

# Assessment of Scoring Content and English in the first CLIL Project at a High School in Sardinia

Chiharu Nakanishi<sup>1</sup> and Hodaka Nakanishi<sup>2</sup>

<sup>1</sup>Music Department, Kunitachi College of Music, Tachikawa, Tokyo, Japan

<sup>2</sup>Technology Transfer Center, Teikyo University, Itabashi-ku, Tokyo, Japan

**Abstract** – *The present study is based on the first experiment of introducing a Content and Language Integrated Learning (CLIL) to a technical high school in Sardinia, Italy. In CLIL there is an assumption that Content teachers and Foreign Language teachers cooperate in teaching and scoring. However, Content teachers and Language teachers have different backgrounds and may have different standards. It is necessary to investigate how a teacher of Content and that of English score students' production in their fields and compare their standards with the criteria, which is made by an analytical rubric. The research questions are: 1) How do a Content teacher and an English teacher score Written and Oral Production in CLIL classes? 2) What is the correlation between the scoring of a Content / an English teacher and the criteria? 3) What do a Content and an English teacher think about the scoring of each other and the correlation of their own scoring? The results show that the scoring of Content and that of English teachers are highly correlated in Written and Oral Production of content but they are not correlated in Oral Production of English. The scoring of contents by English and that of Content teachers, and the scoring of English by the English teacher and the criteria are very highly correlated. However, the correlation of the Content teacher and the criteria is low especially in English. The reason why the Content teacher scores differently in English of Oral Production might be his low English proficiency and his difficulty in hearing. In the future, Content and Language teachers should collaborate to design a rubric to score appropriately when they start regular CLIL classes.*

**Keywords:** CLIL, assessment, scoring, criteria, 1-5 points scale, rubric

## 1 Introduction

Content and Language Integrated Learning (CLIL), where a subject is taught through the medium of a second language, is a growing trend in all levels of education in Europe and in other parts of the world. In CLIL classrooms, children are learning subjects such as history, geography and science in their second or third language in primary and secondary school. CLIL in Europe has been partly due to a commitment of the European Union (EU) to a multilingual Europe. In 1995 EU proposed that every European citizen

should master two other languages in addition to their mother tongue as one of the objectives of the language policy. Following this proposal, EU leaders at the March 2002 Barcelona European Council declared that every child in the EU should be taught at least two foreign languages from an early age. These requirements have given “CLIL” a great importance and have resulted in its establishment in many schools all around Europe. Coyle et al. define, “(CLIL) is a dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language. That is, in the teaching and learning process, there is a focus not only on content, and not only on language. Each is interwoven, even if the emphasis is greater on one or the other in a given time (p.1) [1].”

Though many schools in Europe have adopted CLIL, there are many challenges to implement it. This paper is based on the first experiment of introducing CLIL to a technical high school program in Sardinia, Italy. Among these challenges, we would like to focus on the assessment in CLIL classrooms. In this study, the Italian teachers and the Japanese researchers collaborated to find out how teachers with different backgrounds score students' production.

In this study, the words “score” and “assess” are defined as follows: “Score” is used when Content and English teachers evaluate their students' production. “Assess” is used to evaluate teachers' scoring with the criteria based on analytical rubric designed by one of the authors. In addition, the participants of this study; a Content teacher and an English teacher are abbreviated as follows. The Content teacher who teaches IT is marked as ITT and the English teacher is marked as ET.

## 2 Assessment in CLIL

“The European Framework for CLIL Teacher Education” is published on-line as “a tool for guiding the design of curricula for CLIL teacher professional development” (Marsh et al.) [2]. In this booklet, the target professional competences that the CLIL teacher is expected to acquire are listed. In “Assessing” CLIL teachers are expected to be able to “articulate CLIL-specific assessment needs and goals, and to develop and implement related assessment tools or further develop during the training programme”. However, assessment in CLIL is not a simple issue.

