

## Writing Development in Syntactic Complexity, Accuracy and Fluency in a Content and Language Integrated Learning Class

Hsuan-Yu Tai

Institute of Linguistics  
National Chung Cheng University  
168 University Rd., Min-Hsiung  
Chia-Yi 62102, Taiwan

### Abstract

*This study aimed to explore L2 writing development (as measured by syntactic complexity, accuracy and fluency) in a CLIL class by examining authentic texts from the class. 19 participants joined an 18-week CLIL class and a total of 57 written assignments from the beginning, the middle, and the final phase of the course were collected. By analyzing syntactic complexity, syntactic accuracy, and fluency across three different times, the findings revealed that the participants improved their accuracy and fluency but not complexity, suggesting that a CLIL class was beneficial for L2 writing improvement to some extent. Possible accounts for accuracy and fluency improvement include preference for Standard English, assessment criteria of the course, and practice effects. Further pedagogical implications were provided in response to the findings.*

**Keywords:** second language writing, complexity, accuracy, fluency, content-and-language-integrated-learning, CLIL

### 1. Introduction

Content and language integrated learning, content-based second language teaching or English as a medium of instruction refers to a pedagogical approach that non-linguistic materials such as mathematics or history are taught through a language that is not students' first language (Lyster, 2011). It is assumed that while students are learning the content, they can also absorb the language simultaneously (Krashen, 1981) and studies have shown that language development was seen as taking-for-granted outcome after attending a CLIL class (Arnó-Macià & Mancho-Barés, 2015) even if there was little explicit instructions on language (Dalton-Puffer, 2008). Nonetheless, to what extent students can successfully obtain language and content knowledge together is still a debatable issue among applied linguists. In terms of language learning, the assumption that CLIL can facilitate language proficiency growth is supported by much research. Numerous empirical studies have compared the effectiveness of language learning between CLIL and non-CLIL classes in terms of morphosyntax in speaking, receptive vocabulary (Catalan & Zarobe, 2009; Ibarrola, 2011; Olaizola & Mayo, 2009) and writing (Jexenflicker & Dalton-Puffer, 2010; Ruiz de Zarobe, 2010), claiming that CLIL created a better context for language learning than non-CLIL class. However, only few studies have paid attention to the language development within a CLIL class. Of those studies, few have compared students' listening, grammar and reading by using pre-/post-tests (Aguilar & Muñoz, 2014; Goris & Verhoeven, 2013; Kung, 2013), but research on writing development was still scarce. What's more, research on the effectiveness of CLIL classes so far all assessed learners by adopting tasks that were irrelevant to the class and were in a controlled setting such as narrating a story, writing an email or completing grammar tests. To what extent these findings could reflect language development that associates with CLIL classes are in question. Given these lacunas, the current study served to contribute to the existing body of literature and aimed to investigate L2 writing development within a CLIL class by examining students' authentic written texts in association with the content of the class in a natural setting.

### 2. Complexity, Accuracy and Fluency as Measures in L2 Writing Development Research

In the research on L2 writing development, a plethora of studies have used syntactic complexity, accuracy and fluency as objective measures. Casaneve (1994) explored 16 Japanese EFL intermediate level students' journals in three semesters. The researcher selected two journals from the first, and the end of the first semester.

