Linguistic Research 37(Special Edition): 89-112
DOI: 10.17250/khisli.37..202009.004

Assessing the effects of CLIL on Korean high school students' writing*

Jongbong Lee
(Nagoya University of Commerce & Business)

Lee, Jongbong. 2020. Assessing the effects of CLIL on Korean high school students' writing. *Linguistic Research* 37(Special Edition): 89-112. This study examines the effects of content and language integrated learning (CLIL) on several aspects of the written language competence of learners of English as a foreign language (EFL). Controlling for L2 proficiency and L1, the study compares the English written narratives produced by CLIL (N = 29) and non-CLIL (N = 35) 11th-grade students in terms of syntactic complexity, lexical complexity, cohesion, and writing quality. The results show that the CLIL group outperformed the non-CLIL group according to measures of mean length of clause, lexical sophistication, lexical diversity, and writing quality, while the non-CLIL group's writing showed greater semantic cohesion. The findings contribute to CLIL research, previously limited to a focus on morphosyntactic features or writing quality, by expanding it to investigate CLIL effects on syntactic complexity, lexical complexity, and cohesion in addition to writing quality. **(Nagoya University of Commerce & Business)**

Keywords CLIL, writing, narrative, secondary school

1. Introduction

The educational approach of content and language integrated learning (CLIL), in which a second language (L2) serves as the medium of instruction for content subjects such as mathematics or science, is widely considered an effective means of improving L2 proficiency. Particularly in English as a foreign language (EFL) contexts, many stakeholders (e.g., policymakers, parents, teachers) see CLIL as a mechanism for providing English learners the skills to achieve success in educational environments and gain the competence necessary to participate in international academic communities (Dalton-Puffer 2011; Dalton-Puffer, Nikula,

* The author would like to thank the Editor and anonymous reviewers for their constructive feedback on the manuscript.

and Smit 2010). The basic assumption of CLIL is that it is beneficial because it allows authentic language use in natural settings and thus leads to a natural type of learning (Coyle, Hood, and Marsh 2010). The naturalistic setting provides the learners with intense input and opportunities to produce language; thus, the CLIL is believed to benefit language development in terms of quantity and quality (e.g., Dalton-Puffer 2011).

Research from diverse perspectives has reported CLIL's linguistic benefits. For example, studies have compared the performance of CLIL and non-CLIL students in terms of vocabulary knowledge (Lo and Murphy 2010; Ruiz de Zarobe 2008, 2010; Vidal and Jarvis in press), compensatory strategies (Gallardo-del-Puerto, Basterrechea, and Martínez-Adrián in press), attitudes toward CLIL (Lo 2014), communicative competence (Martinez Agudo, 2019), the role of L1 use (García Mayo and Hidalgo 2017), and overall proficiency (Merino and Lasagabaster 2018).

As noted by Roquet and Pérez-Vidal (2017), studies on CLIL's effects on learners' writing performance, however, are scarce. When previous research on CLIL has looked at production skills, it has tended to focus only on oral communicative competence rather than writing performance. Yet CLIL requires learners to actively produce written text in content classes, and therefore there is a need for research that will help us understand its effects on written production as well (Dalton-Puffer 2009). Furthermore, to be effective, EFL program development, policy making, and pedagogy should be informed by empirically supported theory that addresses the multidimensional nature of the CLIL approach and its effects on writing (e.g., Knoch, Rouhshad, Oon, and Storch 2015; Roquet and Pérez-Vidal 2017).

Given the importance of examining writing performance and understanding the effects of CLIL, and the current lack of studies that have done so, more comparative research (CLIL vs. non-CLIL) from different perspectives is worthwhile. Most studies on CLIL have examined only part of the linguistic outcomes such as overall writing quality or morpho-syntactic features (e.g., Jexenflicker and Dalton-Puffer 2010; Ruiz de Zarobe 2010). In addition, previous research on CLIL has not focused on sentence-level aspects of production such as cohesion in writing. For theory building purposes, the present study is designed to address these gaps in the literature by delving into the effects of

CLIL on writing through examining writing quality and using a range of linguistic measures of syntactic complexity, lexical complexity, and cohesion (e.g., Dalton-Puffer 2011). The study thus will contribute to the body of research on CLIL and to the formation of policy by providing empirical evidence on the effects of CLIL on L2 writing skills.

2. Literature review

2.1 CLIL and second language learning

CLIL is defined as an educational approach in which content is taught through the medium of a target language (Dalton-Puffer 2011). As a type of immersion education, CLIL has been adopted mostly in European contexts, but it is used in other countries. This approach is characterized by the use of the target language, often English in EFL contexts. As Cenoz (2015) noted, the terms "content-based instruction" and "CLIL" could be used interchangeably in that the two share the same properties and are not different in terms of teaching. Both refer to academic programs where the medium of instruction is the target language. CLIL has the advantage of providing L2 learners with more natural and intense input than mainstream EFL contexts (e.g., Merino and Lasagabaster 2018). The purpose of CLIL is to use content learning to motivate foreign language learning (Dalton-Puffer 2011). In CLIL schools, students are expected to be language users who study subjects rather than language learners in foreign language classrooms.

CLIL has gained popularity because of its meaning-oriented and naturalistic features in communicative language teaching. The theoretical basis of the approach comes from language acquisition models such as Krashen's (1985) input hypothesis and cognitive interactionist models (Gass 1997; Long 1996). These models posit that meaning-oriented classrooms offer high quality input and opportunities to negotiate meaning, which are necessary to develop the L2. Based on this theoretical background, CLIL is believed to provide a meaningful environment for L2 learners to learn language.

However, there is mixed evidence on the effects of CLIL (e.g., Bruton 2013).