First of all, as CLIL aims a dual focus of content and

language, assessment in CLIL is inevitably a dual focus (Kiely, 2014) [3]. It tends to happen that content and language are assessed at the same time and through the same tasks and activities. However, it is hard to find out the reason why students are poor in their performance. For example, when students cannot explain some features of a historical event, it might be their poor understanding of content or it might be their poor verbal ability.

Secondly, the purpose of CLIL is varied in each classroom setting. Clegg (2012) wonders if the programme is supposed to develop learners' knowledge of language as well as content, should both content and language be assessed? [4] In CLIL there are many types of programs and modules depending on the situation, context, and aim. According to Ikeda, there is a basic division of the CLIL approach into two types depending on purposes: one in which the teaching and learning is focused primarily on language and the other on the subject content [5]. The former is called Soft CLIL and the latter is called Hard CLIL. Therefore, assessment is different depending on whether it is Soft or Hard CLIL.

Thirdly, it might be difficult for subject teachers to assess learner performance in a foreign language which they may feel themselves not qualified or confident (Clegg, 2012) [4]. At the same time, language teachers might also find it difficult to assess learner performance in content which they may feel themselves not qualified or confident.

Fourthly, it is unknown what assessment tools are the most useful for measuring learner performance. Clegg suggests a useful assessment tool may be a bandscale, a set of performance descriptors which allow the teacher to rank a piece of student performance on several sub-skills of a given task, using a pre-constructed scale(2012) [4].

In the present study, the third and fourth questions are investigated. We would like to see how Content and English teachers score Written and Oral Production in the same activity. Performance descriptors to assess students' performance were made to compare and contrast with the scoring of teachers.

### 3 The Present Study

The research questions of this study are as follows:

- 1) How do a Content teacher and an English teacher score Written and Oral Production in CLIL classes?
- 2) What is the correlation between the scoring of a Content / an English teacher and the criteria?
- 3) What do a Content and an English teacher think about the scoring of each other and the correlation of their own scoring?

#### 3.1 Method

##### 3.1.1 Participants

The participants were 12 Italian high school students at a Technical High School in Sardinia, Italy. Their English proficiency levels are A1, A2, and B1 in Common European

Framework (CEF). Among the six levels of CEF, A1 is the lowest, whereas C2 is the highest. These students took a 5-day-program of IT in CLIL as an elective course. Although it was the first experience for them to learn under CLIL approach, they were used to learn content in English as their textbooks were theme based such as on environment and energy.

Two Italian teachers, one IT teacher (hereafter ITT) and one English teacher (hereafter ET) participated in the study. They were both experienced in teaching their own fields more than 20 years, but it was first time for them to teach in CLIL. The English proficiency level of ITT was between A2 and B1 and that of ET was native speaker's level as she was an examiner of the Cambridge Test examiner.

##### 3.1.2 Materials

The class was designed to use the textbook, "Information technology: a story of people and machines" by Davies, P. A., Oxford Bookworms, OUP. Four Writing Prompts and one Oral Prompt were used. The prompts are as follows.

###### 1) Writing Prompts

The students were asked to write 50-80 words in about 5-10 minutes (Table 1).

Table 1: Writing Prompts

1. Analyze the roles of "People and / vs machines". Compare the roles of people and machines in the development of IT.
2. Draw a pie-chart to show percentages of time devoted to everyday free time activities (besides school and sleeping time) with and without computers. Do you think you depend on computers too much? Explain why / why not.
3. What is the role of Alan Turing in developing a computer?
4. List 5 to 7 key words and draw a mind map.
5. Summarize the text.

###### 2) Oral Prompt

The students were asked to give a presentation for about 5-10 minutes as a final project (Table 2).

Table 2: Oral Prompt

Summarize the text using 5-7 keywords. Draw a mind-map and based on the mind map, give a presentation.
--

##### 3.1.3 Classes

The classes were intensive IT CLIL courses from February 3<sup>rd</sup> to 7<sup>th</sup> in 2014. In 5 days, totally 6 hours were taught by ITT and ET. The classes were managed in English most of the time. The teachers encouraged the students to interact with self, content-material, other students and teachers.

