

# **A study of writing development in a CLIL legal course**

**Hsuan-yu Tai**

Intergrams 15.2(2015):

<http://benz.nchu.edu.tw/~intergrams/intergrams/152/152-tai.pdf>

ISSN: 1683-4186

## **Abstract**

Unlike previous studies that placed their primary focus on the comparison between CLIL and non-CLIL classes, this study aimed to explore L2 writing development in a CLIL class by examining authentic texts from the class. 19 participants participated 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 corrective feedback on grammar, criteria of written assignments, 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

# 學科內容與語言綜合學習課程中

## 第二外語寫作句法複雜度、正確度與流暢度的發展研究

### 摘 要

有別於先前研究主要比較學科內容與語言綜合學習課程(CLIL)與非學科內容與語言綜合學習課程(non-CLIL)的差異，本研究旨在探討第二外語寫作在學科內容與語言綜合學習課程(CLIL)中，句法複雜度、正確度以及流暢度的發展情形。在為期十八週的課程當中，本研究分別蒐集了十九位英語作為外國語學生的學期初，學期中以及學期末的課後寫作，共五十七篇的作業進行分析，在比較學生學期三階段句法複雜度、正確度以及流利度之後，研究結果顯示學生正確度與流暢度有顯著進步，但複雜度並無顯著差異，此結果說明學科內容與語言綜合學習課程對於第二外語寫作有部分的助益，本研究並提出學生正確度與流暢度進步的原因可能是因為評分者對於寫作的糾正回饋、作業評量標準以及練習效應，根據研究結果，文章最後提出相關教學啟示以及方法。

**關鍵詞：**第二外語寫作、複雜度、正確度、流暢度、學科內容與語言綜合學習課程

## **1. Introduction**

Content and language integrated learning (CLIL) 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 also a "dual-focused educational approach that an additional language is used to teach and learn both content and language (Colye, Hood & Marsh, 2010). It is assumed that while students are learning the content, they can also absorb the language simultaneously (Dalton-Puffer, 2008) 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 such as Bruton (2011) who criticized that the positive effects in CLIL research was misleading due to selective participants, extra help from foreign language teachers and assistants and many other factors.

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; Lasagabaster, 2008; Llinares and Whittaker, 2010; Ruiz de Zarobe, 2010), claiming that CLIL created a better context for language learning than non-CLIL class. Despite the fact that previous research has shown that CLIL has been facilitative for language learning, the majority of the studies focused on the comparison of the effectiveness of CLIL and non-CLIL classes. Relatively little attention has been paid to language development, for instance, writing, within a CLIL class (Whittaker, Llinares & McCabe, 2011) at tertiary level. Accordingly, the current research focused on writing development within a CLIL class in one semester through examining syntactic complexity, accuracy and fluency in a university.

## **2. Writing performance and CLIL**

In the research on the writing development and CLIL, a number of studies have shown that CLIL yielded positive influences on students' writing. Lasagabaster (2008) compared speaking, writing, listening, and grammar competence between secondary students who enrolled in CLIL and non-CLIL courses. In terms of writing, the students were asked to write a letter to an English family and their letters were

evaluated based on holistic scales that focused on their content, organization, vocabulary, language usage and mechanics. The results showed that the students in CLIL courses outperformed non-CLIL students in writing. Moreover, third year CLIL students' writing scores were significantly higher than fourth year non-CLIL students' writing scores, indicating that CLIL courses had positive effects on language learning.

Similarly, Ruiz de Zarobe (2010) compared secondary level learners' writing performance between CLIL and non-CLIL classes by asking them to write a letter. The findings revealed that students' in CLIL classes outperformed those who were in non-CLIL class in a written production test in terms of content, vocabulary, organization, language usage and mechanics. Furthermore, the findings showed that CLIL offered a good context for more language exposure, which resulted in language improvement even though students were younger in CLIL classes than those in non-CLIL class.