Also, two journals were chosen at the end of the third semester. Data were analyzed based on T-unit length, complexity (number of clauses per T-unit and percentage of complex T-unit) and accuracy (percent of error-free T-unit and length of error-free T-unit). The quantitative results showed that participants demonstrated longer, more complex and more accurate production when they were seen as a group. In addition, Henry (1996), in an attempt to investigate early writing development, analyzed 67 autobiography essays produced by beginning and intermediate levels of American students who learned Russian. Fluency (total of words and mean T-units length) and accuracy (the number and percentage of correct T-units) were used as measurements. Essays were written in a timed condition without any references to dictionaries. The findings showed that there were no significant differences in terms of accuracy at both level, suggesting that accuracy was not an indicator for differentiating learners' proficiency level. However, the participants showed great improvement in fluency. Storch (2009) investigated whether 25 non-native speakers of English who studied at a University in Australia improved their fluency, accuracy, grammatical complexity, lexical complexity, essay structures and content throughout a semester. The participants were required to produce an essay on a diagnostic test and a pretest-posttest design was used to compare their writing performance. All of the essays were analyzed qualitatively and quantitatively. With regard to the quantitative analyses, grammatical complexity was measured by the ratio of clauses per T-unit (C/T) and the ratio of dependent clauses per clause (DC/C). Fluency was measured by total number of words, T-units and words per T-units (W/T). Accuracy was measured by number of errors per words (E/W), error-free T-units per T-units, and error-free clauses for clauses. The findings show that grammatical complexity, accuracy and fluency did not improve significantly; however, their range of vocabulary, content and rhetoric were better after a semester.

Similar results were found in Deng, Lee, Chitra and Lim's (2010) study where they explored writing development of 31 Chinese speakers of English in a postgraduate EAP course in Singapore. The participants were considered high-intermediate level according their English proficiency test scores for admission. Participants' written products were analyzed according to fluency (total number of words and word per T-unit), accuracy (total number of errors per total words) and academic vocabulary use, and text structure. The findings demonstrated that there was no significant improvement in accuracy, fluency and vocabulary use. However, improvement was found in text organization and content, though content did not reach statistical significance. More recently, Beers and Nagy (2011) conducted a longitudinal study to explore linguistic and discursive literacy development. Participants were divided into two cohorts. One cohort consisted of 83 native speakers of English from grades three and five and the other consisted of 96 speakers of English from grades five and seven. Four genres of essays (narrative, descriptive, compare/ contrast and persuasive) were given on two occasions two years apart to participants. Data were analyzed by means of clauses per T-unit, words per clause and text length in words. The findings showed that there was no main effect of grade but genre effects. For example, persuasive essays had more clauses per T-unit and narrative essays had more words per clause. Further qualitative analyses demonstrated that persuasive essays had more subordination and thus led to higher rate of clauses.

With regard to factors that influence syntactic complexity, fluency and accuracy, Tedick (1990) examined the effects of subject-knowledge on writing performance among three levels of participants. The results, analyzed by T-units, revealed that different course levels showed significant differences in several dependent variables such as mean-length of T-units. Most importantly, participants performed better in response to the field-specific topic than the general topic and hence the researcher concluded that topic familiarity played a part in writing performance. Furthermore, Ishikawa (1995) conducted a longitudinal study on the effects of two writing tasks (guided writing and free writing) on 24 objective measures in relation to T-units and clauses among low-proficiency EFL learners. 28 and 29 Japanese college level students in two separate classes participated in this research. The researcher concluded that a guided writing task was more effective in accuracy improvement. Finally, aside from task effects on writing performance, Lu (2011) also found that institution, genre and timing condition timing had impacts on syntactic complexity in his large sample written corpus data of Chinese learners. To sum up, syntactic complexity, fluency and accuracy have been conventionalized as objective measures for L2 writing development in ESL or EFL settings where English learning was the primary concern. Nevertheless, one special educational context, content and language integrated learning (CLIL) classes is still under-researched. In other words, little attention is given to whether L2 writing will improve within such context where content normally serves as the primary focus. Therefore, in order to fill the gap, there is a need to investigate whether students' syntactic complexity, accuracy and fluency in their writing increase in a CLIL class. The research questions thus are generated as follows.