Lorenzo, Casal, and Moore (2009), for instance, reported that CLIL students may have a higher command of the target language than their non-CLIL peers. However, if the content is too complicated or unfamiliar, CLIL might hinder language development (Tan 2011). In addition, Bruton (2013) suggested that a fair comparison would require many hours of language instruction for both CLIL and non-CLIL students. Since the CLIL programs in previous studies provide the CLIL students extra hours of exposure to the target language (e.g., content classes in the target language and the target language class), it is difficult to compare the effect of CLIL fairly (e.g., Roquet and Pérez-Vidal 2017). Therefore, more rigorous research on this topic is necessary (e.g., Vidal and Jarvis in press).

2.2 Previous research on the effects of CLIL on writing

While a considerable amount of research has been conducted on CLIL's effects, only a handful of studies have examined the effects on written production. These few studies typically take a global approach to assessing writing quality. For instance, Jexenflicker and Dalton-Puffer (2010) examined global language competence by assessing task fulfillment, organization, grammar, and vocabulary in a comparison of CLIL and non-CLIL EFL students' written production. They found that CLIL students did better than non-CLIL students in all the dimensions rated. Similarly, Ruiz de Zarobe (2010) investigated the written production of CLIL and non-CLIL students by assessing writing quality in terms of content, organization, vocabulary, language usage, and mechanics. The study suggested that the CLIL students performed better in content and vocabulary. However, it relied on writing scores derived from Jacob et al.'s (1981) analytic rubric, which has been criticized: Connor-Linton and Polio (2014) pointed out that the descriptors used by Jacob et al. did not match raters' perceptions. More broadly, however, previous studies have relied too heavily on global writing scores, assuming that they reflect linguistic development. Meanwhile, other research has shown that examining more specific aspects of production (e.g., fluency, different types of complexity) in addition to overall writing quality gives a clearer picture of how L2 learners develop their writing skills (e.g., Roquet and Perez-Vidal 2017).

Some studies have explored the effects of CLIL on writing with regard to specific aspects of production such as linguistic complexity (e.g., Navés 2011; Roquet and Pérez-Vidal 2017; Vidal and Jarvis in press). For example, Navés (2011) reported results suggesting that CLIL students outperformed non-CLIL students in terms of fluency, syntactic complexity, and lexical complexity. Navés argued that integrating language and content in class created a more meaningful context and offered learners more intensive language practice than traditional language classes could. However, as Navés acknowledged, the study did not control the hours of instruction because different schools were involved in the project, which reduces the validity of the findings. Vidal and Jarvis (in press) investigated the effects of English medium instruction on lexical diversity and writing quality by comparing first-year and third-year students' argumentative essays. They found effects on writing quality, but not on lexical diversity, suggesting that the language proficiency of the third-year students was still not high enough to lead to an improvement in lexical diversity over the first-year students. On the other hand, Roquet and Perez-Vidal (2017) compared CLIL and non-CLIL groups in a formal instructional setting, examining linguistic complexity, lexical complexity, and writing quality. Their results were inconclusive, although they reported that the CLIL group showed some improvement over a year in terms of writing quality. However, the study used only one measure for each construct (e.g., one coordination measure for syntactic complexity). In addition, at the beginning of the experiment, the CLIL group's proficiency was higher than that of the non-CLIL group. Employing more measures of each construct and controlling for proficiency are necessary to achieve greater validity.

Dalton-Puffer (2011) found significant CLIL effects on morpho-syntactic structures, but only negligible effects on sentence-level language (e.g., cohesion) in written production. Only a few other studies on CLIL have examined the sentence level (Bae 2001; Whittaker, Llinares, and McCabe 2011). Jexenflicker and Dalton-Puffer (2010), for example, found that CLIL students received higher scores on organization than non-CLIL students in the rubric they employed; however, the study did not examine specific cohesive devices. More studies are needed to find the effects of CLIL on cohesion by using specific measures for cohesive devices.

2.3 The present study

The present study contributes to the CLIL research by investigating the effects of CLIL in different linguistic domains by assessing a variety of textual features as well as using writing scores. The study compares the English writing competence of CLIL and non-CLIL 11th-grade EFL students. It thus addresses several gaps in the literature, as discussed in the literature review above, by exploring the effects of CLIL on written performance in terms of textual features such as linguistic complexity. In particular, there is an overall scarcity of research on CLIL learners' writing performance. The existing research on the topic, which has had mixed results, uses very limited textual measures. This study attempts to more fully explore CLIL effects on writing by examining a range of constructs and using several measures for each. In addition, this study selected participants in such a way as to control for two external factors: English proficiency level and school. First, all participants had the same English proficiency level (advanced). Second, according to Roquet and Pérez-Vidal (2017), most of the previous studies that found CLIL students academically outperforming non-CLIL students had been conducted at a single school. Considering that it is mostly the highly motivated students of a school who tend to join the CLIL program, the current study includes students from more than one school to increase the generalizability and validity of the findings. Furthermore, most studies on CLIL have been conducted in European countries. The findings of this study, conducted in South Korea, will contribute to the body of CLIL research by extending the investigation of CLIL effects to a different region of the world.

The study is guided by the following research questions:

1. What are the effects of CLIL on written syntactic complexity, lexical complexity, and cohesion?
2. To what extent does CLIL affect writing quality?

3. Method

3.1 Participants

This study has a total of 64 participants, who were all 11th-grade students in Seoul, South Korea at the time of the data collection. Of the 64 participants, 29 (8 males, 21 females) are from a public international high school and 35 (14 males, 21 females) are from three different public regular high schools. Both groups of students studied English as a second language.

