

Does CLIL Work for Japanese Secondary School Students? Potential for the 'Weak' Version of CLIL

Makoto Ikeda

Sophia University

Abstract

This is a longitudinal research study on teaching evaluation and learning outcomes of a language-driven CLIL course in Japan. The participants were upper secondary school students (16-17 years old) with lower intermediate English proficiency. The measuring instruments were a course evaluation questionnaire completed by the learners and essay writing tests (before and after the course) marked by an automated scoring programme. The results were positive in both: the participants perceived the CLIL course as distinctively different from other language courses and their essay writing skills improved significantly. Qualitative data derived from teacher interviews and their teaching journals were also examined to support interpretation and discussion of the findings. The conclusion drawn from this study is that a 'weak' form of CLIL works in the Japanese context if teachers are trained properly. Future research into CLIL language teachers' professional development is also suggested.

Key words: CLIL, Japanese students, course evaluation, language development, teacher education

1. Introduction

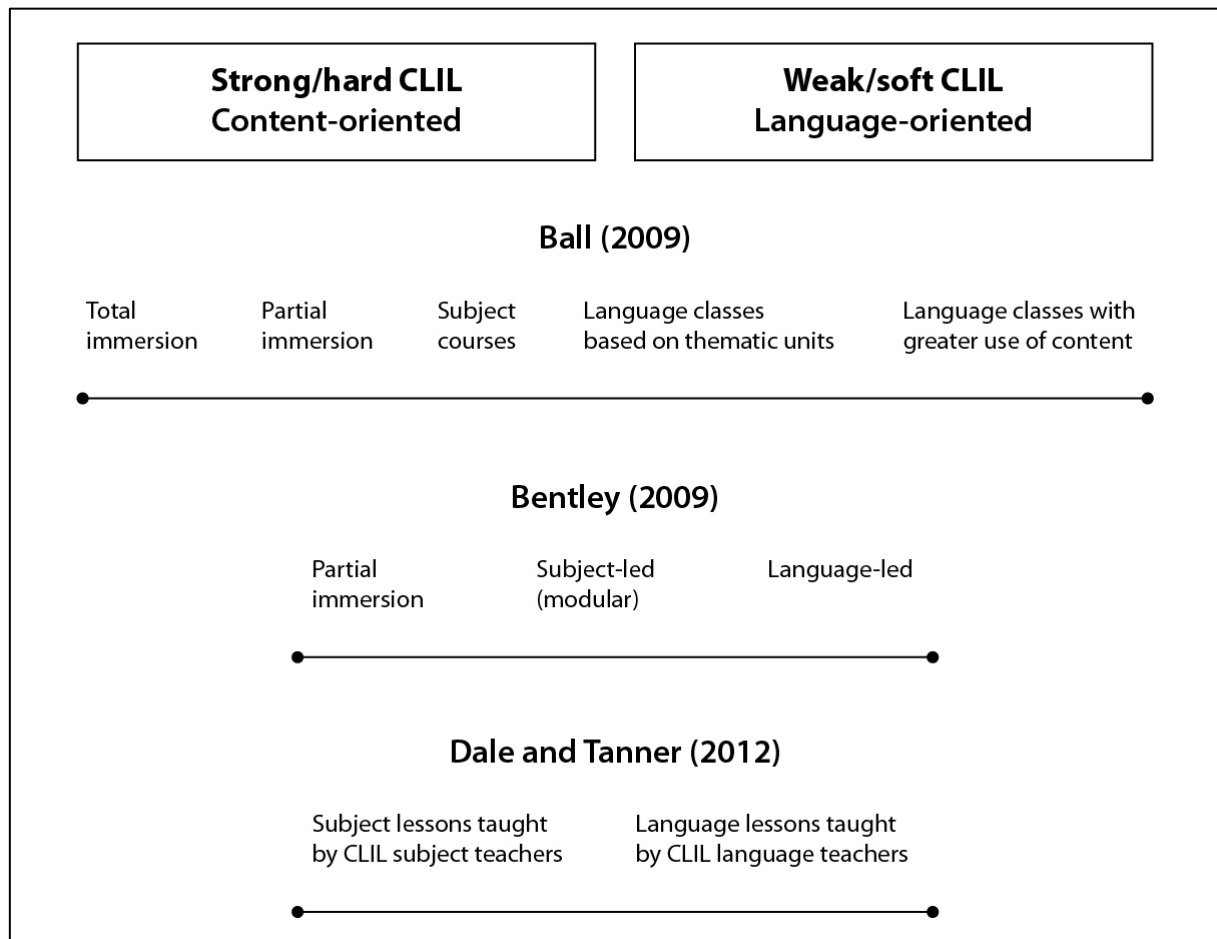
Although the acronym CLIL is sometimes presented as an 'umbrella term' comprising not only immersion and modular subject learning through another language but even student exchanges and homestays (Mehisto *et al.*, 2008: 12-19), its typical context would be classrooms in Europe where academic subjects such as geography or science are taught in English by non-native content teachers. This default type is classified as 'strong' or 'hard' CLIL (Ball, 2009; Bentley, 2009) and all the research outcomes summarised in recent review articles (Dalton-Puffer, 2011; Pérez-Cañado, 2012) are produced in those prototypical settings. This is not surprising, considering the fact that major terms and concepts underpinning CLIL (e.g. the 'dual-focused' definition by Marsh, 2002; the 'four C's' model by Coyle, 2007; Coyle, 2008) have been provided by European bilingual education specialists (as opposed to applied linguists) and are focussed on the teaching and learning of content (e.g. Science, History) through an additional language.

