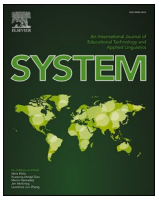
System 123 (2024) 103355

Contents lists available at ScienceDirect

System

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



# A longitudinal test of the impact of CLIL on language emotions and learning motivation☆

Laurence Mettewie [a,*], Luk Van Mensel [a,b], Audrey De Smet [a,c], Benoît Galand [c]

[a] Université de Namur: NaLTT Institute, Rue de Bruxelles 61, BE-5000, Namur, Belgium
[b] Xi'an Jiaotong – Liverpool University: School of Humanities and Social Sciences, Ren'ai Road 111, Suzhou, (Jiangsu), PR China
[c] Université Catholique de Louvain: IPSY, Place Cardinal Mercier 10/L3.05.01, Louvain-la-Neuve, BE-1348, Belgium

ARTICLE INFO

*Keywords:*
CLIL
Longitudinal
Language learning motivation
Emotions
Anxiety
Enjoyment
English
LOTE
Dutch

ABSTRACT

This study relies on a longitudinal design to test the added value of a Content and Language Integrated Learning approach (CLIL) for socio-affective outcomes. 756 French-speaking pupils at primary or secondary school, learning either English or Dutch (as a 'language other than English'), in a CLIL track or in non-CLIL mainstream foreign language classes, participated in the study. The participants twice completed a questionnaire and several tests over an 18-month interval. The questionnaire included items measuring their emotions in the classroom (anxiety and enjoyment) and motivation for language learning (perceived task value, expectancy for success, and perceived cost). A range of individual background characteristics, including initial vocabulary knowledge in the target language, were included in the analyses. The findings of the group comparison between CLIL and non-CLIL (between-subject) showed that the CLIL group reported more favorable emotions and motivation for language learning, in line with previous cross-sectional research on socio-affective outcomes. However, the longitudinal results (within-subject) indicated that the effects of CLIL were limited, particularly when initial vocabulary knowledge was factored in. Our findings thus contradict or moderate the (largely theoretical) claim that CLIL de facto would generate advantages in terms of socio-affective factors such as language emotions and learning motivation.

## 1. Introduction

CLIL (Content and Language Integrated Learning) refers to an instructional approach using a target language to learn subject matters in non-linguistic courses, aiming at acquiring both the language and the content. The integration (the I in CLIL) of these two dimensions is supposed to lead to synergies in the learning process, involving besides content and communication also cognition and cultural awareness (cf. the C 4 model by Coyle, 2011). The target languages can be international languages (English), state languages (French, Dutch or German in Belgium), regional and minority languages (Frisian in The Netherlands, German or French in northern Italy, Sami in Finland), as documented in the Eurydice reports (2006, 2023). Over the past three decades, CLIL has been promoted by the European Council and Commission as a successful method to develop multilingual competences (which are at the heart of the

☆ Submission for https://www.sciencedirect.com/journal/system/publish/guide-for-authors.

* Corresponding author. Université de Namur, LLG/NaLTT Institute, rue de Bruxelles 61, BE-5000, Namur, Belgium.
*E-mail addresses:* laurence.mettewie@unamur.be (L. Mettewie), Luk.VanMensel@xjtlu.edu.cn (L. Van Mensel), benoit.galand@uclouvain.be (B. Galand).

https://doi.org/10.1016/j.system.2024.103355
Received 23 January 2024; Received in revised form 21 May 2024; Accepted 27 May 2024
Available online 28 May 2024

European Union's linguistically and culturally diverse project) and has spread out through Europe and far beyond from pre-primary to higher education (Eurydice, 2023). The outcomes of CLIL as a didactic approach are therefore a main concern and research area. However, not all aspects of CLIL receive the same attention and some results bring more questions than answers.

In CLIL research, socio-affective factors have not been as extensively studied as for instance linguistic outcomes or didactic processes and methods (De Smet, Mettewie, Galand, Hiligsmann, & Van Mensel, 2018). This might come as a surprise, since the socio-cultural and educational dimensions of the multilingual pedagogical approach (including positive attitudes, learning enjoyment and motivation) have been put forward, from the very start, as one of the possible beneficial outcomes of CLIL (see e.g. Dalton-Puffer, 2008; Eurydice, 2006; Marsh, 2000). As detailed below, some cross-sectional data (based on only one measure at a certain point of time), almost exclusively focusing on English, seems to be consistent with these expected socio-affective benefits of CLIL. But globally, results are not unambiguous and call for larger and better controlled studies (Cenoz et al., 2014). A recurrent issue in analyzing the potential role of the CLIL approach is whether CLIL learners have developed better socio-affective profiles by participating in CLIL classes or whether CLIL tracks attract learners with more favorable initial socio-affective profiles. Moreover, few longitudinal studies (including several measures at different points of time) have tested if the development of socio-affective factors over a certain period is higher for CLIL classes than for non-CLIL mainstream classes (Goris et al., 2019). Partly due to differences regarding methodological aspects, their findings point at different directions, thus presenting a rather blurry picture.

Therefore, the aim of the present study is to test if the evolution over time of language learning emotions (anxiety and enjoyment) and motivation (perceived task value, self-efficacy, cost of learning) is different among CLIL and non-CLIL students, controlling for a range of background variables including initial vocabulary knowledge in the target language. We also wanted to check if age cohort (primary vs. secondary school) and target language (English vs. another language) moderate these effects.

## 2. Socio-affective factors in CLIL research

From the early days of CLIL in Europe, both descriptive and theoretical works mention socio-affective aspects as being part of this educational approach. In 2006, the first statistical report of the European Commission on CLIL in Europe mentioned, besides a range of general aims, socio-economic aims (offering internationally better job perspectives), socio-cultural aims (conveying values of tolerance and respect) and improving learning motivation (educational and linguistic aims), through a practical use of the target language (Eurydice, 2006). Earlier on, Marsh (2000) indicated that the natural setting in which the target language is used, together with the combination of language and content learning, would boost youngsters' language learning motivation, as it would emphasize what Baetens Beardsmore and Kohls (1988) coined the 'immediate pertinence' of language learning in multilingual education (see Housen, 2002). This usefulness linked to a sense of fun experienced in an alternative educational setting would in turn help to nurture youngsters' positive attitudes as they see successes achieved, further stimulating their learning motivation. Similarly, when describing the five dimensions for the implementation of CLIL in Europe (culture, environment, content, language and learning), Marsh et al. (2001) considered an increased learning motivation as an outcome of the learning dimension, as CLIL would support the development of individual learning strategies through diversified teaching methods and forms of classroom practices. The immediate relevance of the dual approach in CLIL, and therefore the very nature of the language input, is in this sense theoretically regarded as being more important than the amount of input and is as such considered an added value to the language learning process itself (Maljers et al., 2007). However, it is also acknowledged that the CLIL approach is likely to entail a range of challenges, which then might put socio-affective factors such as motivation, enjoyment and self-esteem at risk (Coyle et al., 2010), for both learners and teachers (see e.a. Hofstadler et al., 2020; Talbot et al., 2021).

Empirical research on the possible socio-affective benefits of CLIL usually found that CLIL learners report (a) higher language learning motivation (Amengual-Pizarro & Prieto-Arranz, 2014; Coyle, 2011; Doiz et al., 2014; Lasagabaster, 2011; Lasagabaster & Doiz, 2017; Mearns et al., 2020; Thompson & Sylvén, 2015), (b) more positive attitudes towards the target CLIL subject, language and community (Amengual-Pizarro & Prieto-Arranz, 2014; De Smet, Mettewie, Hiligsmann, Galand, & Van Mensel, 2023; Mettewie & Lorette, 2014; Navarro-Villarroel, 2011; Pirskanen, 2009), (c) more multicultural openness (Lasagabaster, 2009) and (d) lower anxiety (Coyle, 2011; De Smet et al., 2018; Thompson & Sylvén, 2015). For a more detailed overview of research on emotions and attitudes and motivation in CLIL, we refer to De Smet et al. (2018, 2023).