Jexenflicker and Dalton-Puffer (2010) compared learners writing proficiency between 11<sup>th</sup> –grade students who received CLIL and non-CLIL instruction in Austria. By assessing their writing abilities through a free-writing task, it was found that students in CLIL class outperformed those in CLIL class in terms of grammar, vocabulary, orthographic skills, organization and communicative skills. In addition, the use of connectives was better in CLIL.

In an attempt to see what extent CLIL students could produce subject-specific discourse, Llinares and Whittaker (2010) compared written and spoken production of CLIL secondary school students of history and the students who followed the same syllabus in their first language. Based on Systemic Functional Linguistics, the authors examined how students presented their content. Specifically, they focused on semantic classes of verbs or process types (action, relational, existential, etc.), circumstances (time, place, manner, etc.) and clause complexes (coordination and subordination) as well as the use of modality (information concerning probability and usuality of events). In terms of written production, the students were asked to write a summary after a discussion of a topic for 20 minutes in the class with no resources available. The results showed that the both CLIL and L1 student could use expected process types for the tasks and genre. However, the CLIL students employed a wider variety of processes such as relational, possessive, and identifying in written register. Yet, different from the L1 students who used more diverse circumstances such as manner and cause, the CLIL students produced limited circumstances, primarily temporal and location. The findings thus indicated the L1 students had more awareness of the register of the discipline. In terms of clause complexes, the findings revealed that both L1 and CLIL students tended to use *addition* and *cause* to link clauses. Specifically, when the CLIL students used *cause* for clause-linking, they frequently expressed it

with a subordinate clause. Also, both L1 and CLIL students used elaboration, e.g. relative pronouns, as a clause-combining device. With respect to modality, it was found that the L1 students used only expression of obligation whereas the CLIL students used more expressions, including obligation, ability, and so forth.

Whittaker, Llinares and McCabe (2011) investigated written discourse development from history classes in two junior high schools over four years. Based on Systemic Functional Linguistics, the researchers analyzed students' nominal groups (identification system and tracking system) to examine whether they could create discourse that was both coherent to the register of the academic discipline and coherent in the texts they produced. The participants were required to write a topic that had been discussed few days ago in the class without any references available in 20 minutes. The findings of identification system showed that there was a slight drop in using nominal groups for presenting (introducing new participants such as *a*, *an*, or *one*); however, the nominal groups for presuming (mentioning already known participants such as *the soldier* or *these countries*) increased over time. This indicated that the participants' discourse has moved away from the simple listing of topics to a deeper discussion of topics. Also, the frequency of misuse of nominal groups decreased in the final two years. With respect to tracking system, there was an increase of direct reference and a decrease of exophora, which showed a more cohesive text and a more impersonal register. Finally, the results also revealed that the participants used less pronouns and unmodified nouns but more pre- and post-modified nominal groups in the final two years, showing that the participants progressively produced more complex phrases.

Yang (2015) explored English performance (receptive and productive skills), the relation between performance and language proficiency and learners and teachers' perception of a CLIL program at a polytechnic university in Taiwan longitudinally. With respect to English performance, the participants were assessed according to General English Proficiency Test (GEPT), a locally developed criterion-reference language testing system in Taiwan. The findings showed that the CLIL students did not reveal significant better writing performance in comparison with other test takers nationwide. The researcher thus attributed the results to the lack of practice of productive skills in the program.

Notwithstanding most previous research has shown that CLIL was beneficial for learners' writing, most studies focused on the comparison the effectiveness between CLIL and non-CLIL classes. Little attention was given to how writing develops within a CLIL class only (Whittaker, Llinares & McCabe, 2011). In addition, most previous research on writing performance and CLIL has been conducted in secondary schools; however, studies conducted within higher education are still scarce (Yang, 2015).

Therefore, the present study intended to investigate to what extent tertiary level students' writing performance improved over a semester of CLIL class. Specifically, this study examined accuracy, fluency, and complexity, measurements that have been extensively used in numerous studies to measure writing development (Casaneve, 1994; Henry, 1996; Ishikawa, 1995; Lu, 2011; Tedick, 1990), to see if students improved their writing performance over time in a CLIL class. Based on the research lacunas, the research questions are as follows.