The other version of CLIL, which is termed 'weak' or 'soft' CLIL (Ball, 2009; Bentley, 2009), is taught by native or non-native language teachers with more focus on language learning. As CLIL spreads beyond European boundaries, this ELT (English Language Teaching) type of CLIL is expected to grow and an indication of such a trend has already appeared. For one thing, some updated editions of popular ELT methodology books (e.g. Harmer, 2012; Scrivener, 2011; Ur, 2012), whose target audiences are language teachers throughout the world, have a section describing CLIL as an alternative language teaching methodology. Furthermore, new CLIL titles classified as ELT teaching materials (e.g. the *Oxford Read and Discover* series) are added every year; one coursebook (*World Around* published by Helbling Languages) has even the advertising copy 'CLIL inside!' on its front cover to attract the attention of prospective users. Contrary to this new and expanding phenomenon, research on the 'weak' type of CLIL is difficult to locate, hence this study.

2. Defining 'soft/weak' CLIL

As the 'strong' type is the norm in the present CLIL world, the distinction between the 'strong/hard' version and the 'weak/soft' one has not been discussed in the academic literature. The most comprehensive model in the literature targeting practitioners is offered by Ball (2009: 37-38). On the 'strong/hard' versus 'weak/soft' continuum, he identifies five types of CLIL programmes: total immersion, partial immersion, subject courses, language classes based on thematic units and language classes with greater use of content (Figure 1). For him, any kind of content and language combined education is CLIL. A more restricted view is taken by Bentley (2009: 6), who proposes three types of CLIL depending on class contact time: partial immersion (about half of the curriculum), subject-led/modular (15 hours per term) and language-led (one 45-minute lesson per week). The CLIL coverage proposed by Dale and Tanner (2012: 4-5) looks, at a glance, the same as that of Bentley (2009). What makes their classification unique is that their criterion for differentiating CLIL and non-CLIL programmes lies in the teacher. Thus, immersion and CLIL subject courses are different in that the latter are taught by CLIL subject teachers who can teach their subjects while paying adequate attention to content-specific language; CBLT (Content-based Language Teaching) is distinguished from CLIL language classes, which are taught by CLIL language teachers who can teach not only the target language in general but the target subject language in particular. In short, regardless of content-driven or language-driven, what makes a CLIL course depends on whether or not the teachers are trained in CLIL principles and skills. Drawing on the discussion so far, a working definition of the 'weak' version of CLIL is proposed as this: 'Weak/soft' CLIL is a type of content and language integrated instruction taught by trained CLIL language teachers to help learners develop their target language competency as a primary aim and their subject/theme/topic knowledge as a secondary aim.

Figure 1. Comparison of three CLIL continuum models



3. A feasible type of CLIL in Japanese secondary education

There are several EMI (English-medium instruction) programmes currently available in Japan. In primary and secondary education 24 schools offer IB (International Baccalaureate) programmes in English (MEXT, 2013). At the tertiary level, some universities provide undergraduate and postgraduate programmes where students can earn their degrees only in English. These programmes are few in number and mainly taught by native-speaker content teachers and lecturers. In the mainstream context (e.g. state secondary schools), curricular subjects are taught by Japanese content teachers in their first language. The possibility of those teachers teaching their subjects in English is very slim: many English teachers are struggling to teach in English, much less content teachers with their average English proficiency level of probably A2 on the CEFR (Common European Framework of Reference) scale. In other words, the 'strong' version of CLIL (i.e. subject lessons taught by Japanese content teachers) is unrealisable in the present circumstances, at least until sufficient linguistic training, appropriate teaching materials and language assistants are available to compensate for limited teacher language skills. A more realistic CLIL model that can be implemented without much difficulty is the 'weak' version of CLIL defined in the last section (i.e. trained language teachers teach English and content using CLIL). A question which naturally arises here is 'does it work?'. To be more specific, does CLIL make any difference in students' classroom experience and learning? This leads to the two research questions that this study tries to answer:

- (1) Do the students perceive any CLIL features which are distinct from normal English lessons?
- (2) Does their language proficiency develop as a result of CLIL lessons?

4. This study

4.1 The project

The CLIL project for which this research was conducted started in April 2011 and was ongoing during the writing of this article (August 2013). The context is a state secondary school in Wako City, Saitama Prefecture, which is within commuting distance from Tokyo. The school runs two courses for students in upper secondary education (16-18-year-olds): the mainstream curriculum course and the foreign language track in which the students receive more language lessons (nine English classes plus two French/German/Spanish/Chinese lessons weekly) than the normal stream students (six English lessons per week). Targeted at the students in the foreign language section (about 80 students each year), this project has been carried out by the researcher (the author) and two Japanese teachers of English (a male teacher with 29 years' teaching experience and a female teacher with 20 years of experience as an English instructor). The teachers were trained in CLIL methodology (e.g. principles, lesson planning, materials development, classroom techniques) by the researcher based on two CLIL methodology books (Coyle *et al.*, 2010; Watanabe *et al.*, 2011).