### 3. Research questions

- 3.1 Do participants in a CLIL class increase their syntactic complexity (as measured by the ratio of clause per T-unit (C/T), the ratio of dependent clause per T-unit (DC/T) and the holistic rating scale) throughout the course?
- 3.2 Do participants in a CLIL class increase their syntactic accuracy (as measured by the proportion of error-free T-units (EFT/T), the proportion of error-free clauses (EFC/C), and the total number of errors per total number of words (E/W)) throughout the course?
- 3.3 Do participants in a CLIL class increase their fluency (as measured by the total number of words (W), the number of T-units, and the length of the T-units measured in words per T-unit (W/T)) throughout the course?

### 4. Methodology

#### 4.1 Research Site

Data were collected from an elective CLIL course entitled “Intercultural Communication: Demystifying Courtroom Conversations” taught by a linguistic professor in the first semester in 2012. The course was a general class opens to students from all of the departments. In the class, a legal case was introduced each week and students would learn the association between the case and some linguistic knowledge applied to it. In addition, the class was conducted in English.

#### 4.2 Participants

Initially, there were 26 participants (18 females and 8 males) in this study. However, one participant only submitted one assignment throughout the whole semester and thus he was eliminated from the data. Also, six participants lacked one or two scores from the three assignments. Therefore, they were also excluded from the data. The remaining number of participants was 19 (5 males, 14 females). The English proficiency of the participants was between intermediate to advance. Six had passed intermediate, four had passed high-intermediate and three had passed advance level on GEPT. Two obtained 750-900, one obtained 560-750 and one obtained 900-990 on TEOIC. The rest two reported their proficiency as intermediate. Also, nine of them are English major. 2 majored in adult education, computer science and law. There were also one business administration, finance, mass communication, and engineering student.

#### 4.3 Data Collection

Participants were asked to write weekly assignments in response to class contents. Participants were informed that their assignments would be scored based on organization, content, and accuracy (punctuation, spelling and glaring grammatical errors that impede communication). In order to investigate whether participants' complexity, accuracy and fluency increased throughout the semester, assignments from the beginning, the middle and the last week of the semester were selected for comparisons. Ultimately, a total of 57 assignments were chosen.

#### 4.4 Data Analysis

A T-unit is defined as “one main clause plus whatever subordinate clauses happen to be attached to or embedded within it (Hunt, 1966, p. 735)”. Clauses were separated into dependent and independent clauses. An independent clause is defined as one that can be used on its own (Richards, Platt, & Platt, 1992) and a dependent clause is defined as one that has a finite verb and a subject, including adverbial, nominal, and adjectival clauses (Wolfe-Quintero, Inagaki, & Kim, 1998). Following Casaneve (1994) and Storch (2009), complexity was analyzed by the ratio of clauses per T-unit (C/T) and the ratio of dependent clauses per T-unit (DC/T). Accuracy was measured by the proportion of error-free T-units (EFT/T), the proportion of error-free clauses (EFC/C), and the total number of errors per total number of words (E/W). Errors here included syntactic ones (word order, incomplete sentence), morphology (tense, agreement, use of articles) and errors in word choice. Errors in spelling and mechanics such as punctuations were ignored. Fluency was measured by the total number of words (W), the number of T-units, and the length of the T-units measured in words per T-unit (W/T). Besides the research, another rater was recruited for coding. The inter-coder reliability reached .92 for identification of T-units and .97 for identification of clauses. As for identification of error-free clauses and T-units, the reliability reached .91 and .93 respectively.

### 5. Results

The current study investigated whether syntactic complexity, accuracy and fluency improved in a CLIL Class throughout a semester.

The findings indicated that participants' syntactic complexity did not show significant increase. However, their accuracy and fluency showed significant improvement over time. The detailed results are as follows.

### **5.1 Do Participants in an EMI class Increase their Syntactic Complexity throughout the Course?**

Table 1 indicated that there were no significant differences in the ratio of clause per T-unit ( $F(2, 36) = 2.621$ ,  $p > .05$ ,  $\eta^2 = .127$ ) and the ratio of dependent clause per T-unit ( $F(2, 36) = 2.928$ ,  $p > .05$ ,  $\eta^2 = .14$ ) of the beginning, the middle and the last of the semester, suggesting that the participants' syntactic complexity did not increase over time.