### **3. Research questions**

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?
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. 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**

The present study intended to investigate whether students' writing performance developed in a CLIL class within one semester. To collect the data, the researcher collected students' written assignments to examine their writing development. Further details concerning the research site, participants, and data analysis are presented as follows.

### **5. 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 in English. The course was a class provided by General Education Center and it was open to students from all of the departments. The students were highly exposed to English during the course. Chinese was used limitedly, with only few circumstances where some personal names needed to be mentioned. In the class, a legal case such as voice-print identification of a bomb-threat

caller or the scandal between Clinton and Lewinsky was introduced each week. Students would learn the association between the case and language. Also, they would learn some linguistic knowledge that could apply to the case. Students were encouraged to read the teaching materials before the class. During the class, the professor first introduced a legal case, its controversies, and the role of language played in the case. After that, the professor would provide questions in relation to the controversial parts of the case. The students were asked to form a small group and discuss those questions. Next, the professor would ask for their individual opinions or they would be divided into groups for a debate. Towards the end of the semester, the students needed to form a group for their final presentations. They needed to choose a legal case by themselves and apply the linguistic knowledge they had learned from the class to analyze it for the presentation. Besides the lectures and activities they participated in the class, the students were also required to write weekly assignments. They needed to provide a brief summary of the case, its controversies, and their reflection on the case. They were informed that their assignments would be evaluated based on their content, organization, critical thinking, and language accuracy (see Appendix I and II for a sample of the assignment and the coding). They were also told that two raters (a master's student and a PhD. student in linguistics (the researcher)) would be scored their assignments. The two raters would discuss the scores and feedback each week to ensure the scoring process was reliable and valid. Furthermore, the students could check their scores and feedback on their assignments on an on-line course platform every week. The feedback they obtained included general comments on their contents and correction on their grammar. To assist the student to write better, in week eight, the professor invited the researcher to share his feedback on the students' assignments for 15 minutes. The researcher talked about their frequent errors in language, e.g. lexical choice, grammar, etc., errors in organization, problems with critical thinking and contents.

## **6. 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. According to Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001), seven of them reached B1 level, six were B2 level, three were C1 level

and one was C2 level. The rest two who did not take a standardized language proficiency test reported their proficiency as intermediate. Also, nine of them are English majors. There are two participants from adult education, two from computer science and two from law department, respectively. Business administration, finance, mass communication, and engineering department also had one student enrolled in this class. Given that the participants were from different departments, their exposure to English varied. Except for those who were from English department had high exposure to English, the rest of the participants had only low exposure to English as most of their classes were taught in Chinese, though they sometimes needed to read English materials relevant to their major studies. For the non-English-major participants, this class was the only class that was instructed in English.

## **7. 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, critical thinking and language accuracy (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.

## **8. 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 includes a clause which must be used with another clause to form a complete grammatical construction. Following Casaneve (1994) and Storch (2009), complexity<sup>1</sup> 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

---

<sup>1</sup> Although complexity was not the main objective of this CLIL class, it was a measurement that normally tied with fluency and accuracy in most L2 language development research. Therefore, the researcher followed this tradition and included it in the research. Also, Whittaker, Llinares and McCabe (2011) have found that students in CLIL class produced more complex sentence in their writing development. Thus, this study also looked at whether syntactic complexity increased over time.

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).

Another rater, who was a PhD. student in linguistics, 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.

## 9. 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.

*Do participants in a CLIL 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 ( $F(2,36)= 1.835, p>.05, \eta^2 = .329$ ) and the ratio of dependent clause per T-unit ( $F(2,36)= 2.581, p>.05, \eta^2 = .445$ ) of the beginning, the middle and the last of the semester, suggesting that the participants' syntactic complexity did not increase over time.