Before launching this project in April 2011, the researchers and the teachers had several meetings to develop a CLIL course drawing on the curriculum design format proposed by Nation and Macalister (2010). Here is the summary (Table 1):

Table 1. Overview of the CLIL course development

<p><u>Goals</u></p> <ul style="list-style-type: none"> - Students learn both content (cross-cultural understanding and global issues) and language (English knowledge and skills) with equal focus. - The specific aim in language learning is producing quality discourse, especially expressing opinions in spoken form. - Both students and teachers enjoy using English as a tool to study subject matter.

Environment

- About 80 students take this course each year. They are divided into four groups to limit the class size up to 20 students.
- Each group meets twice a week (50 minutes and 90 minutes each): one session is taught by a Japanese language teacher (JLT) and the other one is taught jointly by a JLT and an assistant language teacher (ALT), who are native speakers of English.
- Classrooms have audio-visual equipment (e.g. projector or monitor, PC and DVD player).

Needs

- Students are relatively good at listening and reading, but weak at writing and speaking.
- Their challenge is building up adequate vocabulary and producing effective discourse.
- They hope to understand global issues and current topics in English, and talk about them in the language.

Principles

- A 'weak' version of CLIL (i.e. a language course taught by trained CLIL language teachers) is adopted.
- Coyle's 'four C's (Content, Communication, Cognition, Culture) are placed at the centre of the course.
- Task-based cooperative learning is emphasised.

Content and sequencing

- Global affairs such as global warming, endangered species, racism, human rights, peace and conflict are explored.
- To motivate students, topics closely related to their lives are taken up at the initial stage of the course.

Format and presentation

- Home-spun hand-outs composed of authentic materials and accompanying tasks are used.
- The lesson format is what is termed 'the new PPP' by the researcher, meaning Presentation (written or spoken input of new content and language), Processing (tasks worked on in pairs or groups using lower-order and higher-order thinking skills) and Production (written or spoken output such as group oral presentations based on manuscripts).
- Other methods/approaches (e.g. oral approach, audiolingualism, lexical approach) are used when appropriate.

Monitoring and assessment

- Students are tested in the knowledge and skills of both content and language they learnt in class.
- Portfolio assessment is employed if necessary.
- 'Learning moments' are carefully monitored while teaching.

Evaluation

- All the documents to record and evaluate the quality of teaching are collected (i.e. syllabuses, teaching materials, teaching journals, learning portfolios, lesson videos).
- Qualitative and quantitative data for evaluating the course are collected (i.e. student questionnaires, student interviews, teacher reflections).
- Essay writing tests are administered at the beginning and the end of the course to gauge students' language development.

4.2 Participants

The participants in this research were 80 students (16-17 years old; 62 female and 18 male students) of the academic year of 2012 (from April 2012 to March 2013). This year group was chosen for study because the teachers were used to CLIL methodology, course content and teaching materials. According to the teachers, the students' average English proficiency was B1 on the CEFR scale, with a few students having achieved an A2 or B2 level. The total hours of CLIL lessons they received for the whole academic year (35 weeks) was 62.5 hours for two groups and 72.5 hours for the other two groups. Apart from these CLIL lessons, the students also had three general English lessons (language knowledge and the four skills) and another three lessons focusing on productive skills (i.e. paragraph writing, presentation skills and debating). In all, they had nine English lessons weekly.

4.3 Instruments

To answer the research questions, the following instruments were devised and administered:

(1) Course evaluation questionnaire

In the last lesson of the whole academic year, the participants were asked to fill in a course evaluation sheet anonymously (they read and answered the questions in Japanese). The questionnaire was composed of two sections: ten multiple-choice questions and one open-ended question. The closed-ended questions (Table 2) were designed to explore the participants' perceptions about teaching and learning offered by the CLIL course, paying attention to the Coyle's four C's: overall evaluation (Q1), Cognition (Q2), Culture (Q3), Content (Q4) and Communication (Q5-Q10). For each of the items, the participants marked one of the five Likert scale responses (5=strongly agree, 4=agree, 3=neither agree nor disagree, 2=disagree, 1=strongly disagree). The pilot version of the questionnaire was administered to 69 students in the previous year and item analysis was made to confirm if the questions worked. Also, in order to measure the internal consistency reliability of the ten-item scale focusing on the participants' perceptions on the teaching and learning in the CLIL course, statistical analysis was carried out and the Cronbach Alpha coefficient of the total scale was 0.91, indicating that the scale had very good reliability.

Table 2. Course evaluation questionnaire items

Q1	This course was 'denser' than other English courses, judging from content, the quality and quantity of English, teaching materials and learning activities.
Q2	In this course, I used more thinking skills to learn about content than other English courses.
Q3	There were more group activities in this course than in other English courses.
Q4	Thanks to this course, my knowledge and interest in global issues (e.g. environmental problems, human rights) were enhanced.
Q5	Thanks to this course, I have picked up the types of English vocabulary that do not appear in ordinary course books (e.g. words about current topics).
Q6	Thanks to this course, my ability to read various documents in English has improved.
Q7	Thanks to this course, my ability to watch and listen to videos in English has improved.
Q8	Thanks to this course, my ability to write my own opinions in English has improved.
Q9	Thanks to this course, my ability to give presentations in English has improved.
Q10	Thanks to this course, my ability to discuss in English has improved.

