

RESEARCH

Open Access



Effects of using the first principles of instruction in a content and language integrated learning class

Hao Hao^{1*}, Hitoshi Susono², Xuewang Geng¹, Li Chen¹ and Masanori Yamada³

*Correspondence:
kakukou@mark-lab.net

¹ Graduate School of Human-Environment Studies, Kyushu University, 744 Motoooka, Nishiku, Fukuoka 819-0395, Japan

² Graduate School of Teacher Education, Mie University, Mie, Japan

³ Faculty of Arts and Science, Kyushu University, Fukuoka, Japan

Abstract

The aim of this study was to examine the effects of Content and Language Integrated Learning (CLIL) designed according to the First Principles of Instruction (FPI). A 15-h Japanese CLIL course was implemented. A total of 16 university students attended the course and data were collected from multiple sources, including learning tests, questionnaire feedback, and dialogues in group discussions, were collected and examined. Analysis showed that students' learning outcomes, including basic Japanese proficiency, intercultural communication content, and writing skills, were statistically significantly improved. Students had a high level of awareness of the elements of FPI designed in the course. In addition, all the FPI elements had a positive impact on basic Japanese proficiency except for the application element: the problem-centered, application, and integration elements positively impacted intercultural communication content and writing skills. The results show that students displayed individual differences in using the worksheet to summarize their writing ideas. Students spent most of the time in the group discussions in their native language. Even when Japanese was used, individual Japanese words were used rather than whole sentences in most cases. The results of the quantitative and qualitative analyses showed that the use of problem-centered theory FPI had a positive impact on the design of the CLIL. However, attention is needed to students' individual differences and the guidance of students in applying basic language knowledge in problem-centered learning activities. Finally, it notes points that should be considered when designing CLIL in the future.

Keywords: Instructional design, Japanese as a foreign language, Content and language integrated learning, First principle of instruction

Introduction

In recent years, following the progress of globalization, research on foreign language learning (FL) has proliferated. Various types of educational methods are being carried out in FL; among these methods, one is called Content and Language Integrated Learning (CLIL). The most prominent feature of CLIL classes is that it is possible to learn a foreign language while acquiring specialized knowledge (Coyle et al., 2010). Also, in the classroom environment, CLIL occupies a position between structure-based instruction and immersion, with intense flexibility that balances the teaching of basics such as grammar and the practical

use of the language (Dale & Tanner, 2012). The CLIL class has made some achievements in FL. CLIL is widely used in various courses in different subjects, such as mathematics, geography, and science (Dourda et al., 2014; Leal, 2016; Ouazizi, 2016). The results of these practices indicate that students' subject knowledge and language knowledge improved. Although research on CLIL practices has been ongoing since its introduction in the 1990s, its advocates understand CLIL and the varied interpretation of this approach in different ways. Coyle et al., (2010, p. 1) provide a succinct definition that refers to CLIL's specific features: "Content and Language Integrated Learning (CLIL) is a dual-focused educational approach in which an additional language is used for learning and teaching of both content and language." According to this definition, CLIL can include a wide range of educational practices. Some studies have treated CLIL as an "educational approach" (Mehisto, 2008; Pérez-Vidal & Juan-Garau, 2010), while others have considered CLIL to be actual instructional techniques and practices used in classrooms (Ball & Lindsay, 2010; Hüttner & Rieder-Bünemann, 2010). Other scholars have considered CLIL primarily from a course perspective (Langé, 2007; Navés & Victori, 2010). Pladevall et al. (2011) point out that CLIL offers flexibility in course design and scheduling, yet the balance between language and content is complex in course design. Meyer (2010) proposes a design sheet to be used when designing CLIL courses, but he also points out that CLIL is not specialized in terms of course design theory. Furthermore, Hao and Yamada (2021) pointed out that few studies have analyzed CLIL practices from instructional design (ID) perspectives during these 20 years. ID is the framework in which teachers follow planned teaching and learning steps (Richards & Lockhart, 1994), including a wide range of fields (Reiser, 2001). ID has been shown in numerous studies to be an effective tool for improving course effectiveness (Hernandez, 2016; McGee & Reis, 2012; Richards & Lockhart, 1994). Therefore, giving consideration to ID, which has become a standard theory in the area of teaching design and syllabus (Honebein, 2019), in CLIL may provide a new perspective for CLIL improvement.

The First Principles of Instruction (FPI), one ID theory, is an integrated one (Merrill, 2002), and it is an effective solution as a principle necessary for the course. The widely endorsed FPI, developed under the influence of constructivism, outlines the requirements necessary for achieving five effective learning environment goals. There have been many empirical studies showing that students' academic performance and satisfaction with the course improve after learning through courses designed using FPI (Gardner, 2011; Lo, 2015; Tu & Snyder, 2017). To explore how to use FPI to improve CLIL course design, it is desirable to establish a pedagogical foundation for the practice of CLIL classes in terms of FPI.

In this study, we designed and implemented a course combining elements of CLIL with FPI in the context of intercultural communication in Japanese and evaluated the effects of the CLIL course designed as a formative evaluation. The purpose is to discuss the effectiveness of CLIL courses designed using FPI, derive FPI elements that positively impact CLIL based on the results, and make recommendations for future CLIL designs.