Table 1 Comparisons of syntactic complexity over time

	Mean	SD	Sig.
C/T	1.94	.34	n.s.
Time 1			
Time 2	1.81	.37	n.s.
Time 3	1.71	.36	n.s.
DC/T			
Time 1	.89	.27	n.s.
Time 2	.77	.28	n.s.
Time 3	.69	.28	n.s.

$p>.05$

*Do participants in a CLIL 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 ( $F(2,36)=4.401$ ,  $p > .05$ ,  $\eta^2 = .485$ ). However, a further examination on the proportion of error-free clauses revealed that there were significant differences across different times ( $F(2,36)= 4.401$ ,  $p < .05$ ,  $\eta^2 = .723$ ). As shown in Table 2, pairwise 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 ( $F(2, 36)= 5.236$ ,  $p < .05$ ,  $\eta^2 = .685$ ). Pairwise 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.

Table 2 Comparisons of accuracy over time

	Mean	SD	Sig.
<b>EFC/C</b>			
Time 1	.66	.17	
Time 1-Time 2			.026*
Time1- Time 3			.036*
Time 2	.76	.13	
Time 2- Time 1			.026*
Time 2- Time 3			.58
Time 3	.75	.10	
Time 3- Time 1			.036*
Time 3- Time 2			.58
<b>EFT/T</b>			
Time 1	.58	.20	n.s.
Time 2	.69	.20	n.s.
Time 3	.67	.15	n.s.
<b>E/W</b>			
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

$p < .05$

*Do participants in a CLIL 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.

Table 3 Comparisons of fluency over time

	Mean	SD	Sig.
<b>Total of T-units</b>			
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*
<b>Total of Words</b>			
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
<b>W/T</b>			
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*

$p < .05$

## 10. Discussion

In response to the first research question, the results revealed that the students' syntactic complexity did not improve in the CLIL class. The reason why syntactic complexity did not show significant enhancement could be because they were not asked to produce sentences as complex as possible. Furthermore, topic familiarity might also play a part in the lack of complexity improvement (Tedick, 1990). The participants might produce less complex sentences because the topics were not familiar, e.g. unfamiliar linguistic knowledge that applied to the case. Unfamiliar topics may not allow them to use their prior knowledge to present their ability to write more complex in L2. In addition, the lack of improvement of complexity might be due to the relatively short time period of the class (Storch, 2009). Syntactic complexity might require up to 12 months of college-level instruction to develop (Ortega, 2003). Since this CLIL class only lasted a semester, the students might not have enough time to develop their syntactic complexity. Finally, 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).

With respect to the second research question, the findings revealed that participants improved on their accuracy throughout the semester. The improvement of accuracy could be attributable to the criteria of their assignments as they were informed their scores would be affected if they had three glaring errors that impede communication. Also, the accuracy improvement could be due to the corrective feedback on the students' assignments. Numerous studies have shown that corrective feedback was beneficial for L2 grammatical accuracy (Bitchener, 2008; Bitchener & Knoch, 2009; Bitchener, Young & Cameron, 2005; Ellis, Sheen, Murakami & Takashima, 2008; Van Beuningen, De Jong & Kuiken, 2012). The students' attention might be directed to the errors they made and

noticed the gap between what they could produce and what they needed to produce (Schmidt, 2001). The detection of such discrepancies between their usage and the target language thus made them restructure their existing second language knowledge and push them moving toward a higher level of learning (Gass & Varonis, 1994; Schmidt, 1990). Moreover, the students were required to produce written assignments every week. Therefore, the ample opportunity to produce output enabled them to try out their own hypothesis of a particular language form. Corrective feedback served as facilitative tool when they attempted to produce comprehensible output that extended their linguistic repertoire to formulate their language to convey precisely, coherently, and appropriately (Swain, 1985, 2005). Besides, 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.

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).

