Effects of input on L2 writing in English and Dutch: CLIL and non-CLIL learners in French-speaking Belgium

Luk Van Mensel[1], Amélie Bulon[2], Isa Hendrikx[2], Fanny Meunier[2], Kristel Van Goethem[2]
[1]Université de Namur & [2] Université catholique de Louvain

Abstract

As part of a project on Content and Language Integrated Learning (CLIL) in French-speaking Belgium, this study aims to explore the impact of formal and informal input on learners' variability in writing, and to compare two target-language conditions (Dutch and English) in CLIL and non-CLIL settings in French-speaking Belgium. A regression model shows that CLIL is a significant predictor of L2 outcomes for both target languages, but that the relative impact of formal and informal input differs depending on the target language. In short, the amount of formal language exposure predicts the outcomes of the English learners' written productions, and the frequency of informal exposure those of the Dutch learners. We argue that this observation is likely related to the difference in status that each of these languages holds among the pupils in our sample. The findings thus highlight the importance of the L2 status in research on CLIL, since different L2s can yield different results.

Key words

English L2 – Dutch L2 – French L1 – CLIL – input – writing skills – French-speaking Belgium

Acknowledgements:

This work was supported by a Concerted Research Action grant (ARC 14/19-061) from the Université catholique de Louvain (UCL) and the Université de Namur (UNamur), awarded to Philippe Hiligsmann (spokesperson; UCL), Benoît Galand (UCL), Laurence Mettewie (UNamur), Fanny Meunier (UCL), Arnaud Szmalec (UCL) and Kristel Van Goethem (UCL). We would like to thank Audrey De Smet and Morgane Simonis for their assistance in the data collection. We also warmly

thank the schools and participants that took part in the study.

1. Introduction

The educational method of Content and Language Integrated Learning (CLIL) aims to improve pupils' L2 skills by teaching a content-based subject in the target language (Dalton-Puffer, 2011). A large body of recent research reports significantly higher L2 competences among pupils in CLIL than those who are learning the L2 in traditional foreign language learning settings (e.g. Admiraal, Westhoff & de Bot, 2006; Lasagabaster, 2008; Lorenzo et al., 2005; Ruiz de Zarobe, 2008, 2010; Zydatiß, 2007; Wesche, 2002). In particular, pupils in CLIL appear to use a more diverse and more complex vocabulary, adapted to the context (Jexenflicker & Dalton-Puffer, 2010; Lo & Murphy, 2010; Ruiz de Zarobe, 2010; Zydatiß, 2007). Also, CLIL pupils are more fluent and more confident in speaking in the target language (Dalton-Puffer et al., 2008), and show better communication skills (e.g. Wode, 1994; Klieme, 2006) in comparison to their peers in traditional foreign language learning settings. This said, whilst time spent on learning a second/foreign language has been recognized as one of the most important factors for successful acquisition/learning of an additional language (Kinsella, 2009; Muñoz, 2011), few studies on CLIL have explicitly controlled for the possible effect of L2 exposure (Saladrigues & Llanes, 2014). The amount and quality of target language (TL) input has become a focus of interest for many SLA researchers (e.g. Kinsella, 2009, and Moyer, 2009, on formal and informal contact; Llanes & Muñoz, 2009, and Pérez-Vidal & Juan-Garau, 2011, on study abroad; Long, 1983, on formal instruction). Saladrigues & Llanes (2014), for example, is one of the few CLIL studies where learners were grouped both depending on the program they belonged to (CLIL or non-CLIL) and on the amount of L2 exposure they received (high or low, according to the number of hours taking curricular and extracurricular classes). The results of this study showed no significant difference between CLIL and non-CLIL learners in terms of written fluency, accuracy, lexical complexity and syntactic complexity. However, the high versus low L2 exposure grouping revealed statistically significant higher gains for the former group.

Considering Saladrigues & Llanes' (2014) results, the present study aims to disentangle the impact of CLIL from other types of target-language input. Based on data collected in a multidisciplinary research project on CLIL in French-speaking

Belgium (Hiligsmann et al., 2017), we wish to examine to what extent the writing of secondary school pupils is influenced by extra target language input through CLIL and/or through informal extra-curricular contact. As our study addresses two target languages (English and Dutch), it also enables us to adopt a contrastive approach. Most research on CLIL typically considers only one target language (mainly English), and by looking at two languages we wish to shed light on the impact of different target language conditions on the potential benefits of the CLIL approach. This contrastive approach is all the more interesting in the context of French-speaking Belgium (*Communauté française*), in which English and Dutch have a different target-language status: Dutch is one of the three official languages in Belgium (besides French and German), whereas English is a foreign language. However, students in the French community of Belgium generally encounter Dutch infrequently in day-to-day life. Dutch is spoken in the Flemish community, and the different communities have relatively much political autonomy – the education systems being organized at communal level instead of national level. Consequently, the students in the French community may be exposed as little to Dutch as to English, or may even be exposed more often to English (as a *lingua franca*) than to Dutch.

The following section provides an overview of the research on input exposure and its impact on language learning, focusing more specifically on writing and on the learning context of CLIL. Section 3 formulates the specific research questions, and describes the participants and data set, as well as the input and outcome measures that we selected. In the fourth section we present the results drawn from our analyses, which are then discussed in section 5. In the final section we wrap up the discussion with some concluding remarks.

2. The influence of TL input on TL proficiency

One of the important debates in research on second and foreign language acquisition concerns the relative influence of starting age, length of TL exposure, and current TL contact on the learners' (ultimate) attainment. Earlier research suggested that early starters reach higher levels of proficiency than late starters in second language learning environments (e.g. Patkowski, 1980; Johnstone & Newport, 1989). In a similar vein, Carroll (1969) has often been quoted as a proponent of the idea that time is the most important variable in the acquisition of an L2. In this view, the amount of competence one achieves is largely a matter of time spent in learning. However,

according to Muñoz (2011), this cannot be applied to foreign language learning settings, where input never ceases to be a determinant factor. In fact, research has indicated that late starters can eventually catch up on younger learners (e.g. Cenoz, 2002, 2003; García Lecumberri & Gallardo, 2003; Miralpeix, 2006; Muñoz, 2006). In other words, starting age does not necessarily seem to be a significant factor of ultimate higher proficiency in foreign language learning contexts, as recently confirmed in two studies conducted by Muñoz (2008, 2011). What appears to be an important indicator of language proficiency outcomes in these contexts is exposure time (Lambelet & Berthele, 2015). Also, studies investigating the impact of the age factor in instructional settings acknowledge the relevance of intensity of exposure in foreign language acquisition (Torras & Celaya, 2001).