Literature review

Content and language integrated learning

CLIL, proposed in Europe in the early 1990s, is defined as an educational approach used for both content and language learning and teaching (Coyle, 2007). CLIL has two central characteristics and consists of four elements (4Cs). Its two central characteristics are

(1) the integration of language and learning content involving a lesson design via which teachers teach educational content in a foreign language and (2) the integration of communication and intercultural understanding to deepen students' cultural understanding through natural language communication. The 4Cs are (a) content, posing a progression in knowledge, skills, and understanding related to specific elements of a defined curriculum; (b) communication, using language to learn while learning to use language; (c) cognition, developing thinking skills that link conceptual formation, understanding, and language; and (d) culture, exposure to alternative perspectives and shared understandings, which deepen the awareness of otherness and self (Coyle, 2006). Coyle (2006) called these four elements the 4Cs and suggested that a successful CLIL class should include them all.

Sabet and Sadeh (2012) pointed out that CLIL classes can improve students' confidence to use a second language and significantly improve their language proficiency and thinking ability. CLIL can promote the use of foreign language learning strategies and geographic knowledge, while at the same time improving students' reading comprehension, vocabulary, and learning satisfaction (Dourda et al., 2014). Kanamura and Miyajima (2016) implemented CLIL in Japanese language education for international students whose primary purpose is to learn Japanese law. Specifically, lessons were held as a one-year lecture series designed to support group activities. At the end of the course, students' legal knowledge and Japanese speaking skills were shown to have improved significantly.

Although the effectiveness of CLIL for content and language learning has been widely validated since its introduction in the 1990s, there is still some controversy with the design of CLIL courses. CLIL is considered a combination of foreign language learning and content-based instruction (CBI) (Cenoz et al., 2013). CBI has been defined as "the teaching of content or information in the language being learned with little or no direct or explicit effort to teach the language itself separately from the content being taught" (Richards & Rodgers, 2001, p. 204). CBI can be used to develop learners' language proficiency by providing them with meaningful content (Crandall, 1999). CBI can be effectively applied to foreign language learning in various contexts, but it is not a specific instructional theory and therefore requires a specific course design when using CBI (Heo, 2006). There is the same issue as CBI when teaching in CLIL. CLIL requires an emphasis on both content and language, but the ratio of these two aspects is not strictly required, which allows it to be used very widely, but at the expense of accuracy (Cenoz et al., 2013). Thus, the definition of CLIL allows it to be widely used in courses that include specialized knowledge learning and language learning, but it does not have a strict design theory (Pladevall et al., 2011). Filice (2020) notes that although the learning effectiveness of CLIL experiment about subject knowledge was positive during the group discussion, there were difficulties on the part of language production, especially in formulating questions and in professional terms. Filice (2020) also points out that simply combining content and language in the CLIL class may not be effective for learning content and language. Thus, further specification and design of the course are needed, which requires further experimentation with CLIL involving different designs to provide more empirical research findings. Agustín-Llach (2016), for

example, conducted a three-year controlled experiment on primary school students between and 9–10 years of age. The CLIL and non-CLIL groups completed a writing task, and he then compared learners' vocabulary scores. The results showed that although the CLIL group's scores were higher than those of the non-CLIL group, there was no significant difference between the two.

Through practice, the CLIL researchers have concluded that students need more opportunities to apply what they have learned and the language in the CLIL classroom (Evnitskaya, 2014; O'Dwyer & de Boer, 2015; Yufrizal & Huzairin, 2017). One commonality among these findings is the focus on active approaches to learning. Yufrizal and Huzairin (2017) pointed out that giving students real-life problems can increase their motivation and give them more opportunities to practice what they have learned. This recommendation coincides with the general prescription of using real-world problems when teaching any complex skill (van Merriënboer, 1997).

Instructional design

Several researchers have attempted to improve CLIL from an ID perspective (Langé, 2007; Navés and Victori, 2010; Meyer, 2010). The ID theory refers to the theory that provide help and guidance for people to learn better (Reigeluth, 1999). ID and Technology contains a wide range of fields. According to the definition of Association for Educational Communication and Technology (AECT), its fields include six categories of activities or practices: (a) design, (b) development, (c) utilization or implementation, (d) management, (e) evaluation, and (f) analysis (Seels & Richey, 1994). And the fields it contains are still expanding (Reiser, 2001).

Although Meyer (2010) designed worksheets for use in CLIL course design based on the four features (4Cs) of CLIL, he also pointed out at the same time that CLIL is not a professional course design theory, and more research is needed to explore the design of CLIL. Authors (2021) reviewed nearly 20 years of CLIL practice papers and revealed that there is little practice in CLIL practice to design courses from the perspective of ID. ID encompasses numerous theories and models and has been widely used across fields (Hernandez, 2016; McGee & Reis, 2012; Reiser, 2001; Richards & Lockhart, 1994). The use of ID for the design of CLIL may provide new ideas for the design of future CLIL.