Before attending high school, both groups of students went to regular middle schools approved by the Ministry of Education, which all followed the same national curriculum. The South Korean national curriculum for high schools specifies that students in both school types (international and regular) must spend the same amount of time in school. However, the international high school students had received content-based language instruction for a year in all subjects except in the Korean language and history. The students took other content classes such as mathematics in English in addition to regular English language classes. The students did not take additional English language classes specifically designed to develop writing. On the other hand, the regular high school students took content classes such as mathematics in Korean. The regular high school students had received non-content-based language instruction consisting of 50-minute English classes four times a week. The students at regular schools all receive the same number of hours of English language instruction per week as directed by the government curriculum. The public international high school is approved by the Ministry of Education, and the students attending the school are Korean nationals; it thus differs from some other international schools in Korea, which accept only foreigners or Korean nationals who have resided in foreign countries for at least three years.

To investigate the effects of CLIL on writing specifically, this study controlled for English proficiency on the basis of the participants' English scores on the Korea Preliminary College Scholastic Ability Test (KPCSAT) administered by the Korean Ministry of Education. The scores of both groups were within the top four percent of the whole population, identifying them as highly advanced EFL learners. Furthermore, a comparison of their average length of residence in

English-speaking countries found no significant difference between the two groups ($t(46) = 1.34$, $p = .185$). Table 1 summarizes the demographic data of the participants.

Table 1. Demographic data of the participants: CLIL and non-CLIL high school students

	CLIL students ($N = 29$)	Non-CLIL students ($N = 35$)
Male	8	14
Female	21	21
Grade level	11th	11th
KPCSAT score	Within top 4%	Within top 4%
Time in Eng.-speaking countries(mean)	16.59 months. (SD = 22.25)	10.11 months. (SD = 14.62)
CLIL instruction	Yes	No

3.2 Materials

Each group of participants was given a written prompt: "my most frightening experience" (Kang 2005), which was intended to elicit a firsthand story, or personal narrative. Personal narratives as a genre can provide information on authors' special lexical and syntactic choices; topic selection; rhetorical patterns; and relation to social, cultural, cognitive, and educational issues (Özyıldırım 2009). Liskin-Gasparro (2000) also suggested that narrating a personal experience is a complex linguistic task, considerably more so, for example, than recounting a film segment. Another reason this study used personal narratives was to avoid favoring one program over another, under the assumption that the CLIL students, due to their English-medium content courses, might have an advantage in certain genres such as argumentative essays, or on specific topics such as science. A background questionnaire was also used to collect the participants' biographic information including age, sex, length of residence in English-dominant countries, and KPCSAT scores.

3.3 Procedure

The experiment was conducted during class time. First, the study was

explained to the students. Next, the writing task was introduced. All of the participants were asked to produce a narrative essay in English during the class time (50 minutes). The students were given paper and asked to hand-write their essays. They were not allowed to use external resources such as dictionaries while writing. The time students spent on writing the essay ranged from 15 minutes to 45 minutes.

3.4 Scoring

The essays were rated using an analytic rubric for narratives based on Connor-Linton and Polio's (2014) study. In the rubric, the full score for mechanics is 10 points, and the full score for each of the four other categories (content, organization, vocabulary, language use) is 20 points; the total score is 90 points. Two native English speakers, who were expert raters and had taught ESL and EFL students, did the rating. Both were instructors at an English language center at a university and were pursuing their master's degrees in TESOL. Before they rated the essays, the raters participated in a two-hour norming session where they rated sample narratives that were not part of this study and discussed their scoring (e.g., Kim, Lee, and You 2019). If a discrepancy in any subscale was larger than two points, the raters discussed and resolved the discrepancy. After the norming session, the raters rated all of the essays independently, and the average scores obtained from the two raters were used for the analysis. If essays received discrepant scores (subscale scores differing by three or more), a third rater rated the essays, and the two closer scores were utilized to find average scores. The interrater reliability between the two raters was calculated by Pearson correlations. The reliability for total score was r = .85 (content: r = .85, organization: r = .80, vocabulary: r = .73, language use: r = .67, mechanics: r = .70). Based on the findings of Brown, Glasswell, and Harland (2004), reliability of 0.70 is a benchmark for structured rubrics, and thus the interrater reliability is acceptable.

3.5 Text analysis

In order to answer the research question regarding the effects of CLIL on

linguistic complexity, this study utilized Lu's (2010) web-based, computerized syntactic complexity analyzer, which has 14 syntactic complexity measures. This system shows high reliability in searching for specific types of syntactic structures, as well as high reliability between the system and human coders (Ai and Lu 2013).

Coh-Metrix 3.0 (McNamara, Graesser, McCarthy, and Cai 2014) was used to assess lexical sophistication with average word length and word frequency. The lexical sophistication measures were generated on the basis of frequency norms of the CELEX lexical database, which provides an analysis of 17.9 million words. From among the measures of the frequency of all words, CELEX word frequency for content words (WF) and average word length (WL) were chosen to prevent rare words from creating a limiting factor. In interpreting WF, a lower value indicates less frequent words and a higher value indicates more frequent words. WL also has been used in L2 studies because average word length tends to increase as L2 proficiency develops. These lexical sophistication measures are believed to predict L2 development (Crossley, Cobb, and McNamara 2013). Lexical diversity, D, was assessed with the VOCD function in Coh-Metrix. The VOCD function automatically calculates and averages type-token ratios for 100 random trials from a transcript of token size (McCarthy and Jarvis 2007; Yu 2010). Before the VOCD function was run, spelling mistakes were corrected (Yu 2010). The lexical measures can show how these lexical complexity indices interact with CLIL. In order to assess cohesion, which shows the links within texts, Coh-Metrix was also used. According to Crossley, Kyle, and McNamara (2016), cohesion can be measured at text, global, and local levels. Causal and temporal connective indices were utilized for text cohesion, and coreferential (argument overlap in adjacent sentences) and semantic (latent semantic analysis overlap in adjacent sentences; level of semantic similarities) features were used to assess global and local cohesion (e.g., Lee 2018).