Similar observations have been made with regard to writing competence more specifically. As learners become more proficient, they write more fluently, more accurately and produce more grammatically and lexically complex texts (Wolfe-Quintero et al., 1998). Yet, growing written proficiency and duration of TL learning are not necessarily linearly correlated. Indeed, some studies looking at foreign language written production showed that an earlier start does not seem to be beneficial in written production development (Burstall, 1975; Torras & Celaya, 2001; Celaya, Torras & Perez-Vidal, 2001; Torras et al., 2006; Celaya & Navés, 2009). However, in an investigation of the link between L2 competence and written production, Cumming (1994, as cited in Lasagabaster & Doiz, 2003, p. 141) concluded that time of exposure when learning a L2 is "a decisive factor in the level of competence attained". Myles (2002, pp. 9-13) also stressed the importance of input in second language writing: "Input and interaction also play important roles in the writing process, especially in classroom settings. […] Instruction should provide students with ample amounts of language input and instruction, as well as writing experience […]".

Target language exposure can be divided into two main categories. A first type of exposure refers to the amount of *formal* instruction the learners receive (often measured in number of years or number of curricular and possibly extracurricular hours). Johnstone (2007), for instance, in a review of findings regarding language education in Scotland, states that length of instruction appears to be positively correlated with successful learning. A second type of exposure refers to the frequency

of *informal* contact with the target language. Contact with native speakers clearly has a positive impact on language learning, and even more so in informal than in formal contexts (Kinsella, 2009; Moyer, 2009). Llanes and Muñoz (2009) and Pérez-Vidal and Juan-Garau (2011), for instance, studied the linguistic gains provided by a study abroad experience and found a positive influence on different aspects of language (fluency, accuracy and listening comprehension). Research by Muñoz (2011) compared the impact of various input measures and revealed that measures of both recent and current exposure to language input (at home and abroad) as well as length of instructed exposure correlated significantly with proficiency scores. In an earlier study where she focused on learners' oral performance (Muñoz, 2008), it was found that the number of years of instruction and current informal contact significantly predicted syntactic complexity, while the number of years abroad was the best predictor of lexical diversity and accuracy. Regarding fluency, the number of years abroad and current informal contact were the best predictors. As for Mitchell et al. (2017), they explored L2 development before, during and after a temporary sojourn abroad, and its relationship with sojourners' personal development, social experience and language practices while abroad. The study focused on British undergraduates learning French or Spanish as a TL. The study offered a better understanding of informal language learning and of the complex triangular relationship between identity, personal and L2 development. All these findings suggest that different input variables affect the various components of proficiency in different ways.

If the distinction between formal and informal input has turned out to be a very fruitful one in traditional foreign language learning situations, the specific context of CLIL education can be regarded as yet another and somewhat different type of input. CLIL environments are believed to facilitate language learning since they offer more naturalistic and input-rich environments than foreign language classrooms. A substantial amount of research in CLIL has focused on learning outcomes, usually comparing the language proficiency of CLIL learners with the one of learners in traditional foreign language classes (non-CLIL learners). It is mostly hypothesized that pupils in CLIL programs will outperform their peers since they benefit from more exposure time to the L2, and this has been shown to be the case for various language aspects (notably oral production and vocabulary) (e.g. Dalton-Puffer, 2008; Jexenflicker & Dalton-Puffer, 2010; Lo & Murphy, 2010; Ruiz de Zarobe, 2010;

Zydatiß, 2007). As far as writing skills are concerned, similar positive results have been found. Lasagabaster (2008), for instance, showed that CLIL had a positive impact on various aspects of writing (content, textual organization, vocabulary, grammar and spelling). Jexenflicker & Dalton-Puffer (2010) identified highly significant differences between the writing of CLIL and non-CLIL pupils for a range of measures of grammar and syntactic complexity, except for the number of subordinate clauses. Whittaker & Llinares (2011) found that CLIL pupils' writing coherence and their choice for the appropriate register improved over time. Gené-Gil et al. (2015) reported significant differences in the development of written complexity, accuracy and fluency of CLIL learners over a 3-year period (and only in accuracy for non-CLIL learners). Lahuerta Martínez (2015) investigated the writing of learners following bilingual and non-bilingual programs and noticed that the bilingual group surpassed the non-bilingual group in all the fluency, accuracy and lexical complexity measures. Pérez-Vidal & Roquet (2015) identified larger gains in accuracy, syntactic and lexical complexity in the writing of learners who received extra CLIL hours. Isidro & Lasagabaster (2018) found that both CLIL and non-CLIL students improved their competence in L2 English after two years target language instruction, with significantly greater progress found in the CLIL group. In addition, the CLIL students also outperformed their non-CLIL peers in both Spanish and Galician. In contrast, Roquet & Pérez-Vidal (2017) could not confirm advantages for the CLIL students. They improved their written productive abilities, but only accuracy improved significantly. To our knowledge, and with the exception of the study by Saladrigues & Llanes (2014) mentioned in the introduction, there are no studies that investigated the impact of other exposure/input measures on written proficiency in a CLIL context.

3. Research questions, participants and data collection

The first purpose of this study is to assess the relative impact of different types of input variables on learners' writing performance. Given the findings in current literature on CLIL, we expect the input provided by the CLIL experience to be an important predictor of writing skills. Incidentally, previous comparative analyses on the dataset used in the present contribution (Bulon et al., 2017) yielded results that support this hypothesis: based on a set of lexical and morpho-syntactic parameters, Bulon et al. (2017) found that the writing skills of the CLIL learners were globally

more advanced for both target languages (Dutch and English), while no significant differences were found for their mother language (French) (see also Section 4.1). However, these results were not compared with native control groups and hence might need to be refined. Moreover, besides CLIL, we wish to include formal and informal target language exposure, since both types of input appear to affect learners' (oral) proficiency in different ways (Muñoz, 2008, 2014).

Second, we adopt a contrastive approach (comparing Dutch and English as L2s) in order to examine whether the selected input variables influence CLIL and non-CLIL learners' writing skills in the same way depending on the target language.