Nevertheless, other studies found no differences or even a disadvantage for the CLIL groups under investigation (Amengual-Pizarro & Prieto-Arranz, 2014; Fernández Fontecha & Canga Alonso, 2014; Heras & Lasagabaster, 2015; Lasagabaster & Doiz, 2017). This is for example the case in Spain (Doiz et al., 2014), where grade 7 pupils learning English through CLIL reported at the same time higher motivation but also higher anxiety when compared to pupils from the non-CLIL group, i.e. those involved in mainstream English as a foreign language instruction class. The authors suggest that the more demanding learning situation in CLIL probably explains higher emotions, whether positive or negative. However, in a second and slightly older cohort (grade 9), the difference in anxiety seemed to disappear, suggesting that CLIL learners became accustomed to the approach and the associated expectations of the CLIL program. In a similar vein, Seikkula-Leino (2007) also warns about a possible detrimental effect linked to the high demands of CLIL, as CLIL learners rated themselves very low as English language learners, even though the motivation to learn a foreign language was higher in CLIL. In Germany, Möller (2021) identified different types of anxiety linked to the language learning (fear of success, activating or inhibiting exam anxiety) of English, with marginal differences between the German-speaking CLIL and non-CLIL grade 11 pupils, though CLIL students were slightly more inhibited by exam anxiety. In Sweden, on the contrary, research by Thompson and Sylvén (2015) in grade 10 indicated lower anxiety in the CLIL group, but more favorable attitudes towards English as a target language amongst the non-CLIL pupils. In French-speaking Belgium, De Smet et al. (2018) observed that CLIL students reported lower classroom anxiety than non-CLIL students, both in grade 5 and grade 11, whereas a difference in enjoyment in favor of CLIL students only appeared in grade 11 but not in

grade 5. Similarly, De Smet et al. (2023) found more positive attitudes towards the target language (English or Dutch) and higher language learning motivation (higher perceived task value and self-efficacy, lower perceived cost of learning) among CLIL learners compared to the non-CLIL learners in mainstream English or Dutch language classes. Interestingly, these differences appeared to be more clear-cut for Dutch than for English, and at secondary level (grade 11) than at primary level (grade 5). These results suggest that the CLIL learning experience could - to a certain extent - mitigate the expected decrease in motivation with age for formal foreign language learning (Chambers, 1999; Williams et al., 2002), and indicate the potential impact of target language on motivational processes, thereby stressing the relevance of studying 'languages other than English' (LOTE), as also noted by Dewaele and Proietti Ergün (2020).

It is worth noting that in all these studies the authors tended to explain the differences in socio-affective outcomes by the CLIL or non-CLIL approaches, even though the design of the study prevents them from identifying whether the (more or less) positive outcomes recorded amongst CLIL learners are the results of the pedagogical approach itself. On the one hand, these studies compared CLIL and non-CLIL students at one moment in time (cross-sectional design) and not changes among CLIL and non-CLIL students, making impossible to know if the observed differences between students reflect initial differences in socio-affective factors (Thompson & Sylvén, 2015) or the added-value of the CLIL approach. To overcome this limitation, studies measuring outcomes at least on two occasions (longitudinal design) are needed, as they allow to compare the evolution of emotions and motivation between CLIL and non-CLIL students. As an illustration, Goris et al. (2019) systematically reviewed longitudinal studies on linguistic outcomes for CLIL vs. non-CLIL learners of English in European contexts and concluded that there was no "unequivocal support for the hypothesis that learners in a CLIL class will develop more EFL proficiency over a certain period than their mainstream counterparts" (Goris et al., 2019, p. 692). On the other hand, observed differences in motivation or emotion between CLIL and non-CLIL groups could reflect other influences than instructional approaches, like the selection processes for accessing CLIL (Fehling, 2009; Rümlich, 2014) and/or differences in learners' socio-economic status, non-verbal intelligence, or initial language proficiency (Van Mensel et al., 2020). One way to overcome this limitation would be to randomize students between CLIL and non-CLIL classes, but this is usually both practically and ethically impossible. Another way to try to make CLIL and non-CLIL students as comparable as possible is to statistically control for several individual background variables. However, gender is often the only individual background variable included in previous studies on emotions and motivation. Finally, another limitation of most available studies is the small number of pupils involved (often less than 50 in CLIL and non-CLIL, sometimes no more than one class in each group), threatening the replicability and generalizability of the findings (Cheung & Slavin, 2016).

## 3. Longitudinal research on socio-affective profiles in CLIL

As already mentioned, longitudinal studies on socio-affective effects of CLIL are particularly scarce. In their overview of second language motivation research published between 2010 and 2019, Hadi Mahmoodi and Yousefi (2022) list only two studies (Lasagabaster & Doiz, 2017; Pladevall-Ballester, 2019) that refer to longitudinal research in a CLIL context. Pladevall-Ballester (2019) followed 138 CLIL and 149 non-CLIL learners of English from 5th to 6th grade in Catalonia. Using a longitudinal regression model with both intergroup and intragroup comparisons, very few significant differences were found between CLIL and non-CLIL group on foreign language motivation. At the last measurement time, only a factor mixing perceived easiness and self-efficacy was favorable for the CLIL group (due to a drop in the non-CLIL group), whereas both groups had similar levels regarding interest and perceived utility. Lasagabaster and Doiz (2017) followed 199 CLIL and 111 non-CLIL Spanish participants learning (through) English from the first/third year of secondary education to the third/fourth year. Contrary to their expectations and previous findings (see Doiz et al., 2014), the authors observed no main differences between motivation of CLIL and non-CLIL pupils in the last phase of the study. The authors suggest this might be due to the hegemonic position of English, sustaining equal motivation in both education types. However, while non-CLIL pupils' motivation to learn English did not decline during the study, the motivation of the younger CLIL pupils seemed to wane over time. This echoes the findings of Papaja and Rojczyk (2013), who found that the pleasure and satisfaction of learning through CLIL waned towards the end of the third year due to pressure of the final examinations (see also Möller, 2021). This makes Lasagabaster and Doiz (2017) wonder whether motivation may also wane once CLIL ceases to be a novelty (see also Dalton-Puffer & Smit, 2013). Nevertheless, it should be stressed that in the study of Lasagabaster and Doiz (2017) only the younger pupils, who had been selected by the schools to participate in CLIL depending on their level of English, suffered a motivational decline. The older ones who had not undergone a selection process maintained high motivation, although they became slightly more anxious to speak English during CLIL classes. Therefore, *age* and *selection* into CLIL seem important factors for motivation, although the authors also suggest the *difficulty of the subjects* taught through CLIL might play a role. In a more recent publication, San Isidro and Lasagabaster (2022) discuss the findings of a two-year longitudinal study looking at the attitudes and motivations towards English of two groups of Spanish students ($n = 20$ and $n = 24$) from the same school, as well as their parents. In this sample, both CLIL and non-CLIL students (and their parents) displayed positive long-term motivation, and the CLIL group even more so. However, the authors not only note the small sample size but also acknowledge the importance of *the teachers' commitment* in this school in obtaining such positive results.