The open-ended question, which read 'please write freely about what you liked and what you think should be improved about this course', was intended to find out the learners' opinions on various aspects of the course which were not covered by the structured closed questions. The completed sheets were collected by the teachers and the data were processed statistically by the researcher.

(2) Essay writing tests (pre-test and post-test)

As stated in 4.1, the linguistic goal of this CLIL course was the production of quality discourse, particularly in oral language. The learning outcomes in language measured in this study were, however, written discourse production, more specifically, essay writing skills. There are three reasons for this. One is that writing effective academic prose is as important as discussing subject topics, debating controversial issues and presenting facts, ideas, opinions, etc. in CLIL learning environments. Another reason is that spoken and written developmental data are reported in the literature to show correlated results, as summarised in Wolfe-Quintero *et al.* (1998: 4). The last reason is practicality: it is much easier to collect, process and analyse written language samples than to deal with spoken data.

In assessing the participants' writing, this research follows the traditional assumption that underlies many studies of second language development in writing: that 'more proficient second language writers are more fluent, accurate, and complex in their writing than less proficient writers' (Wolfe-Quintero *et al.*, 1998: 4). In order to measure these three areas in writing skills (i.e. fluency, accuracy and complexity), two digital tools were used. One is the *Criterion* Online Writing Evaluation Service, an Internet-based, automated scoring system developed by Educational Testing Service (ETS). This pay computer programme provides two kinds of immediate feedback on the essays students typed in and turned in: holistic scores ranging from a 1 (the lowest) to a 6 (the highest) and detailed error analysis on grammar (e.g. subject-verb agreement), usage (e.g. word forms), mechanics (e.g. spelling and punctuation), style (e.g. repetition of words) and discourse structure (e.g. essay organisation). What makes this possible is a scoring engine called *e-rater*, which generates scores based on the comparison in word choice, sentence structure and essay organisation between submitted essays and the model essays pre-marked by human raters (Burstein *et al.*, 2004). According to Attali and Powers (2008), the correlation between *e-rater* and human scores were almost unified (0.97) in the two studies they referred to. The other tool used in this research was the *Range* programme for vocabulary analysis (Heartley *et al.*, 2002). This well-known, free-distribution software compares the words used in submitted texts with three base word lists (i.e. the most frequent 1000 English words, the second 1000 most frequent English words and the words not included in the first two lists but frequently used in academic texts), which were based on 'A general service list of English words' (West, 1953) and 'The academic word list' (Coxhead, 1998; Coxhead, 2000). By running this programme, data on the levels of vocabulary used in essays are available.

At the beginning (April 2012) and the end (March 2013) of the one-year CLIL course, the participants wrote their essays on school computers in the language lab in response to the prompt 'Discuss a simple plan for cutting your country's energy consumption by 60%'. The length of time allowed for writing was 30 minutes and their essays were submitted and marked on the spot by *Criterion* (Grade 11). The essays saved in the *Criterion* account were then copied and pasted to *Word* document sheets to count the number of words in each essay, which in turn were put into *Range* for vocabulary analysis. The results (i.e. the holistic scores for the overall quality of writing, the number of words for fluency, the number of errors per sentence for accuracy, and vocabulary distribution by word frequency level for complexity) were processed statistically.

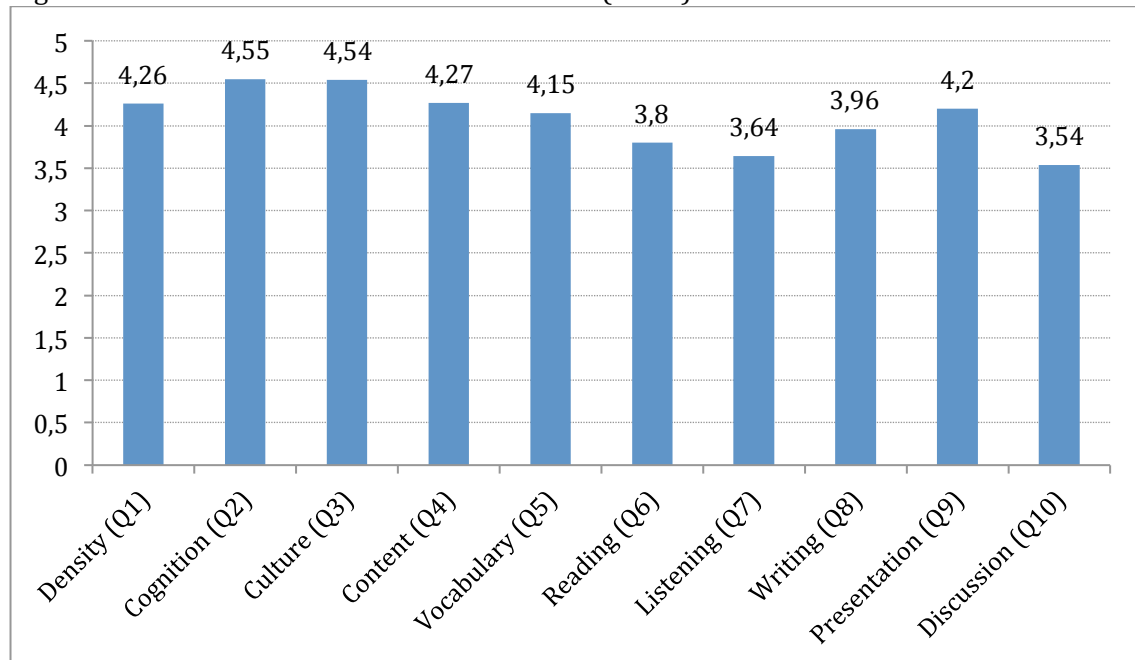
Besides the questionnaire and the tests, follow-up interviews with the teachers were held to consider the results from various angles. Also, the teaching journals the teachers wrote after each lesson and sent to the researcher by email were examined.