More specifically, we will address the following three research questions:

> i) Compared to a control group of native speakers, do students in CLIL display a more 'native-like' writing (in terms of text length, sentence length, word length and lexical diversity) than non-CLIL students? If so, does this more native-like writing performance manifest itself to the same extent for both target languages?
>
> ii) To which extent does informal contact with the target language and/or the duration of target language learning have an impact on the writing performance?
>
> iii) Are the effects of (formal and informal) input similar for both target languages (English and Dutch), or can we observe any significant differences which may be related to their different status in French-speaking Belgium?

3.1 Participants

The participants involved in this study are 438 5th year French-speaking secondary school learners of Dutch and English from nine secondary schools in Wallonia (French-speaking Belgium). The participating schools have contrasted profiles, notably in terms of location (all provinces are covered), socio-economic level, and education authority (official education and publicly subsidized schools) (see also Hiligsmann et al., 2017; Van Mensel et al., *forthc.*). These schools provide CLIL programs in Dutch and/or English, along with traditional instruction (French-medium instruction with foreign language classes): 229 pupils learn English (96) or Dutch (133) in a CLIL setting (52.28%), while 209 learn English (97) or Dutch (112) in traditional settings (47.72%). The participants' ages range from 15 to 18 and their

mean age is 16.5. 207 (46.7%) of the learners are male and 231 (53.3%) female.

As regards the L1-speaker control groups, the Dutch data were collected from 59 5th year Dutch-speaking secondary school students from the Netherlands and Flanders (Dutch-speaking Belgium), of whom 11 (18.64%) were male, and 48 (81.36%) female (average age: 16.7 years old). The data for English were gathered from 65 English L1 speakers from the US (Florida), of whom 11 (16.92%) were male, and 54 (83.08%) female; their average age was slightly higher: 19.4 years old.

3.2 Data collection procedure and instruments

The learners spent one day at the Université catholique de Louvain (in Louvain-la-Neuve, Belgium) between October and November 2015 to perform a variety of computer-administered tasks, including two writing exercises. The writing tasks consisted in writing an e-mail to a friend (min. 15 lines) on two possible topics which were randomly assigned, either their last holidays or a party they attended. The task was timed (max. 25 minutes per e-mail) and we made sure the pupils had no access to online dictionaries or other reference tools. The same writing tasks were collected from the L1 reference groups in similar conditions between November 2015 and January 2016 for the Dutch-speaking group, and one year later for the English-speaking group. A few texts were lost due to technical problems, but as Table 1 shows, we were able to collect a total of 412 learner and 130 L1 productions. [1]

Table 1. Number of texts and number of words collected

	Dutch L1	Dutch CLIL	Dutch non-CLIL	English L1	English CLIL	English non-CLIL
Texts	61	132	100	69	90	90
Words	16 262	37 209	19 399	23 016	29 394	23 747

Since we collected written productions in two different languages, we had to select different appropriate tools to analyze the texts. For the English texts, we used Coh-Metrix (Graesser, McNamara, Louwerse & Cai, 2004, p. 93), for the Dutch texts T-Scan (Pander Maat et al., 2014). As both computational tools offer a large number of measures, some being similar for the two languages and others specific to one language only, a selection of shared indices had to be performed (see also Bulon et

al., 2017). In order to get an overall evaluation of the pupils' writing, we selected the following four measures: (1) text length, measured through the number of words per text, as a proxy of fluency; (2) sentence length, measured through the number of words per sentence, as a proxy of syntactic complexity [2]; (3) word length, measured through the number of syllables per word, as a proxy of morpho-syntactic complexity; and (4) MTLD (Measure of Textual Linguistic Diversity)[3], indicating lexical diversity. Whilst we are well aware that such measures offer only a partial view on writing skills (and that other aspects should be taken into account to obtain a refined view of the learners' proficiency), these measures have been widely used in assessing complexity and fluency in learners' productions (see Mitchell et al., 2017, for more examples and comments). They allowed us to get proxies of proficiency for all our learners and for the two target languages.

The input measures (besides CLIL) used to investigate the potential impact of L2 input on the learners' written proficiency are based on Muñoz's work (2011, 2014) and are derived from extensive questionnaires: (1) length of TL instruction in years, a measure of cumulative exposure to formal input, and (2) current informal contact with the TL, a composite measure consisting of frequency of internet use in the TL, frequency of TL (productive and receptive) use outside school and frequency of contact with native speakers outside school (see Appendix 1). We also included the pupils' nonverbal intelligence (Raven test-score, see Raven, Court & Raven, 1998) as a control variable in our analysis. As discussed in Bulon et al. (2017), among the pupils learning Dutch as a second language, the CLIL and the non-CLIL pupils differed significantly in their Raven scores, whilst this was not the case for the English learners (see also Simonis et al., *submitted*). These differences may be due to an (auto-)selection effect for Dutch CLIL (see Van Mensel et al, *forthc.*, for details). In order to preclude any effects related to a bias in the sample, we thus deemed it wise to add Raven scores as a control variable.

4. Results

4.1 Descriptives

In Tables 2a and 2b, we provide an overview of the descriptive statistics for the two target languages and for all selected variables, broken down according to the two target language conditions.

Table 2a. Descriptive statistics English

	English					
	L1-speakers		*CLIL*		*Non-CLIL*	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Outcome variables Target Language						
Text length (words/text)	333.57	87.99	326.52	63.81	263.86	83.35
Sentence length (words/sentence)	23.39	6.51	15.48	5.93	12.81	5.39
Word length (syllables/word)	1.33	0.06	1.29	.05	1.31	.05
Lexical diversity (MTLD)	95.53	22.07	72.99	16.41	62.07	14.27
Input variables						
Length of target language instruction (in years)	-	-	6.23	3.00	5.05	2.24
Current informal contact with TL[a]	-	-	3.42	.81	3.02	.87
Control variable						
Raven score	-	-	43.67	6.21	41.82	7.78

[a] Frequency of informal contact with the target language outside school (composite measure, Cronbach's alpha .78, see Appendix 1), on a scale from 1 to 5 (never – rarely – sometimes – often – very often)

Table 2b. Descriptive statistics Dutch

			Dutch			
	L1 speakers		*CLIL*		*Non-CLIL*	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Outcome variables Target Language						
Text length (words/text)	222.18	67.298	280.95	57.79	193.99	59.76
Sentence length (words/sentence)	12.49	4.09	11.27	3.13	9.40	2.48