3.6 Statistical analysis

To address the first research question on the effects of CLIL on textual features, multivariate analysis of variance (MANOVA) was conducted. First,

however, a correlation analysis was performed to check multicollinearity ($r > .8$); highly correlated variables were excluded from the analysis. Table 2 summarizes the textual features that were included in the analysis. For the second research question concerning the effects of CLIL on writing quality, another MANOVA was conducted to see whether there are similarities or differences between the CLIL and non-CLIL groups in terms of total score (i.e., writing quality), content, organization, vocabulary, language use, and mechanics.

Table 2. Syntactic complexity, lexical complexity, and cohesion measures

Measures	Definition
Syntactic complexity:	
Length of production	
Mean length of clause (MLC)	Number of words / Number of clauses
Subordination	
Dependent clause ratio (DC/C)	Number of dependent clauses / Number of clauses
Coordination	
Coordinate phrases per clause (CP/C)	Number of coordinate phrases / Number of clauses
Particular structure	
Complex nominals per clause (CN/C)	Number of complex nominals / Number of clauses
Verb phrases per T-unit (VP/T)	Number of verb phrases / Number of T-units
Lexical complexity:	
Word length (WL)	Average length of word
Word frequency (WF)	CELEX word frequency for content words
Lexical diversity (D)	VOCD-D
Cohesion:	
Causal	Causal connective (e.g., because, so)
Temporal	Temporal connective (e.g., first, until)
Coreferential	Argument overlap (e.g., 'The sun was bright. The day was sunny.')
Semantic	LSA overlap (e.g., 'The dog was tired, so was the cat.')

With the explore function in SPSS, descriptive statistics and confidence intervals were obtained. Evaluation of the homogeneity of variance-covariance matrices (Box's *M*), error variances (Levene's test), linearity, non-multicollinearity,

and normality assumptions underlying MANOVAs found no substantial anomalies. Along with p-values, effect sizes (η_p^2, partial eta squared) for inferential statistics are reported. According to Cohen (1969), small, medium, and large effect sizes correspond to values of .0099, .0588, and .1379, respectively.

4. Results

The descriptive statistics show the values of syntactic complexity, lexical complexity, and cohesion in the narrative writing produced by the CLIL and non-CLIL students (see Table 3). For some of the linguistic measures such as WL and WF, 95% confidence intervals for the two groups (CLIL and non-CLIL) did not overlap, which indicates a large difference between the groups. All measures except for those of causal and semantic cohesion appeared to show better outcomes in the CLIL students' writing than in the non-CLIL students' writing.

Table 3. Descriptive statistics of CLIL and non-CLIL students' writing

		CLIL (N = 29)		Non-CLIL (N = 35)	
Variables		M (SD)	95% CI	M (SD)	95% CI
Syntactic complexity	MLC	7.82 (.94)	7.47, 8.18	7.29 (1.03)	6.94, 7.65
	DC/C	.34 (.07)	.31, .36	.33 (.10)	.29, .36
	CP/C	.18 (.07)	.15, .20	.14 (.07)	.11, .16
	CN/C	.64 (.16)	.58, .69	.55 (.15)	.50, .60
	VP/T	1.96 (.26)	1.86, 2.06	1.95 (.37)	1.82, 2.08
Lexical complexity	WL	1.36 (.05)	1.34, 1.38	1.30 (.05)	1.28, 1.31
	WF	2.43 (.13)	2.38, 2.48	2.54 (.09)	2.51, 2.57
	D	78.71 (20.60)	70.88, 86.55	67.25 (13.25)	62.69, 71.80
Cohesion	Causal	26.48 (9.85)	22.72, 30.22	29.27 (11.28)	25.40, 33.15
	Temporal	26.99 (10.23)	23.10, 30.88	25.83 (8.80)	22.80, 28.85
	Coreferential	.63 (.14)	.58, .69	.67 (.16)	.62, .72
	Semantic	.16 (.03)	.15, .17	.20 (.06)	.18, .22

According to the MANOVA results, many linguistic measures were significantly influenced by CLIL (see Table 4). The MANOVA indicated a significant CLIL effect on a length of production unit measure (MLC) with a medium effect size (F(1, 62) = 4.58, p = .036, η_p^2 = .069). It also showed a significant effect of CLIL on the phrase level complexity measures of CN/C (F(1,

62) = 4.76, p = .033, η_p^2 = .07) and VP/T (F(1, 62) = 4.78, p = .033, η_p^2 = .07), suggesting that CLIL students produced more complex language at the phrase level than non-CLIL students (see Figure 1).

Table 4. CLIL effects on linguistic measures (MANOVA)

		CLIL		
Variables		F	p	η_p^2
Syntactic complexity	MLC	4.583	.036*	.07
	DC/C	.009	.923	.00
	CP/C	.293	.590	.01
	CN/C	4.757	.033*	.07
	VP/T	4.778	.033*	.07
Lexical complexity	WL	20.542	<.001***	.25
	WF	15.382	<.001***	.20
	D	7.246	.009**	.11
Cohesion	Causal	1.095	.299	.02
	Temporal	.238	.627	.00
	Coreferential	1.147	.288	.02
	Semantic	8.345	.005**	.12
Wilks' lambda	.613		.007**	.39

