLIL Research Journal*, 1(4), 36–46. Retrieved from <http://www.icrj.eu/14/article4.html>, 30th March 2020.
- McCall, I. (2012). Score in French: Motivating boys with football in key stage 3. *Language Learning Journal*, 39(1), 5–18. <https://doi.org/10.1080/09571731003692928>
- Mearns, T., de Graaf, R., & Coyle, D. (2017). Motivation for or from bilingual education? A comparative study of learner views in The Netherlands. *International Journal of Bilingual Education and Bilingualism*. <https://doi.org/10.1080/13670050.2017.1405906>
- Mearns, T., & de Graaf, R. (2018). Bucking the trend? Motivational differences between boys and girls who opt in or out of bilingual education. *Journal of Immersion and Content-Based Language Education*, 6(1), 1–26. <https://doi.org/10.1075/jicb.17003.mea>
- Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL: Content and language integrated learning in bilingual and multilingual education*. Oxford: Macmillan.
- Merisuo-Storm, T. (2007). Pupils' attitudes towards foreign-language learning and the development of literacy skills in bilingual education. *Teaching and Teacher Education*, 23, 226–235. <https://doi.org/10.1016/j.tate.2006.04.024>
- Meyer, O. (2010). Towards quality-CLIL: Successful planning and teaching strategies. *Pulso: Revista de Educación*, 33, 11–29. Retrieved from <http://hdl.handle.net/10017/7204>, 30th March, 2020.
- Meyer, O., Coyle, D., Imhof, M., & Connolly, T. (2018). Beyond CLIL: Fostering student and teacher engagement for personal growth and deeper learning. In J. de D. Martínez Agudo (Ed.), *Emotions in second language teaching*. Cham: Springer. https://doi.org/10.1007/978-3-319-75438-3_16.
- Moate, J. (2014). A narrative account of a teacher community. *Teacher Development: An International Journal of Teachers' professional development*, 18, 384–402. <https://doi.org/10.1080/13664530.2014.931294>
- Moliner, L., & Alegre, F. (2020). Attitudes, beliefs and knowledge of mathematics teachers regarding peer tutoring. *European Journal of Teacher Education*, 1–20. <https://doi.org/10.1080/02619768.2020.1803271>
- Nikula, T. (2017). CLIL: A European approach to bilingual education. In N. Van Deusen-Scholl, & S. May (Eds.), *Second and foreign language education. Encyclopedia of language and education* (3rd ed.). Cham: Springer. https://doi.org/10.1007/978-3-319-02246-8_10
- Nikula, T., & Moore, P. (2019). Exploring translanguaging in CLIL. *International Journal of Bilingual Education and Bilingualism*, 22(2), 237–249. <https://doi.org/10.1080/13670050.2016.1254151>
- Norton, B., & Pavlenko, A. (2004). *Gender and English language learners*. Alexandria, VA: TESOL Inc.
- Otto, A., & San Isidro, X. (2019). Language as the articulator of a CLIL ecosystem: The Spanish case. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 27, 14–31. <https://doi.org/10.26378/rnlael1327338>
- Otwinowska, A. (2013). CLIL lessons in the upper-primary: The interplay of affective factors and CALP. In D. Gabrys-Barker, & J. Bielska (Eds.), *The affective dimension in second language acquisition* (pp. 211–225). Bristol: Multilingual Matters.
- Otwinowska, A., & Forys, M. (2017). They learn the CLIL way, but do they like it? Affectivity and cognition in upper-primary CLIL classes. *International Journal of Bilingual Education and Bilingualism*, 20(5), 457–480. <https://doi.org/10.1080/13670050.2015.1051944>
- Passolunghi, M. C., Caviola, S., De Agostini, R., Perin, C., & Mammarella, I. C. (2016). Mathematics anxiety, working memory, and Mathematics performance in secondary-school children. *Frontiers in Psychology*, 7, 42. <https://doi.org/10.3389/fpsyg.2016.00042>
- Pastrana, A. (2019). Co-construction of knowledge in Primary CLIL group work activities. In K. Tsuchiya, & M. D. Pérez Murillo (Eds.), *Content and language integrated learning in Spanish and Japanese contexts* (pp. 205–236). Cham: Palgrave Macmillan: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-27443-6_9.
- Pavlenko, A., & Piller, I. (2008). Language education and gender. In S. May, & N. H. Hornberger (Eds.), (2nd ed., vol. 1. *Encyclopedia of language and education* (pp. 57–69). New York: Springer Science+Business Media LLC.