Word length (syllables/word)	1.27	.24	1.28	.05	1.23	.06
Lexical diversity (MTLD)	94.41	26.22	74.05	16.09	61.04	17.99
Input variables						
Length of target language instruction (in years)	-	-	9.20	2.67	7.73	2.48
Current informal contact with TL[a]	-	-	2.54	.78	2.05	.73
Control variable						
Raven score	-	-	46.04	6.90	42.41	7.97

[a] Frequency of informal contact with the target language outside school (composite measure, Cronbach's alpha .78, see Appendix 1), on a scale from 1 to 5 (never – rarely – sometimes – often – very often)

Tables 2a and 2b show that, overall, the CLIL learners of both English and Dutch display more native-like scores than the non-CLIL learners for the examined parameters. Both CLIL groups wrote substantially longer texts than the non-CLIL pupils. In the case of English, the CLIL texts were in this sense very similar to those produced by the control group, whereas for Dutch the CLIL pupils produced even longer texts than the L1 speakers. Next, both the scores for sentence length (number of words per sentence) and lexical diversity (MTLD) are higher and more native-like in the English and the Dutch texts written by CLIL pupils than in the texts written by non-CLIL pupils, suggesting a more advanced syntactic competence and a richer vocabulary. Word length (syllables per word) is more native-like in the Dutch texts written by CLIL pupils than in the texts written by non-CLIL pupils (on average the CLIL pupils produce even slightly longer Dutch words than the natives), but this tendency is not present in the English texts. Previous comparative analyses on these outcomes only (Bulon et al., 2017) indicated that the differences between the CLIL and non-CLIL learner groups are statistically significant, but the effect size appeared to be larger for Dutch than for English. Hence, while there is a more clear-cut difference in terms of writing skills between CLIL and non-CLIL learners of Dutch, this seems to be less the case for learners of English.

We can equally observe in Table 2b that the length of target language instruction in years is higher among the pupils learning Dutch than among those learning English in

our sample (Table 2a). Among the learners of both languages, the CLIL pupils received more years of formal target language exposure. By contrast, the current informal contact with the target language is higher among the pupils learning English than among those learning Dutch. Also, the CLIL pupils of both languages had more current informal contact with the target language than the pupils who are not in CLIL. In the following section we analyze if and how the differences observed in the written production measures, duration of TL teaching, informal TL contact and Raven scores correlate. To do so, we first calculate correlation scores; in a second step we fit the measures in a regression model.

4.2 Correlations

In order to compare the strength of association between the selected measures for written production on the one hand and the independent and control variables on the other hand, we ran Pearson correlations for each of the target languages separately. Table 3 shows the correlations (Pearson r coefficients) between the outcome measures and the independent variables – CLIL, Raven score (nonverbal intelligence), length of target language exposure, and current informal contact with the target language – for the English texts.

Table 3. Correlations for English

	Text length (words/text)	Sentence Length (words/sentence)	Word length (syllables/word)	Lexical Diversity (MTLD)
CLIL (n=180)	.391**	.231**	-.149*	.336**
Raven Score (n=178)	.106	-.080	-.128	.067
Length of TL instruction in years (n=170)	.296**	.144	.152*	.360**
Current informal contact with TL (n=176)	.236**	.128	-.036	.161*

* $p < .05$ (two-tailed)

** $p < .01$ (two-tailed)

CLIL significantly correlates with all the written production measures, something we had hypothesized given the results of our previous analyses discussed above (Bulon et al., 2017). In contrast, the Raven score does not correlate with any of the outcome

variables, suggesting a limited impact – if any – of the pupils' non-verbal intelligence on their written production. Length of target language instruction correlates with text length, word length[4] and lexical diversity, while sentence length (number of words per sentence) does not expand with number of years of exposure to the target language. Finally, current informal contact correlates significantly with two out of four measures, namely words per text and lexical diversity (MTLD). Taken together, the correlations presented here lead us to suspect that for written English production, besides the (expected) impact of CLIL, the two proposed input measures impact text length (fluency) and lexical diversity, while differences at a more granular level of writing (word and sentence length) seem less affected.

Table 4 shows the correlations (Pearson r coefficients) between the outcome measures and the independent variables – CLIL, Raven score, length of target language exposure, and current informal contact with the target language – for the Dutch texts.

Table 4. Correlations for Dutch

	Text length (words/text)	Sentence length (words/sentence)	Word length (syllables/word)	Lexical Diversity (MTLD)
CLIL (n=232)	.594**	.309**	.367**	.357**
Raven Score (n=227)	.159*	-.050	.293**	.268**
Length of TL instruction in years (n=219)	.264**	.150*	.088	.232**
Current informal contact with TL (n=222)	.333**	.136*	.127	.228**

* $p < .05$ (two-tailed)

** $p < .01$ (two-tailed)

Here again, and as hypothesized, the CLIL variable significantly correlates with all four measures of written production. In contrast to the English texts, the pupils' non-verbal intelligence does also correlate with two of the outcome measures, namely word length and lexical diversity, an observation which can be explained by the reported significant differences in Raven scores between the pupils in CLIL and non-CLIL contexts learning Dutch (Simonis et al., *submitted*; Van Mensel et al, *forthc.*), differences that were not found for the English target language learners. Length of target language exposure correlates with text length, sentence length and lexical

diversity. Word length (number of syllables per words) does not increase with the number of years of exposure to the target language. Finally, current informal contact with the TL (Dutch) correlates with three out of four of the measures, namely text length, sentence length and lexical diversity (MTLD). Although the correlation coefficients and the strength of the correlations for these Dutch productions differ somewhat from those presented above for English, the overall picture provided shows some similarities. Putting aside the differences with respect to the pupils' non-verbal intelligence scores, the input measures consistently correlate with text length and lexical diversity for both target languages, whereas sentence length and word length are less affected.

In order to gauge the relative effects of the different input measures on the pupils' written production, we conducted a set of multiple regression analyses whose results are presented in the following paragraphs. All assumptions were met and there was no collinearity between the independent variables.