### **5.2. Do Participants in an EMI class Increase their Syntactic Accuracy throughout the Course?**

In terms of accuracy, repeated-measures one-way ANOVA showed that there were no significant differences in the proportion of error-free T-units ( $F(2, 36) = 2.599$ ,  $p > .05$ ,  $\eta^2 = .485$ ). However, a further examination on the proportion of error-free clauses revealed that there were significant differences across different times ( $F(2, 36) = 4.007$ ,  $p < .05$ ,  $\eta^2 = .679$ ). As shown in Table 2, pair wise comparisons showed that Time 2 and Time 3 were significantly higher than Time 1, but there were no significant differences between Time 2 and Time 3. In addition, the results of the total number of errors per total number of words also indicated significant differences across times ( $F(2, 36) = 5.236$ ,  $p < .05$ ,  $\eta^2 = .685$ ). Pair wise comparisons revealed that the error rate of Time 3 was lower than Time 1, but there were no significant differences between Time 3 and Time 2 as well as between Time 1 and Time 2. The results thus indicated improvement in syntactic accuracy over time.

### **5.3. Do Participants in an EMI Class Increase their Fluency throughout the Course?**

Table 3 revealed that there were significant differences in the number of T-units ( $F(2, 36) = 3.822$ ,  $p < 0.5$ ,  $\eta^2 = .657$ ) and Time 2 was significantly higher than Time 1 and Time 3. In terms of the total number of words, the results indicated that there were significant differences among the three times ( $F(2, 36) = 6.051$ ,  $p < 0.5$ ,  $\eta^2 = .857$ ). Specifically, Time 2 and 3 had significantly more words than Time 1. With regard to W/T, the results demonstrated that there were also significant differences across times ( $F(2, 36) = 8.919$ ,  $p < 0.5$ ,  $\eta^2 = .961$ ). Pairwise comparisons indicated that Time 3 was significantly higher than Time 2 and Time 1.

## **6. Discussion**

In response to the first research question, the findings were in accordance with Storch's (2009) research where syntactic complexity did not improve after one semester. The lack of improvement of complexity might due to the relatively short time period of the study. In a meta-analysis conducted by Ortega (2003), syntactic complexity might require up to 12 months of college-level instruction to develop. It could also be argued that the participants were already fairly advanced and thus improvement for them might be harder or take longer to achieve (Green, 2004). Furthermore, topic familiarity might also play a part in the lack of complexity improvement (Tedick, 1990). The participants might produce more complex sentences because the topics were more familiar. Furthermore, the goal of this CLIL class did not aim to improve syntactic complexity. Therefore, it is reasonable that the participants did not show any increase.

With respect to the second research question, the findings, in consistence with Casaneve (1994), revealed that participants did improve on their accuracy throughout the semester. Despite the fact that the goal of this CLIL class was not the pursuit of accuracy, several factors could be accountable for the findings. First, the participants might have preference for the native English grammatical norm. He (2010) conducted a large scale survey on Chinese students and the findings demonstrated that the majority of the informants preferred Standard English with regard to grammar. In addition, Kirkpatrick and Xu (2002) found that there were significant differences in attitude between English and other majors, with female English major students to be the most likely to pursue Standard English. Thus, in the current study, over half of the participants were female English major; their attitude might result in the increase in accuracy. Third, the assessment criteria in this class might also have impacts on their accuracy improvement. The assessment criteria stated that they would be scored based on their organization and they should not make glaring grammatical errors that impeded communication. Though the initial attempt was to encourage the students to double check their writing so that misunderstanding would not occur, it was likely that the students directed their attention to grammar. Finally, their improvement in accuracy might be because of practice effects. Specifically, the practices that the participants were systematically engaged in could help them develop their second language writing (DeKeyser, 2007) so that they could write more academically and accurately.