*p< .05; **p < .01; ***p < .001

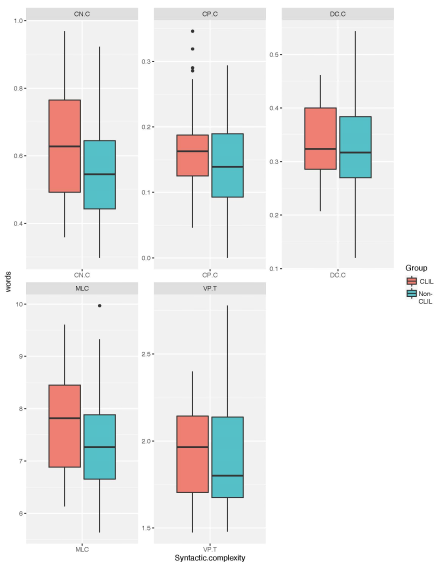

Figure 1. Comparison of syntactic complexity measures in CLIL and non-CLIL students' writing

All lexical complexity measures showed significant differences with very large effect sizes, indicating that CLIL students' vocabulary is more sophisticated (longer words and less frequent words) and diverse than non-CLIL students' (WL: $F(1, 62) = 20.54$, $p = < .001$, $\eta_p^2 = .25$; WF: ($F(1, 62) = 15.38$, $p = < .001$, $\eta_p^2 = .20$; D: ($F(1, 62) = 7.25$, $p = .009$, $\eta_p^2 = .11$) (see Figure 2).

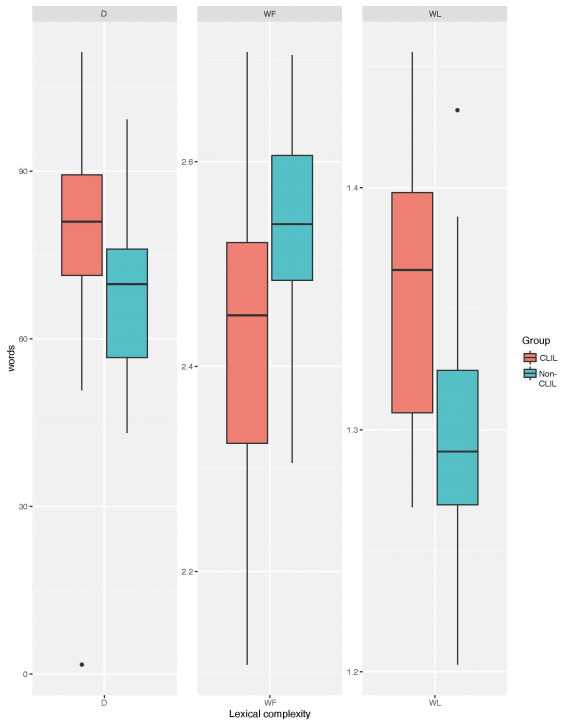

Figure 2. Comparison of lexical complexity measures in CLIL and non-CLIL students' writing

Contrary to the other linguistic measures, semantic cohesion was higher in the non-CLIL writers' language than in the CLIL writers' ($F(1, 62) = 8.35$, $p = .005$, $\eta_p^2 = .12$); in other words, the non-CLIL writers used more semantically related words than the CLIL writers to achieve semantic cohesion (see Figure 3).

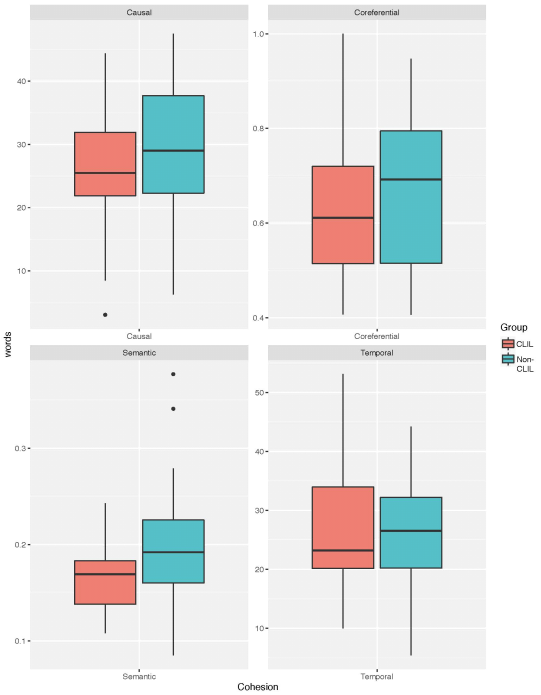

Figure 3. Comparison of cohesion in CLIL and non-CLIL students' writing

To assess the CLIL and non-CLIL students' writing quality, descriptive statistics on the total scores and the scores for content, organization, vocabulary, language use, and mechanics were calculated (see Table 5). The average total score of CLIL students was 79.11 (*SD* = 4.12) whereas that of non-CLIL students was 72.66 (*SD* = 1.03), showing that the CLIL students earned higher scores in writing quality than non-CLIL students.

Table 5. Descriptive statistics of CLIL and non-CLIL students' writing quality

	CLIL (*N* = 29)		Non-CLIL (*N* = 35)	
	M (*SD*)	95% CI	*M* (*SD*)	95% CI
Total	79.11 (4.12)	77.54, 80.68	72.65 (6.09)	70.56, 74.75
Content	17.90 (1.17)	17.45, 18.34	15.99 (1.84)	15.35, 16.62
Organization	17.72 (1.33)	17.22, 18.23	16.09 (1.70)	15.50, 16.67
Vocabulary	17.52 (1.01)	17.13, 17.90	16.04 (1.37)	15.57, 16.52
Language use	17.22 (1.05)	16.83, 17.62	15.89 (1.40)	15.41, 16.37
Mechanics	8.75 (.66)	8.49, 9.00	8.66 (.60)	8.45, 8.87

Table 6. CLIL effects on writing quality (MANOVA)

		CLIL		
		F	p	η_p^2
Total		23.543	<.001***	.28
Content		23.400	<.001***	.27
Organization		17.830	<.001***	.22
Vocabulary		23.002	<.001***	.27
Language use		18.096	<.001***	.23
Mechanics		.342	.561	.01
Wilks' lambda	.661	5.954	<.001***	.34