4.3 Regressions

Table 5 shows the outcomes of the regression model with the CLIL variable, the Raven score, number of years of TL instruction, and the factor of 'input/use' outside school for the English texts. The CLIL variable appears consistently as a significant predictor for all four measured dependent variables, albeit negatively for the number of syllables per word (cf. also the negative correlation in Table 3 above). The pupils' non-verbal intelligence as measured by the Raven test, on the contrary, does not appear to have any impact on any of the production measures for the English texts. Regarding the two input variables, we can observe how the length of target language instruction significantly predicts text length, word length and lexical diversity, whereas current informal contact with English outside school does not predict the pupils' writing skills in a significant way. Interestingly, the model is substantially better at explaining the variance in the pupils' scores for text length and lexical diversity (around 20%) than for the other two measures (slightly more than 7%), which echoes the correlational tendencies presented above.

Table 5. Regression models for the English written productions. (*ns = not significant)*

Measures	**Regression coefficient and t-test statistic**	**Coefficient of determination[5] and F-**

		ratio
Text length	CLIL (ß = .320, *t* = 4.39, p < .001) Raven - *ns* Length of TL instruction (ß = .206, *t* = 2.91, p < .01) Current informal contact with TL - *ns*	(R^2 = .219, F(4,164) = 11.475, p < .001)
Sentence length	CLIL (ß = .210, *t* = 2.64, p < .01) Raven - *ns* Length of TL instruction - *ns* Current informal contact with TL - *ns*	(R^2 = .073, F(4,164) = 3.223, p < .05)
Word length	CLIL (ß = -.167, *t* = -2.11, p < .05) Raven – *ns* Length of TL instruction (ß = .202, *t* = 2.62, p < .05) Current informal contact with TL - *ns*	(R^2 = .075, F(4,164) = 3.311, p < .05)
Lexical diversity	CLIL (ß = .246, *t* = 3.341, p < .01) Raven – *ns* Length of TL instruction (ß = .301, *t* = 4.192, p < .001) Current informal contact with TL - *ns*	(R^2 = .202, F(4,164) = 10.352, p < .001)

The results of the same regression analyses for the Dutch texts provide a somewhat different picture, as can be seen in Table 6. Again, CLIL remains a significant predictor for all selected measures, but also the pupils' nonverbal intelligence shows a significant relationship with three out of four measures: word length, sentence length, and lexical diversity. Furthermore, current informal contact with the target language (Dutch) is significantly related to text length and lexical diversity, whereas length of TL instruction does not predict any of the measures for written production in a significant way, an observation which is the inverse of the outcomes for the English texts. Also, the percentage of variance that the model predicts appears to be higher for the written productions in Dutch when compared to those in English. With the exception of the lexical diversity variable, the percentages are considerably higher, rising up to nearly 40% for text length.

Table 6. Regression models for the Dutch written productions. (*ns = not significant)*

Measures	**Regression coefficient and t-test statistic**	**Coefficient of**

		determination and F-ratio
Text length	CLIL (ß = .513, *t* = 8.43, p < .001) Raven - *ns* Length of TL instruction - *ns* Current informal contact with TL (ß = .164, *t* = 2.79, p < .01)	(R^2 = .382, F(4,204) = 31.544, p < .001)
Sentence length	CLIL (ß = .306, *t* = 4.20, p < .001) Raven (ß = -.142, *t* = -2.08, p < .05) Length of TL instruction - *ns* Current informal contact with TL - *ns*	(R^2 = .113, F(4,204) = 6.510, p < .001)
Word length	CLIL (ß = .317, *t* = 4.52, p < .001) Raven (ß = .216, *t* = 3.30, p < .01) Length of TL instruction - *ns* Current informal contact with TL - *ns*	(R^2 = .184, F(4,204) = 11.522, p < .001)
Lexical diversity	CLIL (ß = .228, *t* = 3.28, p < .01) Raven (ß = .206, *t* = 3.17, p < .01) Length of TL instruction - *ns* Current informal contact with TL (ß = .138, *t* = 2.06, p < .05)	(R^2 = .196, F(4,204) = 12.440, p < .001)

5. Discussion

From the analyses presented above, a number of observations can be made. First, it should be noted that the significant correlations between the independent variables and the outcome measures are overall rather modest, indicating the importance of other variables that contribute to explaining variability in learners' written proficiency. Individual variables, such as language learning motivation (see De Smet et al., *in preparation*) or aptitude, are good candidates, as are classroom practices such as quantity and quality of language input, type of feedback, teaching style, etc. This finding is as such not surprising (Muñoz, 2014; Mitchell et al., 2017). Interestingly, however, the results from the regression analyses show us that the percentage of variance which is unaccounted for differs according to the target language condition. With the exception of lexical diversity, the present model appears to be better at predicting the outcome measures for the Dutch written productions than for the English ones, thereby suggesting that the variables unaccounted for play a larger role

in explaining learners' variability for the latter. Also, the results from both the correlations and regression analyses suggest a larger impact of CLIL and the input measures on text length (fluency) and lexical diversity, whereas word and sentence length are less affected. These last two measures would thus appear to hinge less on the amount of input the pupils receive (be it formally or informally), but rather on other factors that were not included in the present model, such as perhaps the type of - and methods and tools used in - formal (writing) instruction.

In terms of the relative importance of the predicting variables, we can confidently state that CLIL is an important predictor of the learners' written productions. The CLIL variable correlates significantly with all four outcome measures in both target languages, and is often the strongest predictor in the regression analyses. These findings confirm our previous analyses (Bulon et al., 2017) and make a strong case for the CLIL approach as enhancing L2 writing, particularly with respect to fluency and lexical diversity. In this sense, our findings are in line with similar results found internationally (Dalton-Puffer, 2008; Gené-Gil et al., 2015; Jexenflicker & Dalton-Puffer, 2010; Lo & Murphy, 2010; Lahuerta Martínez, 2015; Pérez-Vidal & Roquet, 2015; Ruiz de Zarobe, 2010; Zydatiß, 2007).

Besides CLIL, we wanted to look at the relative impact of two different input measures, viz. the amount of formal target language instruction (measured in number of years), and the degree to which the pupils are in informal contact with the target language outside school. Interestingly, the findings were very different for each of the target language conditions, with the results for the English texts showing the reverse picture of those obtained for the Dutch texts. Whereas the number of years of target language instruction that the learners received significantly predicts three of the four outcome measures for English, this variable was never significant for Dutch. Conversely, current informal contact with the target language is a significant predictor of text length and lexical diversity in the Dutch texts but does not appear to affect the pupils' writing skills in English. At first sight, these are rather puzzling results, since one would expect a somewhat similar picture for both target language learning conditions, also given the fact that we found significant correlations between these input variables and the outcome measures text length and lexical diversity for both Dutch (Table 3) and English (Table 4).