5. Results and discussion

(1) Course evaluation questionnaire

The participants' responses to the questionnaire show that they appreciated the value of the CLIL course taught by the trained teachers (Figure 2).

Figure 2. Student evaluation of the CLIL course (N=74)



Question 1 concerns the overall appraisal of the CLIL course, showing that most of the participants felt that their learning experience on the course was 'denser' than in normal language lessons (4.26). This 'density' seems to owe much to the CLIL's four C's. The following student comment in the open-ended question, which was written in Japanese and translated into English by the researcher, best summarises the features of this CLIL course:

I learnt about the world situation that I hadn't known at all [Content/Culture]. I enjoyed the lessons particularly because I thought about the topics deeply using my head [Cognition], discussed with my friends [Communication/Culture] and gave presentations in groups [Communication/Culture].

This response is endorsed by the high evaluation scores on the questions concerning each of the four C's: Cognition (Q2 more critical thinking, 4.55), Culture (Q3 more cooperative learning, 4.54), Content (Q4. increased interest and knowledge, 4.27) and Communication (Q5 enlarged vocabulary, 4.15; Q 9 improved presentation skills, 4.2).

These characteristically CLIL-specific learning opportunities noticed by the students did not come by accident but were brought about on purpose. Before the teachers saw the questionnaire results, they completed a checklist for detecting 'how CLIL' they were (Dale and Tanner, 2012: 15-16). Out of the 36 statements on the list, the CLIL pedagogical skills that both of the teachers 'always' used precisely coincided with those evaluated positively by the participants (Table 3).

Table 3. The CLIL skills that the teachers always used (based on Dale and Tanner, 2012)

<p>1. At the start of a lesson or topic, I find out what learners know about the topic.</p> <p>3. I use visuals (photos, video, drawings, etc.) to introduce new topics.</p> <p>6. I ask learners to talk to each other when I am activating their prior knowledge.</p> <p>7. I provide different sorts of input (multimodal input) – texts, pictures, real objects, videos, models – to help learners understand the topic.</p> <p>8. I formulate and use different kinds of questions – some related to LOTS (lower-order thinking skills) and others related to HOTS (higher-order thinking skills) - to help learners understand input and process information actively.</p> <p>9. I encourage my learners to interact in my classes and use a lot of pair and group work.</p> <p>12. I work actively with my learners on developing their thinking skills.</p> <p>17. I help my learners learn and use subject-specific terminology.</p> <p>19. Learners often speak in English during my classes, i.e. I encourage spoken output.</p>

Items 1, 3, 6, 7 and 8 are related to Content (various kinds of scaffolding to help the students learn subject matter); Item 9 concerns Culture; Item 12 deals with Cognition; Items 17 and 19 are about Communication (vocabulary and speaking). The agreement between the students' perceptions and the teachers' pedagogical intentions underscores the value of professional development in CLIL as the teachers in this study have all been trained in these CLIL skills.

Although the students highly appraised the CLIL-style teaching based on the four C's model (all the items were above 4.0), their overall perceptions about language learning were not comparatively favourable. The language-related responses except vocabulary acquisition and presentation proficiency were below 4.0: Reading (Q6, 3.80), Listening (Q7, 3.64), Writing (Q8, 3.96) and Speaking (Q10, 3.54). Contrary to their modest self-assessment of language development, the participants improved, at least, their written language skills over the year, as the data in the next section show.

(2) Essay writing tests (pre-test and post-test)

The comparison of the participants' two essays (before and after the course) indicates that their written language proficiency improved to some extent (Table 4).