A study published in an edited volume (Sylvén, 2019) and not included in Mahmoodi and Yousefi's (2022) list, is the longitudinal follow up of a Swedish project comparing 'motivational' profiles (including anxiety, but not enjoyment) of pupils in English CLIL and non-CLIL, mainstream language learning tracks (Thompson & Sylvén, 2015). The data were collected at the very beginning of the CLIL-project (1st semester grade 10, aged 15–16 years) to identify possible preliminary differences in profile (T1) and at the end of high school (T2), roughly two years later (grade 12, aged 18–19 years), to assess the stability of the motivational patterns. The authors found that CLIL pupils were more motivated, even *before starting the program* and that at T2 these more positive profiles of the CLIL pupils ($n = 60$) over their non-CLIL counterparts ($n = 51$) were maintained (Thompson & Sylvén, 2019). The results also revealed that

English anxiety in the CLIL group tended to be lower and to decrease over time, whereas the non-CLIL group had more anxiety, which remained stable at T2. This tends to confirm the general trend of a higher degree of *dynamicity* of socio-affective variables in CLIL, for better or for worse, as the study also indicated a decrease in interest and appreciation of English in the CLIL group, suggesting that the initial enthusiasm might have worn out, despite the positive selection bias at the onset (cf. Lasagabaster & Doiz, 2017). Linked to the initial bias potentially present in analyzing CLIL tracks, Ohlberger and Wegner (2019) conducted a longitudinal intervention study through delimited 'CLIL modules' in Germany. The impetus to investigate CLIL modules (short sequences of teaching hour in CLIL implemented in mainstream, non-CLIL tracks) rather than complete CLIL tracks themselves can be explained by an effort to overcome the 'creaming effect' or selection of high-achieving students, which may be induced by CLIL enrolment procedures in Germany (Rumlich, 2016). The authors measured the affective effects of two biology units taught in English on students aged 15 to 16 (n = 166; control group n = 83). The results indicated that the 'CLIL modules' reduced English anxiety and increased self-efficacy over a period of two years compared to the control group, suggesting a positive effect of the CLIL approach.

Even if the studies reviewed above used a longitudinal design, they included a limited number of background variables. Consequently, like we discussed for cross-sectional studies, it is impossible to know from those studies to what extent initial differences and a change in emotions and motivation reflect the effect of instructional approach (CLIL) or other intake differences among students, like for instance socio-economic status (SES). Most importantly, available longitudinal studies on the link between CLIL and socio-affective variables did not consider initial target language (TL) achievement or proficiency. Specific knowledge and skills in one domain have a strong effect on further learning in this domain (Tricot & Sweller, 2014), but they also predict further motivation and emotions in this domain (Möller et al., 2020; Toste et al., 2020). Students with a high initial proficiency level in a TL thus have a higher probability to feel more enjoyment, less anxiety, higher expectancy of success and higher interest toward learning activities related to that language than students with low initial proficiency level. Initial TL proficiency thus appears to be a more proximal predictor than more distal characteristics like SES or non-verbal intelligence.

## 4. Rationale and research questions

All in all, the studies reviewed above show that the effects of CLIL on socio-affective variables are not as clear-cut as is often believed. In fact, a range of parameters (including age, educational level, learning context, teachers' commitment, exam pressure, selection bias, target language) appear to impact different socio-affective variables in various ways. Additionally, the constructs used to operationalize motivation tend to differ from study to another, depending on the underlying theoretical assumptions, which complicates comparison of the outcomes.

The few longitudinal studies available until now largely support this observation. Instead of providing definite answers, they highlight the variability of motivational patterns over time and underscore the complexity of the matter. Additionally, these longitudinal studies have some limitations, primarily involving relatively small sample sizes and a narrow focus on English as the target language. Moreover, they have applied minimum control for the influence of other features of learners' backgrounds on their motivation, particularly their previous language proficiency, which makes it challenging to filter out the contribution of CLIL as an instructional approach.

Therefore, in the present study, we wanted to address the following research questions:

1. Is the evolution of language learning emotions (classroom anxiety and enjoyment) more positive among students in CLIL classes than in mainstream non-CLIL classes?
2. Is the evolution of language learning motivation (expectancy for success, perceived task value and cost) more positive among students in CLIL classes than in mainstream non-CLIL classes?
3. Are changes over time different depending on the target language (English vs. a language other than English)?
4. Do the effects of CLIL remain the same when the level of initial proficiency in the target language and other background variables are considered?

In order to reduce the effect of a possible initial selection bias amongst CLIL learners and to better understand the role of CLIL in the development of socio-affective reactions toward language learning, it seems crucial (a) to involve larger cohorts of respondents, (b) to control, in the analysis, as much as possible for learners' background by including initial language proficiency and proxies for SES, cognitive intelligence, degree of school retention and multilingualism, and (c) to implement longitudinal designs. This is the rationale behind the present study, in which we followed cohorts of pupils in CLIL tracks and in non-CLIL tracks (control groups learning the target language only through the classical language classes). The cohorts had either English or Dutch as a target language, both at 5th/

**Table 1**
Distribution of participants across educational level, target language and instructional approach at T2 (N = 756).

| | CLIL | | Non-CLIL | | TOTAL |
|---|---|---|---|---|---|
| | Dutch | English | Dutch | English | |
| Primary | 157 | 87 | 61 | 76 | 381 |
| Secondary | 129 | 88 | 81 | 77 | 375 |
| TOTAL | 286 | 175 | 142 | 153 | 756 |

6th grade of primary school and 11th/12th of secondary school. Next to measuring twice their affective profiles (motivation and emotional engagement in the classroom) with an 18-month interval between Time 1 (N = 896) and Time 2 (N = 756, see Table 1), we collected a range of background variables to control for.

## 5. The research context: CLIL in French-speaking Belgium

In the Federal structure of Belgium, the organization of the educational system depends on the linguistic community in charge (Dutch-, French- or German-speaking) and the region it is located in. Due to the socio-political, economic, and historical context, the legislation defining language education and the languages of instruction, as is the case with CLIL, is strictly regulated; this has allowed each language community to forge a tailor-made system (for details see Mettewie & Van Mensel, 2023). For this contribution we will spotlight CLIL in the French-speaking community in Wallonia, which has been allowed since 1998 (Chopey-Paquet, 2007) and was originally called "immersion".[1] These programs allow for the use of another language of instruction (Dutch, German or English), besides French (official language of the school), for nearly all courses other than the language courses. The CLIL programs can be started at different moments (from the last year of kindergarten at the age of ±5 to the second part of secondary school, about ±15 years old) and can be organized until the end of compulsory education (at the age of 18). The percentage of courses in the CLIL language can vary from 25 to 75% at primary level and from 20 to 45% at secondary level, depending on the pedagogical project of the school. Since 1998, CLIL or "immersion" has enjoyed considerable success, with 332 schools in 2019–20 (206 primary/116 secondary) and 42.775 pupils involved.[2] Contrary to the regular foreign language classes, for which English is the most popular language amongst learners (43% at primary level and 66% at secondary level in 2019–20), the dominant target language in the CLIL tracks in respectively kindergarten, primary and secondary schools is Dutch (77, 71 and 54%), followed by English (22, 26, 41%) and anecdotally German. It is also important to note that to allow for equal access to CLIL tracks in French-speaking Belgium no financial contribution is asked from the parents, nor is there an official selection procedure (based on intelligence tests or exam results) allowed to admit a pupil in a CLIL class. Anecdotal evidence, however, suggests that intake interviews and procedures might encourage or discourage some pupils (Van Mensel et al., 2020).

## 6. Method

### 6.1. Participants

At baseline (T1), 896 pupils participated to this study. They were at the beginning of their fifth year of either primary (Grade 5) or secondary (Grade 11) school in French-medium education, learning Dutch or English as a target language (TL) through either CLIL or non-CLIL. Eighteen months later, 756 of them participated to a second data collection, when they were either at the end of Grade 6 or Grade 12. The others were absent at the later data collection, were retained, or moved to another school. The participants came from 13 primary and 9 secondary schools across Wallonia in French-speaking Belgium, which were willing to participate and selected to obtain diversity regarding location, socio-economic level, and organizing authority. Table 1 displays the distribution of the participants across the different educational levels, target languages and instructional approaches.