## **11. 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 class, relatively high English proficiency of the participants or other factors such as topic familiarity. On the other hand, the influences from the corrective feedback, 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. Although content plays a dominant role in terms of teaching objectives, in a CLIL class, students not only need to progress in their content learning, but also their language learning and

using (Coyle, Hood & Marsh, 2010). To achieve this goal, teachers need to provide sufficient opportunities for students to produce their writing in the target language since one of the principles of CLIL class lies in communication. They need to use the language to learn. Namely, students need language to express and support their thinking processes. By doing so, their language progression will be gained during the learning processes. Apart from using the language to learn, it is equally significant for them to learn how to use the language. Therefore, corrective feedback could also be facilitative in language accuracy enhancement. In addition to creating more opportunities for practice and providing corrective feedback, it is also suggested that teachers could offer more challenging tasks for students. In the present study, the participants were required to summarize the cases, provide its controversies and their reflection on them. Since the contents were already discussed during the classes, it is assumed that the participants already had enough prior knowledge and available resources to write about. Thus, to make the tasks more challenging, teachers could choose a similar case and ask students to use the linguistic knowledge they've learned to analyze the case and provide possible solutions to the controversies in the case. Alternatively, teachers could provide extended questions in association with the controversies in a case to generate more critical opinions from students. Students could also be encouraged to search for other relevant resources to support their arguments in order to present a deeper understanding of the controversies of the case.

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 can not be revealed. Namely, some students may show significant improvement while others do not. To tackle this issue, future studies are suggested to adopt a more in-depth analysis and include some personal factors to account for such differences. For instance, researchers can conduct a case study by examining how individual learners improve over time with respect to different linguistic aspects. Also, students' learning experiences in CLIL, their past language learning experiences, attitudes, motivation and expectations towards CLIL are also deemed as significant factors that can influence their language development in CLIL classes. Researchers can conduct individual interviews of students or teachers and collect students' learning diaries as they are valuable resources to further investigate individual differences of language development in CLIL classes.

## References

- 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.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102-118.
- Bitchener, J., & Knoch, U. (2009). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics*, 31(2), 193-214.
- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14(3), 191-205.
- Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, 39(4), 523-532.
- 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* (pp. 81-92). Buffalo: Multilingual Matter.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, England: Cambridge University Press.
- Coyle, D., Hood, P., & Marsh, D. (2010). *Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Dalton-Puffer, C. (2008). Outcomes and processes in content and language integrated learning (CLIL): current research from Europe. In W. Delanoy & L. 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.
- Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused

- and unfocused written corrective feedback in an English as a foreign language context. *System*, 36(3), 353-371.
- Gass, S. M., & Varonis, E. M. (1994). Input, interaction, and second language production. *Studies in Second Language Acquisition*, 16(3), 283-302.
- Green, T. (2004). Making the grade: Score gains on the IELTS writing test. *Research notes*, 16(4), 9–13
- 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(7), 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.
- Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, 1, 31-42.
- Llinares, A., & Whittaker, R. (2010). Writing and speaking in the history class: A comparative analysis of CLIL and first language contexts. In C. Dalton-Puffer, T. Nikula & U. Smit (eds.), *Language use and language learning in CLIL classrooms* (pp. 125-144). Amsterdam: John Benjamins.
- 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(4), 492–518.

- Richards, J. C., Platt, J., & Platt, H. (1992). *Dictionary of language teaching and applied linguistics*. Essex: Longman.
- 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.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129- 158.
- Schmidt, R. (2001). "Attention." In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.
- 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(2), 103-118.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Cambridge, MA: Newbury House.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Eds.), *Handbook of research in second language teaching and learning* (pp. 471-481). Mahwah, NJ: Lawrence Erlbaum.
- Tedick, D. J. (1990). ESL writing assessment: Subject- matter knowledge and its impact on performance. *English for Specific Purposes*, 9, 123-143.
- Van Beuningen, C. G., De Jong, N. H.D., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, 62(1), 1-41.
- Whittaker, R., Llinares, A., & McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research*, 15(3), 343-362.
- Yang, W. H. (2015). Content and language integrated learning next in Asia: Evidence of learners' achievement in CLIL education from a Taiwan tertiary degree programme. *International Journal of Bilingual Education and Bilingualism*, 18(4), 361-382.