- Pavón Vázquez, V., & Ellison, M. (2013). Examining teacher roles and competences in content and language integrated learning (CLIL). *Linguarum Arena*, 4, 65–78.
- Pérez Cañado, M. L. (2016). From the CLIL craze to the CLIL conundrum: Addressing the current CLIL controversy. *Bellaterra Journal of Teaching & Learning Language & Literature*, 9(1), 9–31. <https://doi.org/10.5565/rev/jtl3.667>
- Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present and future. *International Journal of Bilingual Education and Bilingualism*, 15(3), 315–341. <https://doi.org/10.1080/13670050.2011.630064>
- Pérez, A., Lorenzo, F., & Pavón, V. (2016). European bilingual models beyond linguafranca: Key findings from CLIL French programs. *Language Policy*, 15(4), 485–504. <https://doi.org/10.1007/s10993-015-9386-7>
- Pistorio, M. I. (2010). A blend of CLIL and cooperative learning creates a socially constructed learning environment. *Latin American Journal of Content & Language Integrated Learning*, 3(1), 1–10. <https://doi.org/10.5294/laclil.2010.3.1.1>
- Pladevall-Ballester, E. (2015). Exploring primary school CLIL perceptions in Catalonia: Students', teachers' and parents' opinions and expectations. *International Journal of Bilingual Education and Bilingualism*, 18(1), 45–59. <https://doi.org/10.1080/13670050.2013.874972>
- del Pozo, E. (2019). CLIL in secondary classrooms: History contents on the move. In K. Tsuchiya, & M. D. Pérez Murillo (Eds.), *Content and language integrated learning in Spanish and Japanese contexts: Policy, practice and pedagogy* (pp. 125–151). Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-030-27443-6_6.
- Prochazkova, L. T. (2013). Mathematics for language, language for mathematics. *European Journal of Science and Mathematics Education*, 1(1), 23–28. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1108199.pdf>, 30th March, 2020.

- Resch, A., & Isenberg, E. (2018). How do test scores at the ceiling affect value-added estimates? *Statistics and Public Policy*, 5(1), 1–6. <https://doi.org/10.1080/2330443X.2018.1460226>
- Rickard, B., & Mills, M. (2018). The effect of attending tutoring on course grades in Calculus I. *International Journal of Mathematical Education in Science & Technology*, 49(3), 341–354. <https://doi.org/10.1080/0020739X.2017.1367043>
- Roquet, H., Llopis, J., & Pérez-Vidal, C. (2016). Does gender have an impact on the potential benefits learners may achieve in two contexts compared: Formal instruction and formal instruction + content and Language integrated learning? *International Journal of Bilingual Education and Bilingualism*, 19(4), 370–386. <https://doi.org/10.1080/13670050.2014.992389>
- Ruiz de Zarobe, Y. (2015). The effects of implementing CLIL in education. In M. Juan-Garau, & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 51–68). Cham: Springer. https://doi.org/10.1007/978-3-319-11496-5_4.
- Rumlich, D. (2014). Prospective CLIL and non-CLIL students' interest in English (classes): A quasi-experimental study on German sixth-graders. In R. Breeze, C. Martínez Pasamar, C. Llamas Saíz, & C. Taberero Sala (Eds.), *Integration of theory and practice in CLIL* (pp. 75–95). Amsterdam: Rodopi. https://doi.org/10.1163/9789401210614_006.
- Rumlich, D. (2017). CLIL theory and empirical reality: Two sides of the same coin? A quantitative-longitudinal evaluation of general EFL proficiency and affective-motivational dispositions in CLIL students at German secondary schools. *Journal of Immersion and Content-Based Language Education*, 5(1), 110–134. <https://doi.org/10.1075/jicb.5.1.05rum>
- Salvador-García, C., Chiva-Bartoll, O., & Capella-Peris, C. (2019). Bilingual physical education: The effects of CLIL on physical activity levels. *International Journal of Bilingual Education and Bilingualism*, 1–10. <https://doi.org/10.1080/13670050.2019.1639131>
- San Isidro, X. (2018). Innovations and challenges in CLIL implementation in Europe. *Theory Into Practice*, 57(3), 185–195. <https://doi.org/10.1080/00405841.2018.1484038>
- Sánchez-García, R., & Pavón-Vázquez, V. (2021). Students' perceptions on the use of Project-Based Learning in CLIL: Learning outputs and psycho affective considerations. *Latin American Journal of Content & Language Integrated Learning*, 14(1), 69–98. <https://doi.org/10.5294/laclil.2021.14.1.3>
- Seikkula-Leino, J. (2007). CLIL learning: Achievement levels and affective factors. *Language and Education*, 21, 328–341. <https://doi.org/10.2167/le635.0>
- Sierra, J. M. (2011). CLIL and project work: Contributions from the classroom: Content and language integrated learning. In Y. Ruiz de Zarobe, J. M. Sierra, & F. Gallardo del Puerto (Eds.), *Contributions to multilingualism in European contexts* (pp. 211–239). Bern: Peter Lang.