*p < .05; **p < .01; ***p < .001

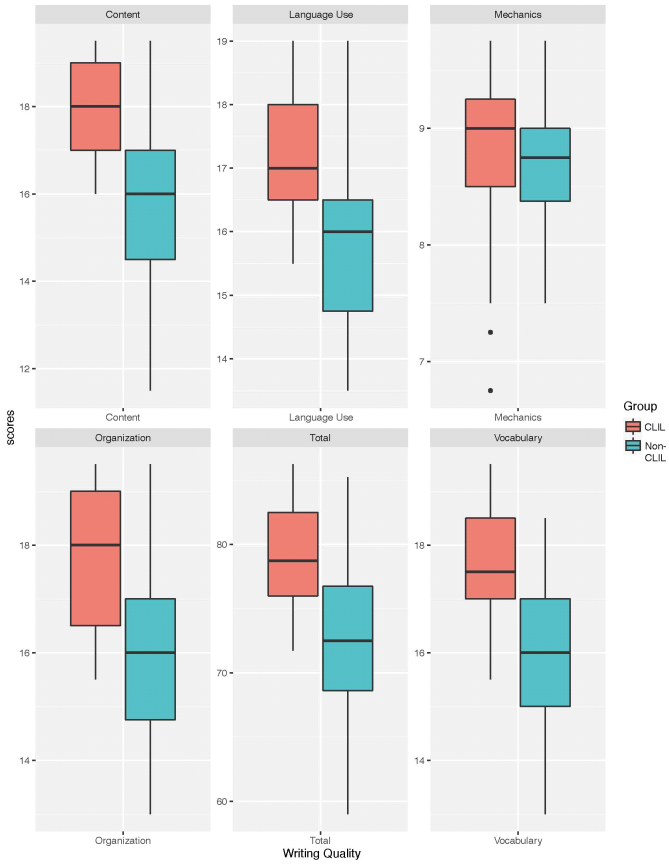

Figure 4. Comparison of writing quality in CLIL and non-CLIL students' writing

As Table 6 shows, there is a significant CLIL effect on total writing quality with a very large effect size (F(1, 62) = 21.22, p = < .001, η_p^2 = .027). The total score and subscale scores on content (F(1, 62) = 23.40, p = < .001, η_p^2 = .27), organization (F(1, 62) = 17.83, p = < .001, η_p^2 = .22), vocabulary (F(1, 62) = 23.00, p = < .001, η_p^2 = .27), and language use (F(1, 62) = 18.10, p = < .001, η_p^2 = .23) demonstrated significant differences with large effect sizes, indicating that CLIL students' narrative writing scored higher in each of these categories than that of non-CLIL students. However, for mechanics, no significant difference was detected, indicating that CLIL students and non-CLIL students received similar scores on mechanics (see Figure 4).

5. Discussion

The aims of this study were twofold. First, it intended to determine the effects of CLIL on syntactic complexity, lexical complexity, and cohesion in writing. Second, it aimed to discover whether CLIL affects writing quality.

The analysis found effects of CLIL on syntactic complexity in one length-of-production measure and two phrasal complexity measures (Navés 2011), with the CLIL students producing longer sentences and more complex phrases than their non-CLIL peers. Although the two groups of students had the same English proficiency level, their writing performance varied, possibly due to the language exposure or language experience that CLIL affords (Martinez Agudo 2019; Merino and Lasagabaster 2018). Perhaps owing to the amount of input obtained in an immersed environment, the CLIL students developed their language skills in extending their sentences as well as in constructing complex sentences. As noted by Navés (2011), the syntactic complexity measures were used to determine the effect of the pedagogical intervention (i.e., CLIL) and the development of language proficiency. The findings of this study support previous research that found that CLIL can help students to produce more complex sentences than non-CLIL students because CLIL provides a great deal of input in natural settings, which can drive the CLIL students to improve their language skills (e.g., Merino and Lasagabaster 2018). In addition, the CLIL

students likely had more opportunities in that curriculum to practice writing tasks than did the non-CLIL students, which may have contributed to the differences in length-of-production-unit and phrasal complexity measures.

Next, the study found greater lexical complexity in the narrative essays of the CLIL students than in those of the non-CLIL students. In other words, the CLIL students produced a more diverse and sophisticated vocabulary than the non-CLIL students. The findings also indicate that the CLIL students used less common vocabulary and showed more lexical variation in their writing than the non-CLIL students. This result is similar to the findings of previous studies (e.g., Jexenflicker and Dalton-Puffer 2010; Navés 2011; Roquet and Perez-Vidal 2017), suggesting that CLIL students have significant advantages in lexicon and grammar as shown by their use of a more diverse vocabulary and by their accurate language usage. As noted by Jexenflicker and Dalton-Puffer (2010), CLIL students may not resort to simple expressions and tend not to be affected by their L1. Because the CLIL students receive ample inputs in their curriculum and have many opportunities to practice their writing, they can develop a diverse and refined vocabulary appropriate to writing. This finding supports Merino and Lasagabaster's (2018) claim that CLIL students produce a more sophisticated, diverse vocabulary because they receive more language input than non-CLIL students.

Unlike in the other measures, greater semantic cohesion was found in the non-CLIL students' writing than in the CLIL students'. To achieve cohesion, the non-CLIL students used more related words in their writing than the CLIL students. Considering the fact that the non-CLIL students showed a less diverse vocabulary in their writing than the CLIL students, the non-CLIL students may have used related words repeatedly to extend their discourse. However, repeating related words did not necessarily produce good writing quality. This finding echoes the results of previous research (e.g., Crossley and McNamara 2010) that found that local cohesion does not correlate or correlates negatively with writing quality. In this study, too, non-CLIL students' writing earned higher scores for semantic cohesion, but their overall writing quality scored lower than the CLIL students' writing.