The sample at T2 is well balanced in terms of gender (53% girls, 47% boys). 65,9% of pupils indicated that they speak exclusively French at home, 27,1% indicated that they speak French at home in combination with another language, and 7% that they speak exclusively another language than French at home. The highest degree of the mother was used as a proxy of participants' socio-economic status (SES) and comparisons revealed significantly higher SES in CLIL and in Dutch as a TL compared to English. Pupils' non-verbal intelligence was measured through Raven's Standard Progressive Matrices (Raven et al., 1998), indicating means in accordance with the norms for the respective ages of the participants, but also significantly higher non-verbal intelligence scores in CLIL and in Dutch compared to English. Regarding the school retention rates, 12.6% of the participants reported having failed at least one year in their curriculum, with again a more favorable profile in CLIL (secondary school) and for Dutch as a TL. Finally, because of the totally free structure of CLIL provisions in French-speaking Belgium (see above), it is impossible to map in a clear and comprehensive way how many hours of CLIL and/or language teaching the 756 pupils have had by the time of the data collections. Each pupil has its own pace and (intra and extra) curriculum input (see Goris et al., 2019, for similar observations in different school systems). However, all pupils from our sample involved in the CLIL tracks had between 3 and 12 years of CLIL, and pupils in the traditional language classes all started at latest in the 5th grade, which means that at T1 the cohorts in primary schools had about 2 months of TL teaching and the cohorts in secondary schools at least 6 years of the same TL. Similarly, because of the large sample, the free CLIL provisions and the individual trajectories of the pupils, it is impossible to include any description of the instructional quality of the CLIL teaching/learning processes the respondents have enjoyed prior or during the data collection. For more details about the background characteristics of the sample, see Van Mensel et al. (2020).

### 6.2. Procedure

Participants were invited by a member of the research team to complete a pen-and-paper questionnaire during school hours at T1

[1] https://www.gallilex.cfwb.be/fr/leg_res_02.php?ncda=22229&referant=l01.
[2] http://enseignement.be - AGE - Dossier rentrée - 2020–2021 (ressource 15,923).pdf.

and an identical online version on campus at T2. The member of the research team explained the aims of the study to the students, answered their questions, informed them about the confidentiality of their answers and reminded them that participation was voluntary. With the approval of school boards and the university's ethical committee, active consent was requested for students, and passive consent (assuming consent unless explicitly declined) was used for parents to maximize the representativeness of the sample (Pokormy et al., 2001). For more details about the design of the larger study and the general data collection, see Hiligsmann et al. (2017).

### 6.3. Measures

There were two slightly different versions of the questionnaire, as items were adapted to pupils' (first) TL (English or Dutch). All items were however formulated in French, the main language of education and most pupils' first language (93%) alone or in combination (see above).

#### 6.3.1. Language learning emotions

Items assessing classroom anxiety were adapted from the Foreign Language Classroom Anxiety Scale (Horwitz et al., 1986) and items assessing classroom enjoyment from the Foreign Language Enjoyment scale (Dewaele & MacIntyre, 2014) in order to fit the specific context and needs of the study. A pilot study showing that CLIL pupils did not distinguish between regular language classes and subject classes taught through the target language when it comes to their answers regarding classroom anxiety and enjoyment, some items were selected and adapted to build a common version for CLIL and non-CLIL pupils (targeting both language and subject classes in the case of CLIL pupils). This resulted in one 9-item scale measuring *foreign language anxiety* and one 5-item scale measuring *foreign language enjoyment* in the classroom. Both were measured through seven-point Likert scales. Exploratory factor analyses supported the distribution of the selected items into their respective scales. Both scales also showed satisfactory internal consistency with $\alpha = 0.85/.88$ for anxiety and $\alpha = 0.78/.80$. for enjoyment, at respectively T1 and T2. For a detailed presentation of the items, see De Smet et al. (2018).

#### 6.3.2. Language learning motivation

Relying on the motivational framework of expectancy-value theory (Eccles & Wigfield, 2020; Wigfield & Eccles, 2000), which is widely used in many fields of education and psychology research, we assessed *expectancy for success*, *perceived task value* and *cost*. This framework complements the dominant models of L2 motivation as its components were found to change throughout school years, to be responsive to interventions, and to predict persistence, achievement and course/study choice (see review by Wigfield & Cambria, 2010). It looks at the relations between the self and the task at hand, integrating components from classic language learning models (instrumental and intrinsic orientations, self-relevance) and enriching them with new components (expectancy, cost).

The scales were constructed from existing adaptations in French (Galand & Hospel, 2015) and measured as 7-point Likert scales, ranging from strongly disagree (1) to strongly agree (7). Exploratory factor analyses supported the distribution of the selected items into their respective scales. Reliability analyses indicated satisfactory internal consistency for all scales. This resulted in one 5-item scale measuring '*expectancy for success*' ($\alpha = 0.76/.79$), one 8-item scale measuring '*perceived task value*' ($\alpha = 0.85/.88$), and one 3-item scale measuring '*cost*' ($\alpha = 0.71/.75$). 'Expectancy for success' refers to the learner's evaluation of how well he/she could perform, whereas 'perceived task value' is an indicator of interest, utility and importance that participants attribute to target language learning. 'Cost' reflects one's estimated effort and time investment she/he needs to achieve success in the learning process. For a more detailed presentation of the items, see De Smet et al. (2023).

#### 6.3.3. Target language vocabulary

At T1, the pupils' *TL receptive vocabulary* was measured through a computerized version of the Peabody Picture Vocabulary Test for English (PPVT-IV; Dunn et al., 2007), and the Dutch version of the same test (PPVT-NL-III; Dunn & Dunn, 2005). The PPVT is a standardized task consisting of a series of four-picture sets from which the participants need to select the drawing corresponding to a

**Table 2**
Means, standard deviations and correlations for the main study variables.

| | M | *SD* | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Anxiety T1 | 3,56 | *1,25* | – | | | | | | | | | |
| 2. Enjoyment T1 | 4,95 | *1,13* | −.32 | – | | | | | | | | |
| 3. Expectancy T1 | 5,32 | *1,00* | −.39 | .55 | – | | | | | | | |
| 4. Task value T1 | 4,97 | *1,17* | −.25 | .61 | .57 | – | | | | | | |
| 5. Cost T1 | 2,53 | *1,33* | .36 | .42 | −.41 | −.46 | – | | | | | |
| 6. Anxiety T2 | 3,28 | *1,35* | **.60** | −.22 | −.31 | −.15 | .22 | – | | | | |
| 7. Enjoyment T2 | 4,72 | *1,15* | −.22 | **.52** | .36 | .36 | −.27 | −.36 | – | | | |
| 8. Expectancy T2 | 5,29 | *1,06* | −.33 | .39 | **.56** | .39 | −.34 | −.44 | .57 | – | | |
| 9. Task value T2 | 4,76 | *1,31* | −.26 | .46 | .41 | **.63** | −.37 | −.23 | .56 | .59 | – | |
| 10. Cost T2 | 2,64 | *1,37* | .27 | −.25 | −.28 | −.29 | **.42** | .40 | −.47 | −.50 | −.52 | – |
| 11. TL vocabulary T1 | 0,00 | *1,00* | −.32 | −.21 | .02 | .01 | −.11 | −.25 | −.15 | .03 | .11 | −.14 |

Note. All correlation coefficients above 0.10 are significant at $p < 0.01$.

word they hear; the score consists of the number of correct answers identified. The task is frequently used to assess receptive vocabulary (Goriot et al., 2021). The raw scores were transformed into standardized scores for each TL separately (English or Dutch). These standardized scores were then used in subsequent analyses. Preliminary analyses showed that TL vocabulary was higher among CLIL students, both in primary (M = −0.51 *vs.* M = −1.32) and in secondary schools (M = 1.01 *vs.* M = 0.17).

For a detailed description of the scales, lay-out, factor analyses and clustering of the items into variables, we refer to De Smet et al. (2018, 2023).

## 7. Results

The data were analyzed using the Statistical Package for Social Sciences (SPSS) 24. Means, standard deviations and correlations for the main variables are presented in Table 2.