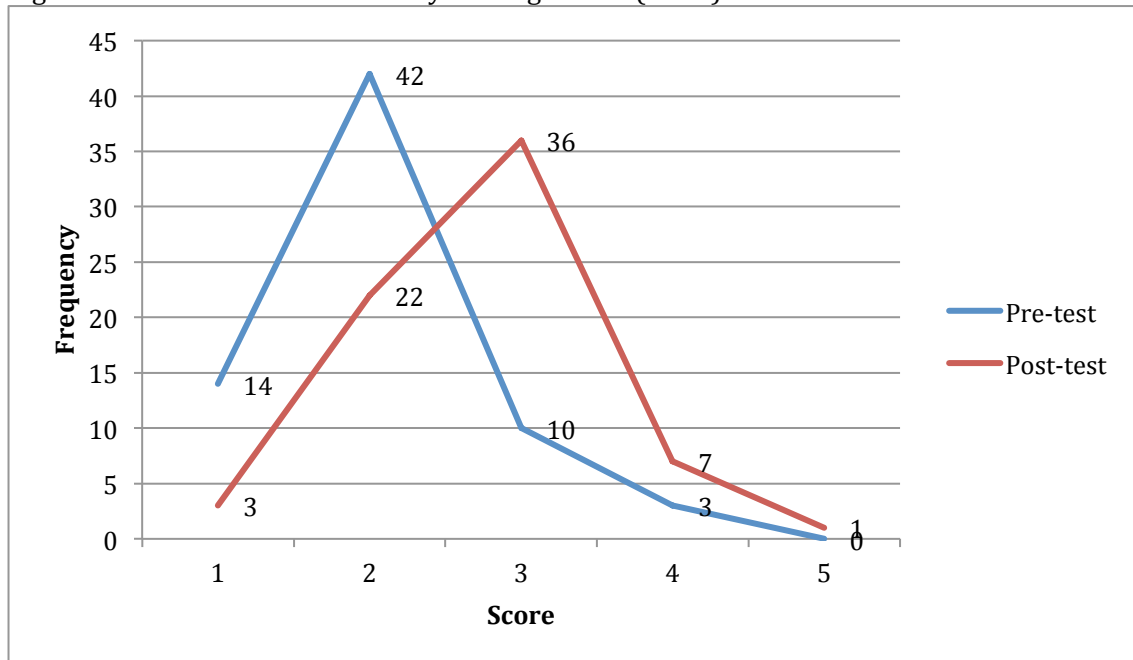
Table 4. Summary of the writing tests means (n=69)

Category	Scale	Test	Mean	SD	t-value
Holistic score	<i>Criterion score</i>	Pre-test	2.03	0.73	-8.10**
		Post-test	2.72	0.77	
Fluency	Number of words	Pre-test	153.67	53.24	-6.47**
		Post-test	196.38	63.99	
Accuracy	Number of errors	Pre-test	15.03	8.30	-3.55**
		Post-test	19.61	10.04	
	Number of errors per sentence	Pre-test	1.07	0.43	-2.98**
		Post-test	1.29	0.54	
Complexity	Number of word types	Pre-test	83.38	21.09	-4.86**
		Post-test	97.33	24.86	
	Percentage of Base list 1 words	Pre-test	83.92	5.14	3.76**
		Post-test	80.88	5.33	
	Percentage of Base list 2 words	Pre-test	4.78	2.48	-1.07
Post-test		5.16	2.48		
Percentage of Base list 3 words	Pre-test	3.00	1.79	-1.75	
	Post-test	3.39	1.37		
Percentage of other words	Pre-test	8.30	4.12	-3.40*	
	Post-test	10.58	4.75		

* $p < 0.05$; ** $p < 0.01$

Overall, the general quality of students' writing based on vocabulary, grammar and organisation improved. The mean score of the pre-test (2.03) and the post-test (2.72) differed significantly. Although the difference in the mean scores looks minute, score distributions between the two tests are clearly distinct (Figure 3): the mode of the pre-test was score 2 (n=42), which went up to score 3 (n=36) in the post-test. A set of score 2 and score 3 essays written by the same student and the scoring guide provided by ETS are appended at the end of this article.

Figure 3. Distributions of the essay writing scores (n=69)



Turning from holistic assessment to more detailed analysis, the participants' fluency in writing developed by 21.93 per cent: the average number of words they used in the pre-test in thirty minutes was 153.67, while in the post-test that figure was 196.38, making the difference statistically significant. The accuracy in writing, however, deteriorated. Not only did the average number of errors in grammar, usage and mechanics increase from 15.03 to 19.61, which is expected as the students wrote more in the post-test (15.52 sentences on average, which was 14.16 sentences in the pre-test), but the average number of errors per sentence went up from 1.07 to 1.29. Only in this respect do the Japanese 'weak' CLIL students seem to differ from their European 'strong' CLIL counterparts in the development of writing abilities: CLIL students in Europe are reported to show 'a higher degree of accuracy' than non-CLIL learners in several studies (Dalton-puffer, 2011; 186).

As to lexical complexity, the mean number of word types (i.e. headwords) in the pre-test (83.38) and the post-test (97.33) differed significantly, indicating that the participants used more varieties of vocabulary. Again, this was predictable because they wrote longer in the post-test: mere increase in word types does not prove they used more 'advanced' vocabulary (i.e. less frequent words) in writing. *Range* categorises all the words appearing in the texts into four levels (words belonging to base lists 1,2,3, and those not in the lists) and generates the percentages each category of words occupies in the texts. According to its analysis, the mean percentage of base list 1 words significantly decreased from 83.92% in the pre-test to 80.88% in the post-test; those of base list 2 and 3 words did not change statistically; and the mean percentage of non-list words significantly increased from 8.30% in the pre-test to 10.58% in the post-test. Although the words classified as 'not in the lists' include misspelled words (e.g. hibrid, erectricity), proper nouns (e.g. Tokyo, Fukushima), blended nouns (e.g. eco-bag, e-mail), acronyms (e.g. DVD, CO2), and even Japanese words (e.g. *uchiwa* [a round fan], *kairo* [a pocket warmer]), there were more academic or topic specific words (e.g. abolish, vital, fossil, dioxin) in the post-test. All in all, the analysis suggests that the learners used less easy words and more difficult words in the post-test essays.