Writing quality, in fact, was found to differ significantly between the two groups according to the scores in all categories except mechanics. This solid

evidence of CLIL's effects corroborates previous research (Jexenflicker and Dalton-Puffer 2010; Roquet and Perez-Vidal 2017; Ruiz de Zarobe 2010; Vidal and Jarvis in press; Whittaker et al. 2011) as the CLIL students outperformed the non-CLIL students in terms of content, organization, vocabulary, and language usage. As shown in previous studies, studying through CLIL for about 1 year seemed to improve students' writing quality. As for the lack of difference in mechanics between the two groups, it is possible that the L2 proficiency level of all the participants was sufficiently high that they were easily able to avoid spelling and punctuation errors. It is worth noting, however, that previous research used a rubric (Jacob et al. 1981) that assigned the same score (20 points) to every category, thus potentially amplifying the effects of any differences in mechanics between the groups compared to this study's rubric (Connor-Linton and Polio 2014), which assigns a smaller number of points to mechanics (10 points).

Overall, the findings suggest that the greater writing skills of CLIL students may result from greater general English writing ability. Although the CLIL and non-CLIL students' English proficiency was controlled, the linguistic complexity and writing quality measures demonstrate the benefits of the CLIL curriculum to the CLIL students' writing. Because the CLIL students use the target language to learn the content and because they receive quality input and ample opportunities for output, they may show more developed language in written performance than their non-CLIL peers (e.g., Navés 2011). In addition, the CLIL curriculum enables the learners to practice language in natural settings, so the students can demonstrate a varied and sophisticated vocabulary. The writing task in this study was a narrative, which is not specific to a content subject, and was designed to demonstrate only the language skills of the two groups of students. The results concur with those of previous research (e.g., Dalton-Puffer 2011), suggesting that CLIL students can manage not only lexical resources, but also morphosyntactic structures that lead to more elaborate syntactic structure and sophistication in their writing.

6. Conclusion

This study contributes to the research on CLIL by investig ating EFL students' written production to determine the effect of CLIL on written production in terms of various linguistic measures and writing quality. The study suggests that the differences between the CLIL and non-CLIL students' written narratives can be attributed to curricular differences: the CLIL students' curriculum provides more diverse language experiences, which are reflected in the greater linguistic complexity and writing quality of the CLIL group's narrative essays.

In addition, similar to the results of previous research that explored the effect of CLIL on written production, this study's results, demonstrating the effectiveness of CLIL on writing, support the hypotheses that CLIL can provide L2 learners with intense input as well as quality time to develop the target language in written production. While in the CLIL program, the L2 learners seemed to develop writing skills with complex structures and vocabulary. The findings also provide empirical evidence that the CLIL approach better facilitates written proficiency, one of the goals of CLIL, than traditional foreign language teaching.

One limitation of this study is the lack of a pre-test, which may have more clearly shown the effects of CLIL. However, by controlling the students' English proficiency and L1, this study was able to focus solely on their writing performance to provide a valid explanation of the effects of CLIL on written language. In addition, longitudinal investigation of CLIL can provide more valid results regarding the effectiveness of CLIL on written performance. Despite these limitations, the findings contribute to CLIL research, which had previously focused only on morphosyntactic features or writing quality, by expanding it to investigate CLIL's effects on syntactic complexity, lexical complexity, and cohesion in addition to writing quality.

References

Ai, Haiyang and Xiaofei Lu. 2013. A corpus-based comparison of syntactic complexity in

NNS and NS university students' writing. In Ana Díaz-Negrillo, Nicolas Ballier, and Paul Thompson (eds.), *Automatic treatment and analysis of learner corpus data*, 249–264. Amsterdam: John Benjamins.

Bae, Jungok. 2001. Cohesion and coherence in children's written English: Immersion and English-only classes. *Issues in Applied Linguistics* 12: 51–88.

Brown, Gavin T. L, Kath Glasswell, and Don Harland. 2004. Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing* 9(2): 105–121.

Bruton, Anthony. 2013. CLIL: Some of the reasons why… and why not. *System* 41(3): 587-597.

Cenoz, Jasone. 2015. Content-based instruction and content and language integrated learning: The same or different?. *Language, Culture and Curriculum* 28(1): 8-24.

Cohen, Jacob. 1969. *Statistical power analysis for the behavioral sciences*. Academic Press.

Connor-Linton, Jeff and Charlene Polio. 2014. Comparing perspectives on L2 writing: Multiple analyses of a common corpus. *Journal of Second Language Writing* 26: 1–9.

Coyle, Do, Philip Hood, and David Marsh. 2010. *CLIL: Content and language integrated learning*. Cambridge, UK: Cambridge University Press.

Crossley, Scott. A, Tom Cobb, and Danielle S. McNamara. 2013. Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System* 41(4): 965–981.

Crossley, Scott. A, Kristopher Kyle, and Danielle S. McNamara. 2016. The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing* 32: 1–16.

Crossley, Scott. A, and Danielle S. McNamara. 2010. Cohesion, coherence, and expert evaluations of writing proficiency. In Stellan Ohlsson and Richard Catrambone (eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 984–989. Cognitive Science Society.

Dalton-Puffer, Christiane. 2009. Communicative competence and the CLIL lesson. In Yolanda Ruiz de Zarobe and Rosa Maria Jimenez Catalán (eds.), *Content and language integrated learning: Evidence from research in Europe*, 197–214. Multilingual Matters.

Dalton-Puffer, Christiane. 2011. Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics* 31: 182–204.