Consequently, their accuracy showed significant improvement. In terms of the third research question, the findings revealed that the participants produced more fluent texts after a semester of the CLIL class. In other words, the number of their T-units and words were much longer. A possible account for this increase might be also practice effects in that repetition of tasks could help learners produce more fluent output because a part of conceptualization, formulation and articulation was stored in the memory and could be used when encountering a similar task for the second time (Bygate, 2001).

### 7. Conclusion

The current study investigated whether students' L2 writing would improve after they participated in an 18-week CLIL class. The findings showed that except for syntactic complexity, the participants' syntactic accuracy and fluency improved over time. The lack of improvement in syntactic complexity might be because of the relatively short length of the current study, relatively high English proficiency of the participants or other factors such as topic familiarity. On the other hand, the preference to Standard English, the influences from the assessment criteria of the course, and practice effects are assumed to be possible explanations for the increase of syntactic accuracy and fluency. Some implications are thus generalized from the findings. From the second language writing perspective, while most of the previous research was conducted in an EFL or ESL setting and the data were collected in testing or timed conditions (Henry, 1996; Storch, 2009; Tedick, 1990), the results indicated that even though a CLIL class did not emphasize forms and the writing was completed in a natural setting, students could absorb linguistic knowledge while they learned the content simultaneously. In other words, the findings of the study support the effectiveness of CLIL classes on language learning, at least on syntactic accuracy and fluency improvement. In addition, provision of ample opportunities to produce written output could be beneficial for students in a CLIL class as practice plays a role in writing development. Finally, although the present research has shown that CLIL classes could be facilitative in L2 writing, one should be cautious that the participants were treated as a homogeneous group. Thus, individual differences in a CLIL class cannot be revealed. Future studies are suggested to adopt a more in-depth analysis and include some personal factors to account for such differences.

**Table 1: Comparisons of Syntactic Complexity over Time**

	Mean	SD	Sig.
C/T	1.91	.299	n.s.
Time 1			
Time 2	1.75	.27	n.s.
Time 3	1.68	.36	n.s.
DC/T			
Time 1	.87	.25	n.s.
Time 2	.75	.25	n.s.
Time 3	.67	.29	n.s.

$p > .05$

**Table 2: Comparisons of Accuracy over Time**

	Mean	SD	Sig.
EFC/C			
Time 1	.67	.174	
Time 1-Time 2			.032*
Time1- Time 3			.058
Time 2	.78	.134	
Time 2- Time 1			.032*
Time 2- Time 3			.452
Time 3	.76	.108	
Time 3- Time 1			.042*
Time 3- Time 2			.452
EFT/T			
Time 1	.58	.20	n.s.
Time 2	.69	.20	n.s.
Time 3	.67	.15	n.s.
E/W			
Time 1	.039	.024	
Time 1-Time 2			.055
Time1- Time 3			.011*
Time 2	.025	.017	
Time 2- Time 1			.055
Time 2- Time 3			.426
Time 3	.023	.009	
Time 3- Time 1			.011*
Time 3- Time 2			.426

p&lt;.05

**Table 3: Comparisons of Fluency over Time**

	Mean	SD	Sig.
Total of T-units			
Time 1	25	6.62	
Time 1-Time 2			.46*
Time1- Time 3			.862
Time 2	30.16	9.42	
Time 2- Time 1			.046*
Time 2- Time 3			.022*
Time 3	24.63	5.87	
Time 3- Time 1			.862
Time 3- Time 2			.022*
Total of Words			
Time 1	384.63	86.11	
Time 1-Time 2			.03*
Time1- Time 3			.00*
Time 2	442.16	107.97	
Time 2- Time 1			.03*
Time 2- Time 3			.567
Time 3	455.95	84.10	
Time 3- Time 1			.00*
Time 3- Time 2			.567
W/T			
Time 1	15.95	4.46	
Time 1-Time 2			.358
Time1- Time 3			.017*
Time 2	15.03	2.22	
Time 2- Time 1			.358
Time 2- Time 3			.000*
Time 3	18.83	2.33	
Time 3- Time 1			.017*
Time 3- Time 2			.000*