To sum up, the participants' writing abilities (vocabulary selection, sentence construction, discourse structure) improved over the year. This was definitely true of fluency (length of writing), but accuracy (grammar, usage, mechanics) did not improve at all (or even suffered) and evidence of growing complexity (word level) was found. These outcomes were observed by the teachers while they were teaching. A comment in the teaching journal written by the male teacher on the 30th of November 2012 (the 58th lesson) reads: 'The amount of [students'] writing seems to have increased more than before; they became able to write much more'. According to him, explicit instructions in writing such as paragraph organisation and process writing were not given during the course. Instead, the students wrote a number of drafts throughout the year to prepare for their group presentations. In other words, emphasis in language teaching was placed on speaking, and writing activities were integrated naturally to enhance spoken production, which incidentally led to fluency in writing (he used the term 'multiplier effects' to explain this). As regards accuracy, the teachers did not correct errors for two reasons: following CLIL principles, they now accepted an SLA (second language acquisition) theory that making mistakes is part of language learning; they wanted the students to notice their errors and correct them by themselves.

These learning outcomes in writing, however, should be interpreted with caution. As previously mentioned, the students received six other English lessons besides the CLIL classes and their language learning gains outside school were unknown. Nonetheless, this study does infer that CLIL contributed to the students' language development in writing.

6. Conclusions

This article has explored the type of CLIL which has not been discussed in-depth in existing primary research papers – its 'weak/soft' version taught by language teachers for dual purposes with slightly more emphasis on language acquisition than on content learning. This is differentiated from CBLT in that characteristically CLIL principles and practice (i.e. the intentional integration of content, communication, cognition and culture) are realised in the classroom by trained CLIL language teachers. To look at it from students' perspectives, '[l]earners view this as part of language teaching but see it as an authentic way in which to use the language to learn non-language content.' (Coyle *et al.*, 2010: 22). In order to find possibilities of this 'gentle' CLIL approach in Japanese secondary education, research questions about the quality of teaching and the outcomes of learning were formulated. The answers are:

- (1) The students noticed there were distinct CLIL characteristics in the course that were different from other English lessons. In particular, they appreciated the cognitive activities, constructive learning, increased content knowledge and enhanced communicative competence (vocabulary and oral presentation) that CLIL teaching offered to them. Their self-evaluation of the development in the four skills (reading, listening, writing and discussion) was relatively modest.
- (2) The learners' overall written English proficiency developed, particularly in terms of fluency (the number of words they can write in a certain amount of time) and complexity (the variation of word types they can use in a limited length of time); accuracy (the number of errors per sentence) did not improve. It must be mentioned, as one of the limitations of this study, that these language learning gains cannot be totally attributable to the CLIL lessons, as the participants also learnt English outside the CLIL classroom.

It would be possible with certain reservations to conclude from these results that a 'weak' form of CLIL does work for Japanese secondary school students if they are taught by trained CLIL English teachers.

This article would not argue, however, that this somewhat watered-down variant of CLIL should be aimed at as the final goal of CLIL implementations. Rather, the intention here is to clarify CLIL's *raison d'être* at this stage in Japan where metaphorically speaking it could be seen as a seedling that could eventually be grown into a mature plant. To put it another way, for CLIL to take root solidly in the Japanese educational climate, the 'soft' version of CLIL should be recognised as an adapted, contextualised breed while, at the same time, its authentic, universal model (i.e. European CLIL) is pursued as the norm. This does not mean any kind of content-based instruction by language teachers is unconditionally accepted as a type of CLIL education. As repeatedly emphasised throughout this article, effective CLIL lessons are most probably conducted by teachers who have been adequately and amply trained in CLIL theories and skills.

The importance of CLIL teacher education provides another advantage of transplanting CLIL in new contexts. Both of the teachers in this study unanimously said in the follow-up interviews that they have developed professionally in terms of teaching skills and language skills (i.e. teaching content through English). As non-native CLIL language teachers are also learners of English and initial content teachers, the effects of CLIL on their professional competence are worth investigating in future research to see from a different angle how CLIL can complement language education in new contexts.

References

- Attali, Y. and Powers, D.: 2008, A developmental writing scale, *ETS Research Report*, 08-19. (Retrieved 11 September 2013 from <http://www.ets.org/Media/Research/pdf/RP-08-19.pdf>)
- Burstein, J., Chodorow, M. and Leacock, C.: 2004, Automated essay evaluation: The *Criterion* online writing service, *AI Magazine* 25/3, 27-36.
- Ball, P.: 2009, Does CLIL work?, in D. Hill and P. Alan (eds.), *The Best of Both Worlds?: International Perspectives on CLIL*. (32-43). Norwich Institute for Language Education, Norwich.
- Bentley, K.: 2010, *The TKT Course: CLIL Module*. Cambridge University Press, Cambridge.
- Coxhead, A.: 1998, An Academic Word List, *English Language Institute Occasional Publication*, 18. Victoria University of Wellington, Wellington, New Zealand.
- Coxhead, A.: 2000, A new academic word list, *TESOL Quarterly* 34/2, 213-238.
- Coyle, D.: 2007, Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies, *The International Journal of Bilingual Education and Bilingualism* 10:5, 543-562.
- Coyle, D.: 2008, CLIL: A pedagogical approach from the European perspective, in N. Van Deusen-Scholl and N. Hornberger. (eds.), *Encyclopedia of Language and Education, Volume 4: Second and Foreign Language Education* (2nd ed.), (97-111). Springer, New York.
- Coyle, D., Hood, P. and Marsh, D.: 2010, *CLIL: Content and Language Integrated Learning*. Cambridge University Press, Cambridge.
- Dale, L. and Tanner, R.: 2012, *CLIL Activities: A Resource for Subject and Language Teachers*. Cambridge University Press, Cambridge.
- Dalton-Puffer, C.: 2011, Content-and-language integrated learning: From practice to principle?, *Annual Review of Applied Linguistics* 31, 182-204.