Dalton-Puffer, Christiane, Tarja Nikula, and Ute Smit. 2010. Language use and language learning in CLIL: Current findings and contentious issues. In Christiane Dalton-Puffer, Tarja Nikula, and Ute Smit (eds.), *Language use and language learning in CLIL classrooms*, 279–291. John Benjamins.

Gallardo-del-Puerto, Francisco, María Basterrechea, and María Martínez-Adrián. in press. Target language proficiency and reported use of compensatory strategies by young CLIL learners. *International Journal of Applied Linguistics*.

García Mayo, María del Pilar, and María de los Ángeles Hidalgo. 2017. L1 use among young EFL mainstream and CLIL learners in task-supported interaction. *System* 67: 132–145.

Gass, Susan. 1997. *Input, interaction and the second language learner*. New York: Routledge.

Jexenflicker, Silvia, and Christiane Dalton-Puffer. 2010. The CLIL differential: Comparing the writing of CLIL and non-CLIL students in higher colleges of technology. In Christiane Dalton-Puffer, Tarja Nikula, and Ute Smit (eds.), *Language use and language learning in CLIL classrooms*, 169–190. Amsterdam: John Benjamins.

Jacobs, L. Holly, Stephen A. Zinkgrap, Deanna Wormuth, V. Faye Hartfiel, and Jane B. Hughey. 1981. *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.

Kang, Jennifer Yusun. 2005. Written narratives as an index of L2 competence in Korean EFL learners. *Journal of Second Language Writing* 14(4): 259–279.

Kim, Hyun Jung, Junkyu Lee, and Hyun-Jo You. 2019. Analysis of rater effect in the evaluation of second language grammatical knowledge in the context of writing: Application of a generalized linear model. *Linguistic Research* 36(Special Edition): 25–57.

Knoch, Ute, Amir Rouhshad, Su Ping Oon, and Neomy Storch. 2015. What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing* 28: 39–52.

Krashen, D. Stephen. 1985. The input hypothesis: Issues and implications. Longman.

Lee, Younghee Cheri. 2018. The hallmarks of L2 writing viewed through the prism of translation universals. *Linguistic Research* 35(Special Edition): 171–205.

Liskin-Gasparro, Judith. 2000. The use of tense-aspect morphology in Spanish oral narratives: Exploring the perceptions of advanced learners. *Hispania* 83(4): 830–844.

Lo, Yuen Yi. 2014. Collaboration between L2 and content subject teachers in CBI: Contrasting beliefs and attitudes. *RELC Journal* 45(2): 181–196.

Lo, Yuen Yi and Victoria A. Murphy. 2010. Vocabulary knowledge and growth in immersion and regular language-learning programmes in Hong Kong. *Language and Education* 24: 215–238.

Long, Michael H. 1996. The role of the linguistic environment in second language acquisition. In Ritchie W. C. William and Tej K. Bhatia (eds.), *Handbook of second language acquisition*, 413–468. San Diego, CA: Academic Press.

Lorenzo, Francisco, Sonia Casal, and Pat Moore. 2010. The effects of content and language integrated learning in European education: Key findings from the Andalusian bilingual sections evaluation project. *Applied linguistics* 31(3): 418-442.

Lu, Xiaofei. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15(4): 474–496.

Martinez Agudo, J. D. D. 2019. Which instructional programme (EFL or CLIL) results in better oral communicative competence? Updated empirical evidence from a mono-

lingual context. *Linguistics and Education* 51: 69–78.

McCarthy, Philip M and Scott Jarvis. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing* 24(4): 459–488.

McNamara, Danielle S., Arthur C. Graesser, Philip M. McCarthy, and Zhiqiang Cai. 2014. *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press.

Merino, Jon Ander and David Lasagabaster. 2018. The effect of content and language integrated learning programmes' intensity on English proficiency: A longitudinal study. *International Journal of Applied Linguistics* 28(1): 18–30.

Navés, Teresa. 2011. How promising are the results of integrating content and language for EFL writing and overall EFL proficiency? *Linguistic Insights: Studies in Language and Communication* 108: 155–186.

Özyıldırım, Işıl. 2009. Narrative analysis: An analysis of oral and written strategies in personal experience narratives. *Journal of Pragmatics* 41: 1209–1222.

Roquet, Helena and Carmen Pérez-Vidal. 2017. Do productive skills improve in content and language integrated learning contexts? The case of writing. *Applied Linguistics* 38(4): 489–511.

Ruiz de Zarobe, Yolanda. 2008. CLIL and foreign language learning: A longitudinal study in the Basque country. *International CLIL Research Journal* 1: 60–73.

Ruiz de Zarobe, Yolanda. 2010. Written production and CLIL: An empirical study. In Christiane Dalton-Puffer, Tarja Nikula, and Ute Smit (eds.), *Language use and language learning in CLIL classrooms*, 191–212. John Benjamins.

Tan, May. 2011. Mathematics and science teachers' beliefs and practices regarding the teaching of language in content learning. *Language Teaching Research* 15: 325–342.

Vidal, Karina and Scott Jarvis. in press. Effects of English-medium instruction on Spanish students' proficiency and lexical diversity in English. *Language Teaching Research*.

Whittaker, Rachel, Ana Llinares, and Anne McCabe. 2011. Written discourse development in CLIL at secondary school. *Language Teaching Research* 15(3): 343–362.

Yu, Guoxing. 2010. Lexical diversity in writing and speaking task performances. *Applied Linguistics* 31(2): 236–259.

Jongbong Lee
Assistant Professor
Faculty of International Studies
Nagoya University of Commerce & Business
4-4, Sagamine Komenoki, Nisshin, Aichi, Japan 470-0193
E-mail: jongbong_lee@nucba.ac.jp

Received: 2020. 04. 28.
Revised: 2020. 06. 10.
Accepted: 2020. 06. 11.