We would like to argue that a possible explanation could be found in the different status that English and Dutch enjoy as L2 languages in French-speaking Belgium.[6] As in other contexts, English in Belgium is regarded as the most important international language and is omnipresent in daily life (music, social media, etc.). Consequently, most of the pupils learning English have, at least to some extent, extracurricular contact with the language. On average, and as shown in Table 2a, the English learners are at least 'sometimes' in contact with English outside school. Dutch, on the other hand, being the (non-international) language of the Flemish language community in Belgium, is on average less frequently used in informal contexts (e.g. on the internet), as can be seen from Table 2b. Therefore, the discriminatory power of the 'current informal contact' variable may be greater for Dutch – those pupils that are in contact with Dutch outside the classroom experience significant gains from this extra input – but neutralized for English since all English learners are to some extent in contact with English outside school. This would suggest that the potential impact of extracurricular input on learners' writing skills is subject to a ceiling effect: once a certain threshold of activity is reached, the variable ceases to be distinctive, at least as fluency (text length) and lexical diversity are concerned. Beyond that point, the amount of input received through formal instruction emerges as the predicting input variable, as reported here for the written productions of the English learners in our sample.

6. Conclusions

The aim of this study was to explore the impact of formal and informal input on learners' variability in writing, and to compare two target language conditions (Dutch and English) in CLIL and non-CLIL settings in French-speaking Belgium. Overall, we can conclude that the CLIL approach results in increased L2 written proficiency[7], even when controlling for other types of input. Our findings thus contradict those reported on by Saladrigues & Llanes (2014), a contrast that can perhaps be partially explained by the difference in sample size, since their study was conducted on a very small sample (total n = 39) of pupils from the same school. The CLIL advantage in our sample is most evident in terms of fluency and lexical diversity, which would confirm international results regarding the effect on oral proficiency (Dalton-Puffer, 2008). The results for sentence length and word length are less straightforwardly interpretable, but our analyses indicate that these variables are clearly more dependent

on other factors than the amount of input the learners were exposed to, factors which were outside the scope of the present study. Possible candidates of these additional factors are related to the type and quality of instructed input (such as for instance the teacher's TL proficiency), specific educational strategies, or types of teaching materials. However, variables such as these remain hard to quantify, and studies combining a quantitative and a qualitative approach would therefore be of great value. Möller's (2017) work on the acquisition of the passive, for instance, is one of the few studies that also analyses the teaching materials in CLIL and non-CLIL settings, thus combining a quantitative and a qualitative approach to input.

Regarding the influence of formal and informal input on written proficiency, we found that both types of input correlate with the pupils' writing skills, but that their relative impact on these skills differs starkly depending on the target language: the amount of formal language exposure predicts the outcomes of the English learners' written productions (written fluency, word length, and lexical diversity), and the frequency of informal exposure those of the Dutch learners (written fluency and lexical diversity). We argued that this observation can perhaps be explained by the difference in status that each of these languages holds among the pupils in our sample (and in the whole of French-speaking Belgium at that). A more favorable stance toward the 'cool' and 'international' language English as opposed to Dutch, generally regarded as less attractive and more difficult (De Smet et al. 2018), may explain why, on average, all English learners report regular informal exposure to the target language, resulting in a loss of discriminatory power of this variable. By contrast, those Dutch learners who are in contact with the target language outside school clearly benefit from the additional exposure they enjoy in informal contexts. In any case, the fact that our results differ according to the target language has clear implications for (SLA) research looking at the effect of input on L2 learning. In fact, different input variables do not only affect the various components of proficiency in different ways (e.g. Muñoz, 2008), but also the relative impact of these different input variables on the learners' proficiency apparently varies depending on the target language. Given the fact that most research in this field is conducted with English as a target language, as is research on CLIL, we believe this is an important finding and are prompting studies on TL other than English.

Notes

[1] The texts were semi-automatically corrected for punctuation mistakes (missing spaces after commas and full stops were added) to increase the accuracy of the measures. As correcting the spelling and grammatical errors did not significantly improve the accuracy of the measures (Bulon et al., 2017), these were not corrected.

[2] We are aware of the fact that T-unit has been recognized as a better measure to assess complexity at the sentence level. However, neither Coh-Metrix nor T-scan provided that measure.

[3] MTLD is calculated as the "mean length of sequential word strings in a text that maintain a given TTR value" (McCarthy & Jarvis, 2010, p. 385). Koizumi (2012) found that MTLD was least affected by text length compared to TTR and other recent indices (e.g. Guiraud index and D), when used with texts of at least 100 tokens. Furthermore, the measure appears to be a good predictor of overall L2 proficiency (Treffers-Daller, 2013; Crossley, Salsbury & McNamara, 2014).

[4] We should, however, note that the p value of the correlation between word length and length of TL exposure is only just below the cut-off point of .05 (.048).

[5] The coefficient of determination indicates the proportion of data explained by the model.

[6] See De Smet et al. (2018) for some evidence regarding more positive attitudes of the pupils toward English in our sample.

[7] The present study did not involve any qualitative analysis of the written productions, such as an investigation of the errors. We believe this may add to our analysis of the learners' writing skills and ongoing research is being carried out accordingly (among others, Bulon, *forthc.*).

References

Admiraal, W., Westhoff, G., & de Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational Research and Evaluation*, *12*(1), 75-93. doi: 10.1080/13803610500392160

Bulon, A. (*forthc.*). QQQ

Bulon, A., Hendrikx, I., Meunier, F. & Van Goethem, K. (2017). Using global complexity measures to assess second language proficiency: Comparing CLIL and non-CLIL learners of English and Dutch in French-speaking Belgium. *Papers of the Linguistic Society of Belgium*, *11*(1), 1-25. Retrieved from https://sites.uclouvain.be/bkl-cbl/wp-content/uploads/2017/04/Bulon_et_al_2017.pdf

Burstall, C. (1975). Factors Affecting Foreign-Language Learning: A Consideration of Some Relevant Research Findings. *Language Teaching and Linguistic Abstracts, 8,* 105-125. doi: https://doi.org/10.1017/S0261444800002585

Carroll, J. B. (1969). Psychological and educational research into second language teaching to young children. In H.H. Stern (Ed.), *Languages and the young school child* (pp. 2–68). London: Oxford University Press.