ETS (Educational Testing Service), *Criterion* online writing evaluation: Scoring guide. (Retrieved 30 August 2013 from <http://www.ets.org/Media/Products/Criterion/topics/topics.htm>)

Harmer, J.: 2012, *Essential Teacher Knowledge: Core Concepts in English Language Teaching*. Pearson, Harlow.

Heatley, A., Nation, P. and Coxhead, A.: 2002, Range and frequency programs. (Retrieved 19 August 2013 from http://www.vuw.ac.nz/lals/staff/Paul_Nation)

Laufer, B. and Nation, P.: 1995, Vocabulary size and use: Lexical richness in L2 written production, *Applied Linguistics* 16/3, 307-322.

Marsh, D.: 2002, *CLIL/EMILE-The European Dimension: Actions, Trends and Foresight Potential*. Public Services Contract EG EAC, European Commission, Brussels.

Mehisto, P., Marsh, D. and Frigolos, M.: 2008, *Uncovering CLIL: Content and Language Integrated Learning in Bilingual and Multilingual Education*. Macmillan, Oxford.

MEXT (Ministry of Education, Culture, Sports, Science and Technology), Accredited International Baccalaureate schools in Japan. (Retrieved 05 September 2013 from http://www.mext.go.jp/a_menu/kokusai/ib/1307999.htm)

Nation, P. and Macalister, J.: 2010, *Language Curriculum Design*. Routledge, New York.

Pérez-Cañado, M.: 2012, CLIL research in Europe: Past, present, and future, *The International Journal of Bilingual Education and Bilingualism* 15/3, 315-341.

Scrivener, J.: 2011, *Learning Teaching: The Essential Guide to English Language Teaching* (3rd ed.). Macmillan, Oxford.

Ur, P.: 2012, *A Course in English Language Teaching* (2nd ed.). Cambridge University Press, Cambridge.

Watanabe, Y., Ikeda, M. and Izumi, S. (eds.): 2011, *CLIL (Content and Language Integrated Learning): New Challenges in Foreign Language Education at Sophia University, Volume 1: Principles and Methodologies*. Sophia University Press, Tokyo.

West, M.: 1953, *A General Service List of English Words, with Semantic Frequencies and a Supplementary Word-List for the Writing of Popular Science and Technology*. Longman, Harlow.

Wolfe-Quitero, K., Inagaki, S., and Kim, H.: 1998, *Second Language Development in Writing: Measures of Fluency, Accuracy and Complexity*. University of Hawaii, Hawaii.

Appendix

Sample essay and descriptor for *Criterion Score 2 (Insufficient)*

Japan use a lot of energy. What we should do for decrease energy consumption? I have two ideas. First, we stop using oil. Japan uses too much oil. Recently, however, Japan developed many kinds of solar cars. I think we should use solar cars more than oil but price of solar car is too expensive. If we decrease solar car's price, we needn't use much oil. Next, we should turn off the lights and TV when we don't use it. It's also good for own light's cost for a month. Second idea is easy to do now but, first idea is difficult to do because we need discuss with car company. I think government should work for this problem and decide an average price. If we can do these two things, it's good for environment and us. [137 words]

- Provides little information and makes little attempt at development.
- Is very disorganized or too brief to detect organization.
- Exhibits little control over sentence structure.
- Contains inaccurate word choices in much of the essay.
- Is characterized by misspellings, missing words, and incorrect word order; errors in grammar and conventions are severe enough to make understanding very difficult in much of the essay.

Sample essay and descriptor for *Criterion Score 3 (Uneven)*

My plan for cutting our country's energy consumption by 60% are 2 points. First idea is using bicycle or on foot when we go somewhere near by. If we do this plan, we can cut down oil energy because we do not need to drive a car. Moreover, we can help environmental problems like polluting air. Less using oil energy means protect clean air. Second idea is turn off the light when we do not need to. If we do not use lights everytime, we can cut down nuclear energy. It means we can use natural energy instead of nuclear energy because we do not need to use so much energy. Natural energy is less cost than nuclear energy. Also, not danger. This is an important point. Citizens who living near nuclear gas plant are always feel fear because of 3.11 [the devastating earthquake and tsunami that hit the north east coast of Japan on the 11th of March 2011]. So we can solve these problems if we do not use so much lights. In conclusion, accepted my plans means cut down energy consumption. At the same time, good effect for environment and people. My plans are everyone easy to do. [182 words]

- Provides limited or incomplete information; may be list-like or have the quality of an outline.
- Is disorganized or provides a disjointed sequence of information.
- Exhibits uneven control over sentence structure.
- May have some inaccurate word choices.
- Contains errors in grammar and conventions that sometimes interfere with understanding.