p&lt;.05

## References

- Aguilar, M., & Muñoz, C. (2014). The effect of proficiency on CLIL benefits in engineering students in Spain. *International Journal of Applied Linguistics*, 24(1), 1-18.
- Arnó-Macià, E., & Mancho-Barés, G. (2015). The role of content and language in content and language integrated learning (CLIL) at university: Challenges and implications for ESP. *English for Specific Purposes*, 37, 63-73.
- Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: syntactic complexity and genre differentiation. *Reading and Writing: An Interdisciplinary Journal*, 24, 183-202.
- Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp.23-48). London: Longman.
- Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing*, 3(3), 179-201.
- Catalan, R. M. J., & Zarobe, Y. R. (2009). The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL instruction. In Y. R. Zarobe & R. M. J. Catalan (Eds.), *Content and language integrated learning: Evidence from research in Europe*. Buffalo: Multilingual Matter.
- Dalton-Puffer, C. (2008). Outcomes and processes in content and language integrated learning (CLIL): current research from Europe. In W. Delanoy & L. Volkman (Eds.), *Future perspectives for English language teaching* (pp.139-157). Heidelberg: Carl Winter.
- DeKeyser, R. (2007). Introduction: Situating the concept of practice. In R. DeKeyser (Eds.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp.1- 18). Cambridge: Cambridge University Press.
- Deng, X., Lee, K. C., Chitra, V., & Lim M. L. (2010). Academic writing development of ESL/EFL graduate students in NUS. *Reflections on English Language Teaching*, 9(2), 119-138.
- Goris, J., Denessen, E., & Verhoeven, L. (2013). Effects of the content and language integrated learning approach to EFL teaching. *Written Language & Literacy*, 16(2), 186-207.
- Green, T. (2004). Making the grade: Score gains on the IELTS writing test. *Research notes*, 16, 9-13
- He, D.Y. (2010). Native speaker norms and china English: From the perspective of learners and teachers in china. *TESOL QUARTERLY*, 44(4), 769-789.
- Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university-level students of Russian. *The Modern Language Journal*, 80(3), 309-326.
- Hunt, K. (1966). Recent measures in syntactic development. *Elementary English*, 43, 732-739
- Ibarrola, A. L. (2011). Faster and further morphosyntactic development of CLIL vs. EFL Basque-Spanish bilinguals learning English in high-school. *International Journal of English Studies*, 12(1), 79-96.
- Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, 4(1), 51-59.
- 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). Amsterdam: John Benjamins.
- Kirkpatrick, A., & Xu, Z. (2002). Chinese pragmatic norms and "China English." *World Englishes*, 21, 269-279.
- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.
- Kung, F. W. (2013). The more the merrier? Bilingualism in an academic perspective: Exploring the implementation of English-medium instruction in Taiwanese tertiary education. *Asian EFL Journal*, 15(4), 8-36.
- Lyster, R. (2011). Content-based second language teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*, Vol. 2 (pp. 611-630). New York: Routledge.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices as college-level ESL writer language development. *TESOL Quarterly*, 45(1), 36-62.
- Olaizola, I. V., & Mayo, M. P. G. (2009). Tense and agreement morphology in the interlanguage of Basque/Spanish bilinguals: CLIL versus non-CLIL. In Y. R. Zarobe & R. M. J. Catalan (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp.157-175). Buffalo: Multilingual Matter.

- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24, 492–518.
- Ruiz de Zarobe, Y. (2010) Written production and CLIL. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp.191-209). Amsterdam: John Benjamins.
- Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, 18, 103-118.
- Tedick, D. J. (1990). ESL writing assessment: Subject- matter knowledge and its impact on performance. *English for Specific Purposes*, 9, 123-143.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu: University of Hawaii at Manoa.