Our study design involved two measurement points among CLIL and non-CLIL students. This design allows us to test the hypothesis that the evolution of language emotions and motivation would be more positive with CLIL versus non-CLIL instructional approaches (i.e. an interaction between time and instructional approach) and to test whether this difference is moderated by educational level and target language. We thus first performed repeated measures ANOVAs with the three grouping variables 'instructional approach (non-CLIL/CLIL)', 'target language (English/Dutch)' and 'educational level (primary/secondary)' as between-subject factors, and language emotions and motivation as within-subject factors. Significant interactions between these grouping variables and time would indicate that changes in language learning emotions or motivation from T1 to T2 varied based on the group to which the students belong. Of particular interest for this study are the interactions between time and instructional approaches, as previously mentioned. In a second step, the level of TL receptive vocabulary at T1 was introduced as a covariate in the analyses. Change in the effects of instructional approach between step 1 and step 2 would indicate that these effects could simply reflect the impact of initial differences in TL proficiency. Finally, we tested the effect of adding the other background variables (gender, SES, bilingualism, school retention, non-verbal intelligence) to the analyses.

Systematic between-subject effects were found for target language, many were found for educational level, and some for instructional approach. Those between-subject effects were entirely consistent with the results of cross-sectional comparisons found in previous studies (De Smet et al., 2018; 2023) and were not informative for the present research questions. For the sake of clarity and parsimony, they are not repeated in this section (see section on Socio-affective factors in CLIL research above).

Results for within-subjects analyses showed that, between T1 and T2, anxiety, enjoyment and task value decreased, expectancy remained unchanged, while cost increased (see Tables 3 and 4).

The results for emotions are presented in Table 3. A three-way interaction between instructional approach, target language and time was found in the first step for anxiety (partial $\eta^2 = 0.02$). This interaction remained similar when T1 TL vocabulary is controlled for (partial $\eta^2 = 0.02$). Decomposition of this interaction (see Fig. 1) showed that in English, CLIL students reported a reduction in classroom anxiety from T1 to T2, while this anxiety remained stable for non-CLIL students. In Dutch, a different pattern was found, classroom anxiety slightly decreased among non-CLIL students and remained more stable among CLIL students. In other words, within

**Table 3**
Within- and between-subjects effects for emotions (anxiety and enjoyment).

| | Anxiety | | | | Enjoyment | | | |
|---|---|---|---|---|---|---|---|---|
| | Without T1 TL vocabulary | | With T1 TL vocabulary | | Without T1 TL vocabulary | | With T1 TL vocabulary | |
| ***Within-subjects*** | ***F*** | | ***F*** | | ***F*** | | ***F*** | |
| Time | 44,36 | a | 36,39 | a | 37,62 | a | 28,64 | a |
| Time * Educational level | 11,16 | b | 5,97 | c | 1,92 | | 0,06 | |
| Time * Target language | 1,59 | | 1,42 | | 0,64 | | 0,07 | |
| Time *Instructional approach | 2,56 | | 0,48 | | 6,13 | c | 1,57 | |
| Time * Educational level * Target language | 0,93 | | 0,59 | | 0,06 | | 0,50 | |
| Time * Educational level Instructional approach | 0,00 | | 0,16 | | 2,32 | | 1,17 | |
| Time * Target language * Instructional approach | 14,07 | a | 15,04 | a | 0,09 | | 0,04 | |
| Time * Educational level * Target language * Instructional approach | 1,21 | | 2,13 | | 0,12 | | 0,01 | |
| ***Between-subjects*** | | | | | | | | |
| Educational level | 27,72 | a | 1,71 | | 103,76 | a | 69,73 | a |
| Target language | 7,70 | b | 5,09 | c | 98,66 | a | 96,22 | a |
| Instructional approach | 30,46 | a | 2,03 | | 16,89 | a | 4,66 | c |
| Educational level * Target language | 2,17 | | 2,11 | | 13,29 | a | 11,24 | b |
| Educational level * Instructional approach | 2,39 | | 3,36 | | 6,06 | c | 8,77 | b |
| Target language * Instructional approach | 4,05 | c | 1,16 | | 1,09 | | 0,02 | |
| Educational level * Target language * Instructional approach | 0,07 | | 0,01 | | 0,18 | | 0,21 | |

Note. Educational level: primary vs. secondary; Target language: English vs. Dutch; Instructional approach: non-CLIL vs. CLIL.
[a] $p < .001$.
[b] $p < .01$.
[c] $p < .05$.

**Table 4**
Within- and between-subjects effects for motivation (expectancy of success, task value and cost).

| | Expectancy | | | | Task value | | | | Cost | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Without T1 TL vocabulary | | With T1 TL vocabulary | | Without T1 TL vocabulary | | With T1 TL vocabulary | | Without T1 TL vocabulary | | With T1 TL vocabulary | |
| Within-subject | F | | F | | F | | F | | F | | F | |
| Time | 0,72 | | 0,50 | | 34,03 | a | 27,76 | a | 7,06 | b | 5,41 | c |
| Time * Educational level | 0,95 | | 4,92 | c | 3,50 | | 0,16 | | 0,31 | | 1,78 | |
| Time * Target language | 0,12 | | 0,58 | | 6,30 | c | 7,78 | b | 0,13 | | 0,14 | |
| Time * Instructional approach | 1,07 | | 0,01 | | 12,48 | a | 4,95 | c | 6,44 | c | 1,40 | |
| Time * Educational level * Target language | 0,37 | | 0,00 | | 0,74 | | 2,41 | | 0,31 | | 0,00 | |
| Time * Educational level * Instructional approach | 0,30 | | 0,98 | | 2,73 | | 3,50 | | 2,83 | | 2,53 | |
| Time * Target language * Instructional approach | 0,05 | | 0,02 | | 0,56 | | 0,68 | | 0,17 | | 2,08 | |
| Time * Educational level * Target language * Instructional approach | 1,21 | | 0,23 | | 0,38 | | 0,67 | | 0,66 | | 0,01 | |
| ***Between-subjects*** | | | | | | | | | | | | |
| Educational level | 19,30 | a | 42,74 | a | 5,33 | c | 26,53 | a | 0,17 | | 16,78 | a |
| Target language | 41,31 | a | 34,38 | a | 203,02 | a | 180,07 | a | 62,74 | a | 45,55 | a |
| Instructional approach | 20,03 | a | 0,48 | | 50,53 | a | 12,12 | b | 61,99 | a | 10,71 | b |
| Educational level * Target language | 17,81 | a | 14,59 | a | 12,22 | a | 12,36 | a | 19,97 | a | 19,05 | a |
| Educational level * Instructional approach | 8,57 | b | 12,12 | b | 4,43 | c | 6,11 | c | 2,45 | | 4,59 | c |
| Target language * Instructional approach | 0,11 | | 0,41 | | 4,44 | c | 6,69 | c | 0,38 | | 0,75 | |
| Educational level * Target language * Instructional approach | 0,15 | | 0,12 | | 0,16 | | 0,98 | | 0,27 | | 0,07 | |

Note. Educational level: primary vs. secondary; Target language: English vs. Dutch; Instructional approach: non-CLIL vs. CLIL.
[a] $p < .001$.
[b] $p < .01$.
[c] $p < .05$.

a global trend of anxiety reduction, the difference in reported anxiety between CLIL and non-CLIL learners of English grew apart, whereas it tended to be reduced for CLIL and non-CLIL learners of Dutch.

For enjoyment (see Table 3), the interaction between instructional approach and time was significant at the first step (partial $\eta^2 = 0.01$), indicating that non-CLIL students, contrary to CLIL-students, reported a slight decrease in classroom enjoyment from T1 to T2. But this effect became non-significant and close to zero (partial $\eta^2 = 0.002$) once T1 TL vocabulary was controlled for, suggesting that this effect is due to initial differences in TL proficiency and not to the CLIL instructional approach.