## Appendix I: Sample of the assignment

### Reflection on Voice-print Identification

In this case, we talked about the voice-print identification case: Bomb threat caller. Paul Prinzivalli, a New Yorker, was accused of making a series of threatening telephone calls to the Pan American ticket office in LA. As a suspect, he originally had no evidence to prove himself innocent. However, through the key evidence: vowel stability, he eventually proved that he didn't do the illegal and dangerous things to harm people.

The linguist William Labov found that the recorded telephone threats showed the vowel system of Eastern New England with a single low back vowel in which bomb, on and off are merged. His knowledge of the NYC dialect helped himself to solve the problem. He found that New Yorker's accent vs. Bostonian's accent are different in many ways. When Labov played the tapes, he was sure that Prinzivalli was innocent as a New Yorker. On the other hand, the bomb threat caller was from Eastern New England. Although the differences is not very clear and obvious toward untrained listeners to recognize, phonetician and non-phonetician who knew the Boston dialect all came to the same conclusion. As a result, Labov hoped to help Prinzivalli prove he innocent. The problem is that: How to provide objective evidence for untrained listeners? How to convey this linguistic knowledge to a judge in the Los Angeles area? The judge liked almost West Coast people who heard the two tapes as very similar. Labov knew that there was a serious danger that an innocent person would be convicted of a major crime, with a heavy prison sentences. So, Labov tried hard to find objective evidence and make Judge Ringer understand the differences.

Many people doubt that whether Prinzivalli can disguise himself as a Bostonian to escape from the sentence. However, Labov replies that people can only imitate phonetic sounds but not phonological structures, and no one can imitate a phonological system perfectly! Fortunately, Prinzivalli was offered his job back. It is very useful that forensic phonetician's had big contribution to resolving a legal case of bomb threat caller.

I never heard about voice point, not to mention the identification. I only know about the difference between female and male's voice: high or low, fast or slow, all very tiny part. After the class, I knew the phonological method to categorize different voice. During my sophomore in department FLLD, I studied in introduction in Linguistic, the professor also talked about a little bit about dialect. However, I don't know that dialect and phonological structure can have so big and significant influences on dealing with crime.

## Appendix II: Coding

<p>Organization (2 points)</p>	<p>Follow the summary + reflection organization. Separate them into paragraphs instead of listing your points.</p> <ol style="list-style-type: none"> <li>1. overall organization (1 point)               <ol style="list-style-type: none"> <li>(1) Introduction is clear with relevant background information</li> <li>(2) Links between ideas are appropriate.</li> <li>(3) Conclusion effectively completes the review.</li> </ol> </li> <li>2. paragraph organization (1 point)               <ol style="list-style-type: none"> <li>(1) Begin with the topic sentence to provide the big picture of a paragraph.</li> <li>(2) Illustrate with supporting examples</li> </ol> </li> </ol>
<p>Content: Summary of events (1 point)</p>	<p>The content sticks to the topic without too much digression. Also the summary shows:</p> <ol style="list-style-type: none"> <li>1. All events are summarized succinctly.</li> <li>2. All the key aspects are included.</li> </ol>
<p>Content: Critical engagement with the text (1 point)</p>	<p>The critical engagement is relevant to the events without too much digression. Also, the critical engagement includes the following:</p> <ol style="list-style-type: none"> <li>1. Evidence of a critical engagement with the event by considering different aspects.</li> <li>2. Good evaluative statements to support your stance.</li> <li>3. Compare and contrast your point of view before and after class</li> </ol>
<p>Language: grammatical accuracy (1 point)</p>	<p>Spellings (typos should be avoided) and grammar (e.g., basic tense and agreement, word choice, sentence structure) are generally correct. If you make more than 3 glaring grammatical errors to impede communication, your overall grade will be affected.</p>