## 3.2 Instruments

### 3.2.1 Teachers' Scoring Scale

ITT and ET scored the Written and Oral Production on a 1-5 point-scale from the viewpoint of content and English (Table 3).

Table 3: Teachers' Scoring Scale

	Content	English
Scores	5	5
	4	4
	3	3
	2	2
	1	1

The criteria were made by one of the authors, Chiharu NAKANISHI, an educational researcher who was the observer of CLIL classes. After observing CLIL classes, she made the criteria of assessment by constructing analytical rubrics; a band scale, which is a set of performance descriptors. In her daily classes, she uses analytical rubrics to evaluate her students' production. She thinks the analytical rubric is a convenient and objective tool to share the criteria of evaluation with her students. She also thinks the analytical rubric can be used to show the goal to her students. Based on her experiences of making and using the analytical rubric, which is the criteria, she evaluated all Written and Oral Production. The criteria for Written Production (Table 4) were adapted from Tabak et al (2012) and Merli (2012). The criteria for Oral Production were adapted from Tabak et al (2012) and (<http://rubistar.4teachers.org/>) (Table 5).

Table 4: The Criteria for Written Production

	<b>5 (Excellent)</b>	<b>4 (Good)</b>	<b>3 (Satisfactory)</b>	<b>2 (Almost satisfactory)</b>	<b>1 (Unsatisfactory)</b>
<b>Content</b>	Student shows a complete and thorough knowledge of the subject.	Student shows a complete knowledge of the subject.	Student has essential knowledge of the subject.	Student is lacking necessary background knowledge of the subject.	Student shows no knowledge of the subject.
<b>English</b>	Consistent grammatical control and appropriate use of vocabulary.  No mistakes except for some articles (1-2).	Good grammatical control and generally appropriate use of vocabulary.  Some mistakes may be present (2-4).	A few mistakes in grammar and vocabulary use that do not lead to misunderstanding.  Mistakes in structures and verb forms may be present (5-8).	Systematically makes mistakes in grammar and vocabulary use but the message is generally clear.  Mistakes in structures and verb forms (5-8) are disturbing the meaning.	Systematic grammar mistakes and narrow range of vocabulary make the message meaningless.  Mistakes prevent understanding of content.

Table 5: The Criteria for Oral Production

Scores Descriptions	5 (Excellent)	4 (Good)	3 (Satisfactory)	2 (Almost satisfactory)	1 (Unsatisfactory)
<b>Content</b>	Student summarizes the text using appropriate 5-7 keywords. S/he draws an appropriate mind-map and gives a presentation based on the mind map. S/he shows a complete and thorough knowledge of the subject.	Student summarizes the text using 5-7 keywords. S/he draws a mind-map and gives a presentation based on the mind map. Student shows a complete knowledge of the subject.	Student summarizes the text using more or less than 5-7 keywords. The choice of keywords is satisfactory. S/he draws a satisfactory mind-map and gives a presentation based on the mind map. S/he has essential knowledge of the subject.	Student summarizes the text using more or less than 5-7 keywords. The choice of keywords is not satisfactory. S/he draws an almost satisfactory mind-map and gives a presentation based on the mind map. Student lacks necessary background knowledge of the subject.	Student summarizes the text using more or less than 5-7 keywords. The choice of keywords is unsatisfactory. S/he draws an unsatisfactory mind-map and gives a presentation based on the mind map. Student shows no knowledge of the subject.
<b>English</b>	Speaks clearly and distinctly all (100-90%) the time, and mispronounces no words.	Speaks clearly and distinctly all (90-80%) the time, but mispronounces no more than 2 words.	Speaks clearly and distinctly most (80-70%) of the time. Mispronounces more than 4 words.	Speaks clearly and distinctly most (70-60%) of the time. Mispronounces more than 6 words.	Often mumbles or can not be understood or mispronounces more than 8 words.