Regarding motivation, the results presented in Table 4 indicate no effect regarding instructional approach for expectancy of success, while a significant interaction between time and instructional approach emerged for perceived task value (partial $\eta^2 = 0.015$). This effect was reduced, but remained significant when T1 TL vocabulary was controlled for (partial $\eta^2 = 0.007$). Means for task value decreased from T1 to T2 and this decrease was slightly smaller among CLIL students than among non-CLIL students (Fig. 2).

The results for cost (Table 4) revealed at first a significant interaction between instructional approach and time (partial $\eta^2 = 0.01$), suggesting that cost increased over time among non-CLIL students, but not among CLIL students. However, this effect did not remain significant once TL vocabulary was introduced as a covariate (partial $\eta^2 = 0.002$).

Finally, complementary analyses indicated that including gender, home bilingualism, SES, school retention rate, or non-verbal intelligence as co-variates did not modify the within-subject results once initial vocabulary knowledge is taken into account. Only initial TL proficiency changed the effects related to instructional approach.

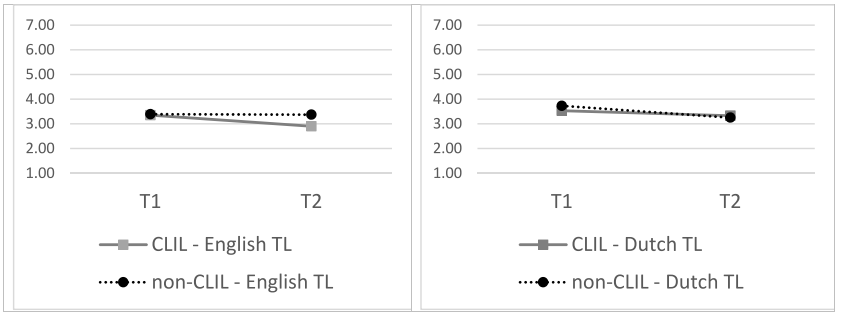



**Fig. 1.** Classroom anxiety means depending on time and instructional approach (corrected for T1 TL vocabulary) for English and Dutch.

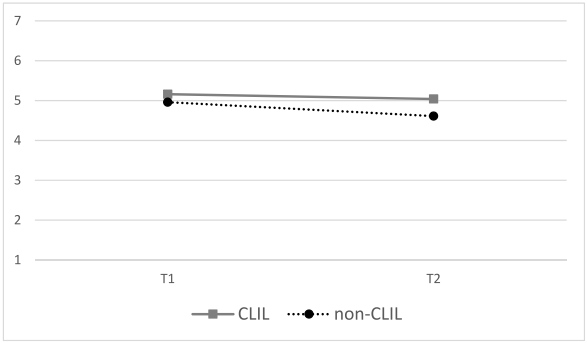



**Fig. 2.** Task value means depending on time and instructional approach (corrected for T1 TL vocabulary).

## Discussion and conclusion

The literature review on socio-affective factors in CLIL, including the limited number of available longitudinal studies, does not present a clear and consistent picture of positive outcomes for CLIL learners compared to non-CLIL controls. Instead, the research reveals a rather blurry picture, with contradictory effects or no effects. Various factors, such as the educational level (primary/secondary), the age of the learners, the duration of time spent in CLIL tracks, the selection process for CLIL classes, the type of subjects taught through CLIL or the type of CLIL approach, exam pressure, the CLIL target language, and individual differences in profile (socio-affective, cognitive or SES), have been noted to potentially impact motivational processes and emotional states.

In an attempt to sharpen the view on the effective contribution of CLIL as an instructional approach, we conducted a longitudinal study comparing the socio-affective states of a representative cohort of French-speaking pupils (n = 756) in Wallonia, Belgium over an 18-month period. The pupils were learning either English or Dutch (the "other" main language of Belgium) at primary or secondary school level, either through a CLIL track or mainstream foreign language classes. To further clarify the differences between the CLIL and non-CLIL groups, we also controlled for individual background characteristics such as initial vocabulary knowledge, gender, bilingualism, school retention rate, socio-economic status, and non-verbal intelligence in the analyses.

First, we can note that the between-subject results at T2 (global comparison of the mean differences between CLIL and non-CLIL with no consideration for the longitudinal effects) are consistent with our previous cross-sectional studies (De Smet et al., 2018; 2023): CLIL students reported more positive socio-affective profiles. However, the analyses detailed in the results section show that few of the within-group effects actually involve the instructional approach (CLIL). Indeed, for the first research question regarding language learning emotions we notice a general decreasing trend over time (less anxiety and less enjoyment), but interactions with CLIL are very limited. The apparently slightly more stable enjoyment in CLIL compared to non-CLIL became indistinguishable when controlling for initial vocabulary knowledge in TL. For anxiety, the interaction remains significant with the initial vocabulary knowledge as a covariate and involves time, CLIL and the target language. For English, we find a pattern similar to the one reported by Thompson and Sylvén (2019), namely that the slightly lower classroom anxiety for learners in CLIL decreases even more, whereas it remains more important and stable for non-CLIL learners. For Dutch, we see the reverse pattern: learners of Dutch in CLIL had slightly lower anxiety than non-CLIL, but that difference disappears at Time 2 as the anxiety of the non-CLIL learners has decreased over time. These findings suggest that the dynamicity of classroom emotions is rather linked to the target language and prior vocabulary knowledge than to the instructional approach.

For the second research question, regarding motivational aspects, the analyses over time indicate that expectancies remain stable within the 18 months of the study, whereas perception of the task value tends to decrease and the perceived cost of the learning process tends to increase. For cost, we notice a similar pattern to enjoyment, as the effect of CLIL (interaction between instructional approach and time) disappears after controlling for prior vocabulary knowledge. For task value, the pattern is comparable to anxiety as CLIL learners show a slightly more favorable profile with no decrease in the perception of task value, compared to a small decrease for non-CLIL pupils. This is the case regardless of the target language and controlling for prior vocabulary knowledge. This pattern of results is similar to the one observed by Pladevall-Ballester (2019) on another motivational variable (self-efficacy).

Regarding the third question, which focused on the differences between the target languages and CLIL over time for emotions and motivation, we only see significant interactions for anxiety in favor of the English learners (vs. Dutch), whose negative classroom emotions continue to decrease significantly between Time 1 and Time 2. For the other variables, longitudinal effects are similar for Dutch and English, suggesting that except for anxiety, TL does not moderate the effect of time on language learning emotions and motivation (Hornstra et al., 2016).

Finally, the fourth question addressed the importance of controlling for background variables. In this set of analyses, the factor more directly related to the reach of the instructional approach was neither gender or instructional level (primary/secondary), nor SES or non-verbal intelligence or school retention rates, but prior target vocabulary knowledge. Hence, these results underline that it is crucial to control for background variables, especially TL proficiency, in order to rule out differences in initial profile (selection bias in CLIL) and to better identify the role of the instructional approach. The results of the present study suggest that initial level of TL proficiency (even measured as simply as through receptive vocabulary knowledge) has an effect on the evolution of TL emotions and

motivation beyond the initial level of those socio-affective variables, highlighting the importance of controlling for TL proficiency in further research on language motivation and emotions (see also Van Mensel & Galand, 2022).

Some limitations of the present study should be kept in mind when interpreting these results. As explained in the Research context and Participants sections, we were unable to trace the precise timing and amount of exposure to the TL for CLIL students, while there was likely variability in exposure among them. However, it is worth noting that participating CLIL students attended at least twice as many TL lessons as non-CLIL students over the course of several years, indicating a substantial difference in exposure between the two groups. While the language qualifications of teachers were similar between these two groups, factors such as the quality of teacher input and out-of-school exposure were not included in the analyses given the large quantitative approach and could differ. Nevertheless, these factors are not necessarily directly related to the CLIL instructional approach itself.