- Sierra, J. M. (2016). Cooperative projects in a CLIL course: What do students think? In D. Lasagabaster, & A. Doiz (Eds.), *CLIL experiences in secondary and tertiary education: In search of good practices* (pp. 69–97). Peter Lang: Bern.
- Simpson, A. (2019). Separating arguments from conclusions: The mistaken role of effect size in educational policy research. *Educational Research and Evaluation*, 25(1–2), 99–109. <https://doi.org/10.1080/13803611.2019.1617170>
- Siqueira, D. S. P., Landau, J., & Paraná, R. A. (2018). Innovations and challenges in CLIL implementation in south America. *Theory Into Practice*, 57(3), 196–203. <https://doi.org/10.1080/00405841.2018.1484033>
- Somers, T., & Llinares, A. (2021). Students' motivation for content and language integrated learning and the role of programme intensity. *International Journal of Bilingual Education and Bilingualism*, 24(6), 839–854. <https://doi.org/10.1080/13670050.2018.1517722>
- Stoet, G., Bailey, D. H., Moore, A. M., & Geary, D. C. (2016). Countries with higher levels of gender equality show larger national sex differences in Mathematics anxiety and relatively lower parental mathematics valuation for girls. *PLoS One*, 11(4). <https://doi.org/10.1371/journal.pone.0153857>
- Sunderland, J. (2010). Theorizing gender perspectives in foreign and second language learning. In R. M. Jiménez Catalán (Ed.), *Gender perspectives on vocabulary in foreign and second languages* (pp. 1–19). London: Palgrave Macmillan.
- Sylvén, L. K. (2017). Motivation, second language learning and CLIL. In A. Llinares, & T. Morton (Eds.), *Applied linguistics perspectives on CLIL* (pp. 51–65). Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.47.04sylv>.
- Sylvén, L. K., & Thompson, A. S. (2015). Language learning motivation and CLIL: Is there a connection? *Journal of Immersion and Content-Based Language Education*, 3(1), 28–50. <https://doi.org/10.1075/jicb.3.1.02sylv>
- Sytsma, M., Panahon, C., & Houlihan, D. D. (2019). Peer tutoring as a model for language and reading skills development for students who are English language learners. *Journal of Applied School Psychology*, 35(4), 357–379. <https://doi.org/10.1080/15377903.2019.1597796>
- Topping, K. (2015). Peer tutoring: Old method, new developments/tutoría entre iguales: Método antiguo, nuevos avances. *Infancia Y Aprendizaje*, 38(1), 1–29. <https://doi.org/10.1080/02103702.2014.996407>
- Torgerson, C. J., & Torgerson, D. J. (2007). The need for pragmatic experimentation in educational research. *Economics of Innovation and New Technology*, 16(5), 323–330. <https://doi.org/10.1080/10438590600982327>
- Ushioda, E., & Dörnyei, Z. (2017). Beyond global English: Motivation to learn languages in a multicultural world: Introduction to the special issue. *The Modern Language Journal*, 101(3), 451–454. <https://doi.org/10.1111/modl.12407>
- Watzek, V., Anselmann, V., & Mulder, R. H. (2019). Team learning and emotions during teamwork: A qualitative study. *Research Papers in Education*, 34(6), 769–789. <https://doi.org/10.1080/02671522.2019.1568525>
- Zeneli, M., & Tymms, P. (2015). A review of peer tutoring interventions and social interdependence characteristics. *International Journal for Cross-Disciplinary Subjects in Education*, 5(2), 2504–2510. <https://doi.org/10.17220/ijpes.2018.03.004>
- Zirkel, S., García, J. A., & Murphy, M. C. (2015). Experience-sampling research methods and their potential for education research. *Educational Researcher*, 44(1), 7–16. <https://doi.org/10.3102/0013189X14566879>
- Zydatiá, W. (2012). Linguistic thresholds in the CLIL classroom? The threshold hypothesis revisited. *International CLIL Research Journal*, 1(4), 17–28.