### 3.3 Procedures

- 1) ITT and ET scored 4 Written and 1 Oral Production separately on a 1-5 point-scale.
- 2) ITT and ET held a meeting for scoring students' Production, when they evaluated two Writing Production in one day.
- 3) One of the authors, Chiharu NAKANISHI (hereafter CN) observed and videotaped the classes. After observing CLIL classes, she made the analytical rubric, which was used as the criteria. After she read through all Written Production and watched Oral Production, she evaluated the Written and Oral Production by using the criteria. The analysis of scoring by ITT and ET is based on CN's assessment.
- 4) ITT and ET were interviewed about their scoring and the correlation between their scoring by CN.
- 5) Hodaka NAKANISHI, the other author, calculated and examined the correlation of scoring by ITT, ET and of assessing by CN.

### 4 Results

All scoring by ITT, ET and assessing by the criteria were analyzed from the viewpoints of content and English.

*Research Question 1: How do a Content teacher and an English teacher score Written and Oral Production in CLIL classes?*

Correlation of scoring by ITT and ET in Written and Oral Production is shown in Table 6 & 7. Between ITT and ET, the average of correlation for scoring Content is higher (Written Production: 0.70, Oral production: 0.74) than that of English (Written Production: 0.57, Oral Production: 0.39).

On both Feb. 5th and 7th, ITT and ET had a short scoring meeting between Activity 1(A1) & 2(A2) and Activity 3(A3) & 4(A4). After the meeting, both the correlation of scoring content and that of scoring English increased.

Table 6: Correlation of Scoring by ITT and ET in Written Production

Evaluator	Viewpoints	Date and Activity				
		Feb.5 A1	Feb.5 A2	Feb.7 A3	Feb.7 A4	Average
ITT and ET	Content	0.59	0.82	0.74	0.73	0.70
	English	0.49	0.76	0.39	0.72	0.57

Table 7: Correlation of Scoring by ITT and ET in Oral Production

Evaluator	Viewpoints	Average
ITT and ET	Content	0.74
	English	0.39

*Research Question 2: What is the correlation between the scoring of a Content / an English teacher and the criteria?*

The correlation of scoring by ITT & CN and ET & CN in Written and Oral Production is shown in Table 8 and 9. The result shows that the correlation of scoring by ET and CN is high both in Content and English in Writing and Oral

Production (Content: Written Production 0.90, Oral Production 0.88) (English: Written Production 0.94, Oral Production 1.0). When to see scoring by ITT and CN, it is relatively highly correlated in Content (Written Production 0.83, Oral Production 0.74). However, the correlation of English scoring by ITT and CN is low (Written Production 0.62, Oral Production 0.39).

Table 8: Correlation of Scoring by ITT &amp; CN and ET &amp; CN in Written Production

Evaluators	Viewpoints	Date and Activity				
		Feb.5 A1	Feb.5 A2	Feb.7 A3	Feb.7 A4	Average
ITT & CN	Content	0.82	0.87	0.91	0.75	0.83
	English	0.50	0.81	0.40	0.81	0.62
ET & CN	Content	0.91	0.94	0.92	0.85	0.90
	English	0.95	0.94	0.90	0.95	0.94

Table 9: Correlation of Scoring by ITT &amp; CN, and ET &amp; CN in Oral Production

Evaluators	Viewpoints	Average
ITT & CN	Content	0.74
	English	0.39
ET & CN	Content	0.88
	English	1.00

*Research Question 3: What do a Content and an English teacher think about the scoring of each other and the correlation of their own scoring?*

Both ITT and ET were asked what they thought about their scoring. ITT and ET said that they were not used to score by 1-5 point-scale, but were glad to see the high correlation of their scoring in Content. They said that this outcome may be the consequence of a long and deep collaboration of teaching and managing classrooms.

ITT confessed that his low English proficiency affected his scoring English. He also said that he had difficulty in hearing in his left ear. He said he was not confident of scoring students' English especially in Oral Production.

ET said that English teachers had to keep studying subject contents when they teach advanced students and when they teach in CLIL. She also said it was not possible to teach English without contents. She mentioned that it was not easy for her to study a variety of contents. She said though she made an effort to study content, she was not confident enough of scoring students' production in content.