In summary, the longitudinal effects of CLIL on classroom emotions and language learning motivation, as measured in this study, appear to be limited. The evolution of students on those measures was largely similar in both the CLIL and non-CLIL approach, echoing the results of other longitudinal studies (Lasagabaster & Doiz, 2017; Pladevall-Ballester, 2019). Among the five indicators used to measure emotions and motivation, four exhibited an evolution, indicating that the 18-month period was sufficient to capture variation. However, the evolution related to the instructional approach only slightly differs for two out of five factors, namely anxiety and perceived task value, with a non-declining task value among CLIL students and mixed findings depending on TL for anxiety. These two effect-sizes are very small, and it is not sure that they have practical significance. Even at the between-subject level, the effect-sizes for TL and educational level were stronger than those for instructional approach (see also De Smet et al., 2018; 2023).

Our findings thus contradict or moderate the (largely theoretical) claim that the CLIL approach de facto would generate advantages in terms of socio-affective factors. A plausible explanation for this lack of impact could stem from a 'partial' implementation of CLIL, where insufficient emphasis is placed on the 'Integration' aspect (the I in the acronym). Issues related to this aspect with a possible impact on learning emotions and motivation include an inadequate integration between CLIL content and language instructors, insufficient content scaffolding in the target language, or the omission of socio-cultural and socio-affective objectives within CLIL teaching. Additionally, the lack of standardized organizational provisions in the Belgian context, leading to heterogeneous CLIL classes in terms of experience and language exposure, might equally contribute to the limited benefits observed. Nonetheless, it is also possible that the CLIL approach simply fails to deliver the anticipated benefits. Relatedly, it is worth noting that many of the longitudinal studies reviewed in Goris et al. (2019) similarly fail to demonstrate any discernible benefits of the CLIL approach in terms of target language proficiency.

The results of the present study call for two important and related observations. Firstly, the effect of CLIL does not have to be overestimated but should rather be cautiously analyzed as being embedded in a broader language learning context in which many different factors are at play. Secondly, when using CLIL as an independent variable, it is crucial to include a longitudinal perspective in the data collection and to control for different contextual and individual factors, including (prior) language knowledge. This appeared to be a determinant feature in accounting for the dynamicity in both language learning emotions and motivational aspects in this study amongst learners of English and Dutch in French-speaking Belgium. As such, we hope that both researchers and practitioners will get a clearer and more nuanced picture of the effect of CLIL as an instructional approach.

## CRediT authorship contribution statement

**Laurence Mettewie:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Luk Van Mensel:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Data curation, Conceptualization. **Audrey De Smet:** Resources, Methodology, Investigation. **Benoît Galand:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

## Acknowledgments

This work was supported by a Concerted Research Action grant (ARC 14/19-061) awarded to Philippe Hiligsmann (spokesman; UCLouvain), Benoît Galand (UCLouvain), Laurence Mettewie (UNamur), Fanny Meunier (UCLouvain), Arnaud Szmalec (UCLouvain) and Kristel Van Goethem (UCLouvain). We thank Amélie Bulon, Isa Hendrikx and Morgane Simonis for their assistance in the data collection, Mrs Albarello and Morue of the Administration Générale à l'Enseignement for making demographic data available and the reviewers for their time, interest and constructive comments.

## References

Amengual-Pizarro, M., & Prieto-Arranz, J. I. (2014). Exploring affective factors in L3 learning: CLIL vs. Non-CLIL. In M. Juan-Garau, & J. Salazar- Noguera (Eds.), *Content-based Language Learning in multilingual educational environments* (pp. 197–220). Springer.

Baetens Beardsmore, H., & Kohls, J. (1988). Immediate pertinence in the acquisition of multilingual proficiency: The European Schools. *The Canadian Modern Language Review, 44*(2), 240–260.

Cenoz, J., Genessee, F., & Gorter, D. (2014). 2014. Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics, 35*(3), 243–262.

Chambers, G. N. (1999). *Motivating Language learnersMultilingual Matters*.

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283–292.

Chopey-Paquet, M. (2007). Belgium (French-Speaking). In A. Maljers, D. Marsh, & D. Wolff (Eds.), *Windows on CLIL. Content and Language integrated learning in the European spotlight* (pp. 28–32). European Centre for Modern Languages.