Celaya, M. L., & Navés, T. (2009). Written production in English as a Foreign Language: age-related differences and associated factors. In R. Manchón (Ed.), *Learning, Teaching and Researching Writing in Foreign Language Learning Contexts* (pp. 130-155). Clevedon: Multilingual Matters.

Celaya, M. L. & Ruiz de Zarobe, Y. (2010). First language and age in CLIL and non-CLIL contexts. *International CLIL Research Journal, 1*(3), 60-66. Retrieved from http://www.icrj.eu/13/article6.html

Celaya, M. L., Torras, M. R., & Vidal, C. P. (2001). Short and mid-term effects of an earlier start: An analysis of EFL written production. *Eurosla Yearbook*, *1*(1), 195-209. doi: 10.1075/eurosla.1.15cel

Cenoz, J. (2002). Age differences in foreign language learning. *ITL Review of Applied Linguistics, 135/136,* 125-142. doi: 10.1075/itl.135-136.06cen

Cenoz, J. (2003). Facteurs déterminant l'acquisition d'une L3 : Age, développement cognitif et milieu. *Acquisition et interaction en langue étrangère*, *18,* 38-51. Retrieved from http://journals.openedition.org/aile/1151

Crossley, S. A., Salsbury, T. & McNamara, D. S. (2014). Validating lexical measures using human scores of lexical proficiency. In S. Jarvis & M. Daller (Eds.), *Human ratings and automated measures* (pp. 47-105). John Benjamins, Amsterdam.

Cumming, A. (1994). Writing expertise and second-language proficiency. In A. Cumming (Ed.), *Bilingual performance in reading and writing* (pp. 173–221). Ann Arbor, MI: Language Learning /John Benjamins.

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of applied linguistics*, *31*, 182-204. doi: 10.1017/S0267190511000092

Dalton-Puffer, C. (2008). Outcomes and processes in Content and Language Integrated Learning (CLIL): current research from Europe. In W. Delanoy & V. Lauren (Eds.), *Future Perspectives for English Language Teaching* (pp. 139-157). Heidelberg: Carl Winter.

De Smet, A. et al. (*in preparation*). QQQ

De Smet, A., Mettewie, L., Galand, B., Hiligsmann, Ph., & 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*(1), 47-72.

García Lecumberri, M.L., & Gallardo, M.P. (2003). English FL sounds in school learners of different ages. In M.P. García Mayo & M. L. Garcia Lecumberri (Eds.), *Age and the Acquisition of English as a Foreign Language* (pp. 115-135). Clevedon: Multilingual Matters.

Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2015). Development of EFL writing over three years in secondary education: CLIL and non-CLIL settings. *The Language Learning Journal*, *43*(3), 286-303. doi: 10.1080/09571736.2015.1053278

Graesser, A. C., McNamara, D. S., Louwerse, M. M. & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments & computers*, *36*(2), 193-202. doi: 10.3758/BF03195564

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 in the French-speaking Community of Belgium: linguistic, cognitive and educational perspectives.

Cahiers du GIRSEF, 109, 1-24. Retrieved from https://cdn.uclouvain.be/groups/cms-editors-girsef/demey/Cahier_109_final.pdf

Isidro, X. S. & Lasagabaster, D. (2018). The impact of CLIL on pluriliteracy development and content learning in a rural multilingual setting: A longitudinal study. *Language Teaching Research*, 1-19. doi: 10.1177/1362168817754103

Jexenflicker, S., & Dalton-Puffer, C. (2010). The CLIL differential: Comparing the writing of CLIL and non-CLIL students in higher colleges of technology. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 169-190). John Benjamins, Amsterdam.

Johnstone, R. (2007). Nationally-sponsored innovations at school in Scotland: issues of evidence, generalizability and sustainability. *International Journal of Innovation in Language Learning and Teaching*, *1*(1), 111-128. doi: 10.2167/illt046.0

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60-99. doi: 10.1016/0010-0285(89)90003-0

Kinsella, C. (2009). *An investigation into the proficiency of successful late learners of French* (Unpublished doctoral thesis). Trinity College, Dublin, Ireland.

Klieme, E. (2006). *Zusammenfassung zentraler Ergebnisse der DESI Studie. Deutsches Institut für Internationale Pädagogische Forschung.* Retrieved from https://www.dipf.de/de/forschung/aktuelle-projekte/pdf/biqua/DESI_Ausgewaehlte_Ergebnisse.pdf

Koizumi, R. (2012). Relationships between text length and lexical diversity measures: can we use short texts of less than 100 tokens? *Vocabulary Learning and Instruction, 1*(1), 60-69. doi: 10.7820/vli.v01.1.koizumi

Lahuerta Martínez, A.C. (2015). Analysis of the Written Competence of Secondary Education Students in Bilingual and Non-Bilingual Programmes. In *Conference proceedings. ICT for language learning* (pp. 499-503) Padova: Libreriauniversitaria.it Edizioni.

Lambelet, A., & Berthele, R. (2015). *Age and foreign language learning in school*. Basingstoke: Palgrave MacMillan.

Lasagabaster, D. (2008). Foreign Language competence in Content and Language Integrated Courses. *The Open Applied Linguistics Journal, 1,* 31- 42. doi: 10.2174/1874913500801010030

Lasagabaster, D., & Doiz, A. (2003). Maturational constraints on foreign language written production. In M.P. García Mayo & M.L. Garcia Lecumberri (Eds.), *Age and the Acquisition of English as a Foreign Language* (pp. 136-160). Clevedon: Multilingual Matters.

Llanes, À. & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, *37*(3), 353-365. doi: 10.1016/j.system.2009.03.001

Lo Y.-Y., & Murphy V. A. (2010). Vocabulary knowledge and growth in immersion and regular language-learning programmes in Hong Kong. *Language and Education*, *24*, 215–238. doi: 10.1080/09500780903576125

Long, M. H. (1983). Does second language instruction make a distinction? A review of research. *TESOL quarterly*, *17*(3), 359-382. doi: 10.2307/3586253

Lorenzo, F., Casal, S., & Moore, P. (2005). *Orientaciones para la elaboración del currículo integrado de las lenguas en los centros bilingües*. Seville: Consejería de Educación (Junta de Andalucía).