ITT and ET said that when they score students' production in next CLIL program, it might be possible that ITT would focus on content, IT and ET would focus on English. They thought it would be easy and save their energy to care their own fields.



## 5 Conclusions

The present study showed that scoring content by ITT was correlated significantly with the scoring by the criteria, whereas scoring English by ET was correlated significantly with the criteria. The correlation of scoring English between ITT and ET, or ITT and the criteria in Oral production is low. It could be caused by ITT's low English proficiency and his hearing problem. The interviews showed that scoring is not easy for both ITT and ET. For ITT to score English was not easy. For ET to score content was not easy.

In order to score more appropriately in CLIL, there are 2 possible cues. First, when the gap between Content and Language teachers is more than 2 points on the 1-5-point-scale, the scoring should be checked to find out the reasons of the gap and how each teacher actually did the scoring. Secondly when they implement the CLIL program, Content and Language teachers and other related staffs should collaborate with designing a rubric. According to Stevens and Levi, rubrics can be shared to determine whether or not there is a consensus on what is being taught, how it is being taught, and why (2013) [6]. They also state, "sharing rubrics can also reveal the degree to which grading is consistent" (2013, p.24). They insist that rubrics cannot indicate what to do when teachers grade differently, but rubrics can at least make us aware of the situation.

There is a limit to the present study. The 1 set of teachers, Content and English, and 12 students are not enough to generalize about scoring and assessing.

For our future study, we'd like to improve on the rubrics to allow both Content and English teachers to score easily. Then we'd like to create and provide a variety of CLIL rubrics for teachers. Moreover we are also interested in CLIL teacher training in their scoring.

According to key data on teaching languages at school in Europe (2012), CLIL seemed to have been implemented in primary and secondary education in Sardinia, Italy (p.39) [7]. However, in the present study's technical high school it was completely the first experience for the teachers to implement CLIL. Though the teachers were working hard to start CLIL, they were not trained enough as CLIL teachers. How to score appropriately had never been discussed. Moreover, the Content teacher's English proficiency is too low to teach in English. The education of CLIL teachers and the improvement of Content teachers' English proficiency are two essential issues.

## 6 Acknowledgements

This work was supported by Grants-in-Aid for Scientific Research, Grant Number 25350294, awarded by Japan Society for the Promotion of Science. We show our special thanks to the IT and English teachers and the students of the Sardinian high school who participated in this study.

## 7 References

- [1] D. Coyle, P. Hood & D. Marsh. "Content and Language Integrated Learning". Cambridge: Cambridge University Press, 2010.
- [2] D. Marsh, "The European Framework for CLIL Teacher Education" Viewed on 5 April, 2015
- [3] R. Kiely (2014). "CLIL-The question of assessment" Developing Teachers.com  
<[http://www.developingteachers.com/articles\\_tchtraining/clilp\\_f\\_richard.htm](http://www.developingteachers.com/articles_tchtraining/clilp_f_richard.htm)> Viewed on 5 April, 2015
- [4] J. Clegg. "Preface" In Quartapelle (Ed.) Assessment and Evaluation in CLIL, 2012
- [5] M. Ikeda. "CLIL no kihongenri (Principles of CLIL)" In Y. Watanabe, M. Ikeda, and S. Izumi (Eds.), CLIL New Challenges in Foreign Language Education at Sophia University Volume 1: Principles and Methodologies, Sophia University Press. 2011
- [6] D. Stevens & A. Levi Introduction to Rubrics, An assessment tool to save grading time, convey effective feedback, and promote student learning, VA: Stylus. 2013
- [7] European Commission, Key data on teaching languages at school in Europe, 2012,  
<[http://eacea.ec.europa.eu/education/eurydice/documents/key\\_data\\_series/143en.pdf](http://eacea.ec.europa.eu/education/eurydice/documents/key_data_series/143en.pdf)> Viewed on 5 April, 2015