Coyle, D. (2011). ITALIC research report. *Investigating student gains: Content and Language integrated learning*. University of Aberdeen. https://www.abdn.ac.uk/italic/documents/ITALIC_Report_-_Complete_Version.pdf.
Coyle, D., Hood, P., & Marsh, D. (2010). *Clil: Content and Language integrated learning*. Cambridge University Press.
Dalton-Puffer, C. (2008). Outcomes and processes in content and Language integrated learning (CLIL): Current research from Europe. In W. Delanoy, & L. Volkmann (Eds.), *Future perspectives for English language teaching* (pp. 139–157). Carl Winter.
Dalton-Puffer, C., & Smit, U. (2013). Thinking allowed: Content and language integrated learning: A research agenda. *Language Teaching, 46*(4), 545–549.
De Smet, A., Mettewie, L., Galand, B., Hiligsmann, P., & Van Mensel, L. (2018). Classroom anxiety and enjoyment in CLIL and non-CLIL: Does the target language matter? *Studies in Second Language Learning and Teaching, 8*, 47–71. https://doi:10.14746/ssllt.2018.8.1.3.
De Smet, A., Mettewie, L., Hiligsmann, P., Galand, B., & Van Mensel, L. (2023). Does CLIL shape language attitudes and motivation? Interactions with target languages and instruction levels. *International Journal of Bilingual Education and Bilingualism, 26*(5), 534–553. https://doi:10.1080/13670050.2019.1671308.
Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching, 2*, 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5
Dewaele, J.-M., & Proietti Ergün, A. L. (2020). How different are the relations between enjoyment, anxiety, attitudes/motivation and course marks in pupils' Italian and English as Foreign languages? *Journal of the European Second Language Association, 4*(1), 45–57.
Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014). CLIL and motivation: The effect of individual and contextual variables. *Language Learning Journal, 42*(2), 209–224. https://doi:10.1080/09571736.2014.889508.
Dunn, L. M., & Dunn, L. M. (2005). *PPVT-III-NL: Peabody picture vocabulary test-III-NL*. Pearson Assessment and Information B.V.
Dunn, L. M., Dunn, D. M., & Assessments, P. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.
Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, Article 101859.
Eurydice. (2006). *Content and Language integrated learning (CLIL) at school in Europe*. Eurydice European Unit.
Eurydice. (2023). *Key data on teaching languages at school in Europe* (2023 edition). Publications Office of the European Union. https://data.europa.eu/doi/10.2797/529032.
Fehling, S. (2009). Lernprozesse und kognitive Entwicklung im bilingualen Unterricht: Bericht aus einer zweijährigen Longitudinalstudie. In D. Caspari, W. Hallet, A. Wegner, & W. Zydatiß (Eds.), *Bilingualer Unterricht macht Schule. Beiträge aus der Praxisforschung* (2nd ed., pp. 51–63) Peter Lang.
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.
Galand, B., & Hospel, V. (2015). Facteurs associés au risque de décrochage scolaire : vers une approche intégrative. *L'Orientation Scolaire et Professionnelle, 44*(3). https://doi.org/10.4000/osp.4604
Goriot, C., Van Hout, R., Broersma, M., Lobo, V., McQueen, J. M., & Unsworth, S. (2021). Using the peabody picture vocabulary test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism, 24*(4), 546–568. https://doi.org/10.1080/13670050.2018.1494131
Goris, J. A., Denessen, E. J. P. G., & Verhoeven, L. T. W. (2019). Effects of content and language integrated learning in Europe A systematic review of longitudinal experimental studies. *European Educational Research Journal, 18*(6), 675–698.
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
Hiligsmann, P., Van Mensel, L., Galand, B., Mettewie, L., Meunier, F., Szmalec, A., Van Goethem, K., Bulon, A., De Smet, A., Hendrikx, I., & Simonis, M. (2017). Assessing content and Language integrated learning (CLIL) in French-speaking Belgium: Linguistic, cognitive and educational perspectives. *Les Cahiers de Recherche du Girsef, 109*, 1–24.
Hofstadler, N., Talbot, K., Mercer, S., & Lämmerer, A. (2020). 2. The thrills and ills of content and Language integrated learning. In C. Gkonou, J. Dewaele, & J. King (Eds.), *The emotional rollercoaster of language teaching* (pp. 13–30). Multilingual Matters. https://doi.org/10.21832/9781788928342-006.
Hornstra, L., van der Veen, I., & Peetsma, T. (2016). Domain-specificity of motivation: A longitudinal study in upper primary school. *Learning and Individual Differences, 51*, 167–178.
Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign Language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132. https://doi.org/10.2307/327317
Housen, A. (2002). Processes and outcomes in the European schools. Model of multilingual education. *Bilingual Research Journal, 26*(1), 45–64.
Lasagabaster, D. (2009). The implementation of CLIL and attitudes towards trilingualism. *ITL - International Journal of Applied Linguistics, 157*, 23–43. https://doi.org/10.2143/ITL.157.0.2042586
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:10.1080/17501229.2010.519030.
Lasagabaster, D., & Doiz, A. (2017). A longitudinal study on the impact of CLIL on affective factors. *Applied Linguistics, 38*, 688–712. https://doi:10.1093/applin/amv059.
Mahmoodi, M. H., & Yousefi, M. (2022). Second language motivation research 2010-2019: A synthetic exploration. *Language Learning Journal, 50*(3), 273–296. https://doi:10.1080/09571736.2020.1869809.
Maljers, A., Marsh, D., & Wolff, D. (2007). *Windows on CLIL: Content and language integrated learning in the European spotlight*. European Platform for Dutch Education.
Marsh, D. (2000). *Using languages to learn and learning to use languages: An introduction to content and language integrated learning for parents and young people*. TIE-CLIL.
Marsh, D., Maljers, A., & Hartiala, A. K. (2001). Profiling European CLIL classrooms. *Languages open doors*. University of Jyväskylä.
Mearns, T., de Graaff, R., & Coyle, D. (2020). Motivation for or from bilingual education? A comparative study of learner views in The Netherlands. *International Journal of Bilingual Education and Bilingualism, 23*(6), 724–737. https://doi:10.1080/13670050.2017.1405906.
Mettewie, L., & Lorette, P. (2014). Tweetalig onderwijs als Belgische "philtre d'amour". Attitudes van Franstalige CLIL- leerders van het Nederlands. In L. Degand, P. Hiligsmann, L. Rasier, M. Sergier, S. Vanasten, & K. Van Goethem (Eds.), *In het teken van identiteit: Taal en cultuur van de Nederlanden* (pp. 195–201). Presses Universitaires de Louvain.
Mettewie, L., & Van Mensel, L. (2023). Understanding foreign language education and bilingual education in Belgium: A (surreal) piece of cake. *International Journal of Bilingual Education and Bilingualism, 26*(5), 639–657. https://doi.org/10.1080/13670050.2020.1768211
Möller, V. (2021). Promoting bilingualism at the primary and secondary level: The role of intelligence, motivation and anxiety. *International Journal of Bilingual Education and Bilingualism, 24*(9), 1306–1323. https://doi.org/10.1080/13670050.2018.1559795
Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A meta-analysis of relations between achievement and self-concept. *Review of Educational Research, 90*(3), 376–419.
Navarro-Villarroel, C. (2011). *Young students' attitudes toward languages [doctoral dissertation, Iowa state university]. Dr.lib.iastate.edu*. https://dr.lib.iastate.edu/handle/20.500.12876/26182.
Ohlberger, S., & Wegner, C. (2019). CLIL modules and their affective impact on students with high English anxiety and low self-efficacy. *Apple: Applied Language Studies, 13*(3), 1–15.
Papaja, K., & Rojczyk, A. (2013). Motivation from the perspective of a CLIL teenage learner. In D. Gabrys-Barker, & J. Bielska (Eds.), *The affective dimension in second language acquisition* (pp. 241–253). Multilingual Matters.
Pirskanen, H. (2009). *Attitudes towards the English language: A comparison between a CLIL class and a regular class*. University of Jyväskylä. https://jyx.jyu.fi/dspace/handle/123456789/20134.
Pladevall-Ballester, E. (2019). A longitudinal study of primary school EFL learning motivation in CLIL and non-CLIL settings. *Language Teaching Research, 23*(6), 765–786.
Pokormy, S. B., Jason, L. A., Schoeny, M. E., Townsend, S. M., & Curie, C. J. (2001). Do participation rates change when active consent procedures replace passive consent. *Evaluation Review, 25*(5), 567–580. https://doi.org/10.1177/0193841x0102500504

Raven, J. C., Court, J. H., & Raven, J. (1998). *Progressive coloured Matrices*. Oxford Psychologists Press.
Rumlich, D. (2016). *Evaluating bilingual education in Germany*. Peter Lang.
San Isidro, X., & Lasagabaster, D. (2022). Students' and families' attitudes and motivations to language learning and CLIL: A longitudinal study. *Language Learning Journal, 50*(1), 119–134. https://doi:10.1080/09571736.2020.1724185.
Seikkula-Leino, J. (2007). CLIL learning: Achievement levels and affective factors. *Language and Education, 21*(4), 328–341. https://doi.org/10.2167/le635.0
Investigating content and language integrated learning. Insights from Swedish High School. In Sylvén, L. K. (Ed.), *Multilingual Matters*, (2019).
Talbot, K. R., Gruber, M. T., & Nishida, R. (2021). The psychological experience of integrating content and language. In *Multilingual Matters*.
Thompson, A. S., & Sylvén, L. K. (2015). Does English make you nervous? Anxiety profiles of CLIL and non-CLIL students in Sweden. *Apple: Journal of Applied Language Studies, 9*(2), 1–23. https://doi.org/10.17011/apples/urn.201512093950
Thompson, A. S., & Sylvén, L. K. (2019). CLIL and motivation revisited: A longitudinal perspective. In L. K. Sylvén (Ed.), *Investigating content and Language integrated learning. Insights from Swedish high school* (pp. 76–95). Multilingual Matters.
Toste, J. R., Didion, L., Peng, P., Filderman, M. J., & McClelland, A. M. (2020). A meta-analytic review of the relations between motivation and reading achievement for K–12 students. *Review of Educational Research, 90*(3), 420–456.
Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26*(2), 265–283.
Van Mensel, L., & Galand, B. (2022). Testing the predictive power of executive functions, motivation, and input on second language vocabulary acquisition: A prospective study. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-022-00597-x
Van Mensel, L., Hiligsmann, P., Mettewie, L., & Galand, B. (2020). CLIL, an elitist Language Learning approach? A background analysis of English and Dutch CLIL pupils in French-speaking Belgium. *Language Culture and Curriculum, 33*(1), 1–14. https://doi:10.1080/07908318.2019.1571078.
Wigfield, A., & Cambria, J. (2010). Expectancy-value theory: Retrospective and prospective. In T. C. Urdan, & S. A. Karabenick (Eds.), *Advances in motivation and achievement* (vol. 16, pp. 35–70). Emerald Group Publishing Limited. https://doi.org/10.1108/S0749-7423, 2010)000016A005.
Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi:10.1006/ceps.1999.1015.
Williams, M., Burden, R., & Lanvers, U. (2002). French is the language of love and stuff: Student perceptions of issues related to motivation in learning a Foreign Language. *British Educational Research Journal, 28*(4), 503–528. https://doi:10.1080/0141192022000005805.