McCarthy, P. M. & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381-392. doi: 10.3758/BRM.42.2.381

Miralpeix, I. (2006). Age and vocabulary acquisition in English as a foreign language. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 89-106). Clevedon: Multilingual Matters.

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Identity, Social Relationships and Language Learning During Residence Abroad*. London and New York: Routledge.

Möller, V. (2017). *Language Acquisition in CLIL and Non-CLIL Settings: Learner corpus and experimental evidence on passive constructions*. Amsterdam/Philadelphia: John Benjamins.

Moyer, A. (2009). Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In T. Piske & M. Young-Scholten (Eds.), *Input Matters in SLA* (pp. 159–174). Bristol: Multilingual Matters.

Muñoz, C (2006). The effects of age on foreign language learning: The BAF Project. In C. Muñoz (Ed.), *Age and the Rate of Foreign Language Learning* (pp. 1-40).

Clevedon: Multilingual Matters.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics, 29,* 578-96. doi: 10.1093/applin/amm056

Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *IRAL-International Review of Applied Linguistics in Language Teaching*, *49*(2), 113-133. doi: 10.1515/iral.2011.006

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics, 35*(4), 463-482. doi: 10.1093/applin/amu024

Myles, J. (2002). Second language writing and research: The writing process and error analysis in student texts. *Tesl-Ej, 6*(2), 1-20. Retrieved from http://www.tesl-ej.org/wordpress/issues/volume6/ej22/ej22a1/?wscr

Pander Maat, H., Kraf, R., Dekker, N., Sloot, K. van der, Bosch, A. van den, Gompel, M. van & Klein, S. (2014). T-Scan: a new tool for analyzing Dutch text. *Computational Linguistics in the Netherlands Journal, 4,* 53-74.

Patkowski, M. S. (1980). The Sensitive Period for the Acquisition of Syntax in a Second Language. *Language Learning, 30,* 449–468. doi: 10.1111/j.1467-1770.1980.tb00328.x

Pérez-Vidal, C. & Juan-Garau, M. (2011). The effect of context and input conditions on oral and written development: A Study Abroad perspective. *VIAL, Vigo international journal of applied linguistics, 4,* 157-185. doi: 10.1515/iral.2011.008

Pérez-Vidal, C., & Roquet, H. (2015). The linguistic impact of a CLIL Science programme: An analysis measuring relative gains. *System, 54,* 80-90. doi: 10.1016/j.system.2015.05.004

Raven, J. C., Court, J. H., & Raven, J. (1998). *Progressive Coloured Matrices.* Oxford: Oxford Psychologists Press.

Roquet, H. & Pérez-Vidal, C. (2017). Do productive skills improve in Content and Language Integrated Learning contexts? The case of writing. *Applied Linguistics, 38*(4), 489-511. doi: 10.1093/applin/amv050

Ruiz de Zarobe, Y. (2008). CLIL and foreign language learning: A longitudinal study in the Basque Country. *International CLIL Research Journal, 1*(1), 60-73. Retrieved from http://www.icrj.eu/11/article5.html

Ruiz de Zarobe, Y. (2010). Written production and CLIL: An empirical study. In C. Dalton-Puffer, T. Nikula & U. Smit (Eds.), *Language Use and Language Learning in CLIL Classrooms* (pp. 191-212). Amsterdam/Philadelphia: John Benjamins.

Saladrigues, G. & Llanes, À. (2014). Examining the impact of amount of exposure on L2 development with CLIL and non-CLIL teenage students. *Sintagma, 26*, 133-147. Retrieved from http://www.raco.cat/index.php/Sintagma/article/view/293008/381457

Simonis, M., Galand, B., Hiligsmann, Ph., Van der Linden, L., & Szmalec, A. (submitted). Executive control in bilingual immersion education. No measurable evidence for a bilingual executive control advantage in language immersion education. *Bilingualism: Language and cognition.*

Torras, M. R., & Celaya, M. L. (2001). Age-related differences in the development of written production. An empirical study of Efl school learners. *International Journal of English Studies*, *1*(2), 103-126. Retrieved from http://revistas.um.es/ijes/article/view/48211/46181

Torras, M. R., Teresa Navés, M. Luz Celaya, & Carmen Pérez-Vidal. 2006. Age and IL development in writing. In Carmen Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 156–182). Clevedon: Multilingual Matters.

Treffers-Daller, J. (2013). Measuring lexical diversity among L2 learners of French: an exploration of the validity of D, MTLD and HD-D as measures of language ability. In S. Jarvis & M. Daller (Eds.), *Vocabulary Knowledge: Human Ratings and Automated Measures* (pp. 79-105). Amsterdam/Philadelphia: John Benjamins.

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

Wesche, M. B. (2002). Early French immersion: How has the original Canadian model stood the test of time? In P. Thorsten, A. Rohde, H. Wode, & P. Burmeister (Eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode* (pp. 357-378). Trier: Wissenschaftlicher Verlag Trier.

Whittaker, A., & Llinares, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research, 15*(3), 343-362. doi: 10.1177/1362168811401154

Wode, H. (1994). *Bilinguale Unterrichtserprobung in Schleswig-Holstein. Band II Analytische Auswertungen*. Kiel: l&f Verlag.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity.* Hawaii: University of Hawaii Press.

Zydatiß, W. (2007). *Deutsch-Englische Züge in Berlin (DEZIBEL). Eine Evaluation des bilingualen Sachfachunterrichts in Gymnasien: Kontext, Kompetenzen, Konsequenzen.* Frankfurt am Main: Peter Lang.

Appendix 1

Frequency of informal contact with the target language outside school (Cronbach's alpha .78), on a scale from 1 to 5 (never – rarely – sometimes – often – very often), based on the following items:

[Original in French]

1. Sur internet, à quelle fréquence utilises-tu l'anglais / le néerlandais?
2. A quelle fréquence parles-tu l'anglais / le néerlandais en dehors de l'école?
3. A quelle fréquence entends-tu l'anglais / le néerlandais en dehors de l'école?
4. As-tu des contacts avec des anglophones / des néerlandophones en dehors de l'école ou de la maison?

[English translation]

1. How often do you use English / Dutch on the Internet?
2. How often do you speak English / Dutch outside school?
3. How often do you hear English / Dutch outside school?
4. Do you have any contact with English-speakers / Dutch-speakers outside school or home?