As described in the introduction section, the basis for the proposed CLIL is constructivism (Coyle et al., 2010). In the existing ID theories, some theories have been proved to be the principles required in course design. In 2002, Merrill tried to integrate these principles and proposed the First Principles of Instruction (FPI) theory. As a strategy common to numerous ID models and theories, FPI proposed strongly under the influence of constructivism, it summarizes the five requirements necessary for realizing an effective learning environment. Also, FPI is a problem-centered theory that promotes active student learning. Building instruction around life problems, or ones that the learner will face after the class is complete, is key to the design and delivery of effective instruction (Merrill, 2002, 2013; van Merriënboer, 1997).

Therefore, considering FPI, which is a problem-centered theory based on constructivism, in CLIL may provide a new perspective for CLIL improvement.

First principles of instruction

In 2002, Merrill proposed the FPI, which integrates many ID and learning models. As a common strategy for many ID models, the FPI, which was strongly influenced by constructivism, is a compilation of the five requirements necessary to realize an effective learning environment. Specifically, it is organized as follows: “(a) Problem-centered: Learning is promoted when learners are working to solve real-world problems. (b) Activation: Existing knowledge learning is promoted when activated as a basis for new knowledge. (c) Demonstration: Learning is promoted when new knowledge is presented to the learner. (d) Application: Learning is promoted when the learner applies new knowledge. (e) Integration: Learning is promoted when new knowledge is integrated into the learner’s world” (Merrill, 2002, pp. 43–44) (Fig. 1).

FPI emphasizes the development of the course in 5 principles (Merrill, 2002, 2013). At the beginning of the course, the teacher will give a problem for the students to solve. Students are expected to solve the problem through learning activities that include activation principle, demonstration principle, application principle and integration principle. As students become more capable, the problems they solve become progressively more difficult. This is to ensure that the new knowledge students have learned can be applied. Gardner (2011) performs practices based on the FPI and compares an experimental group using FPI with a control group not using it; the test results of the experimental group improved significantly. Students who studied microevolution using problem-centered instruction were more confident in their ability to solve problems in the future. In addition, FPI has been used in the design of FL courses. Lo et al. (2018) developed a reversal course applying FPI and showed that students’ scholastic ability in mathematics, physics, and the Chinese language improved through practice.

Research questions

This study aims to design, conduct, and evaluate FPI-based CLIL classes in order to provide more practical experience in the design of CLIL courses to improve teaching and learning. In this study, CLIL using FPI was designed and its educational effects, such as the learning performance of vocabulary and grammar, the understanding of content, and the improvement of writing ability, examined. In addition, Filice (2020) pointed out that most past research on CLIL has been analyzed from the teacher’s point of view, often neglecting students’ perceptions, yet student feedback is vital for future course design. This prompted us to investigate in this study how the students perceive it. Therefore, we analyze the students’ perspectives as well as learning behaviors and suggest improvements for future CLIL. This study sought to answer the following research questions:

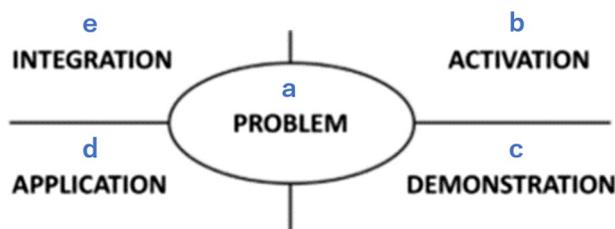


Fig. 1. 5 principles of FPI (Merrill, 2002, p. 45)

RQ 1. How effective is the learning outcome of CLIL courses designed using FPI?

RQ 2. How aware are students of the designed FPI elements?

RQ 3. What elements of FPI have a positive impact on CLIL courses?

Methodology

FPI-based CLIL

In this study, the Japanese CLIL course following the FPI element was designed based on the framework for problem-centered instruction by Merrill (2013). The process of CLIL in this study is presented in Tables 1 and 2. Some authors have demonstrated the successful application of FPI elements in a variety of settings (Gardner, 2011; Gardner et al., 2009; Lo et al., 2018; Mendenhall, 2012).

The Japanese CLIL in this study is designed following FPI elements. There were two main parts involved in this study setting procedure. First, according to the requirements of the course, the students' learning content and learning objectives were set. The learning content was intercultural communication. The learning objective was divided into CLIL's content objective, to understand the concepts of intercultural communication presented, and CLIL's language objective, to learn the basics of the Japanese language and use Japanese to write.

Next, every lesson in this study is designed according to the five elements of FPI: Problem-centered, Activation, Demonstration, Application, and Integration. Table 1

Table 1 Example of a 90-min lesson

FPI principle	Definition	Course design	Time (percentage)
Problem-centered	The quality of being problem-centered, which means basing learning experiences on real-world problems	Each lesson is designed with a problem related to intercultural communication that exists in life. Students will need to solve these problems in class For example, if you come to Japan, how do you make friends with the foreigners you meet for the first time?	5 min (6%)
Activation	Include having students recall or describe their prior experience related to the content to be taught	When presenting problems to students, use what has been covered in class before, or what students have experienced in their lives to explain	15 min (17%)
Demonstration	Include modeling examples of the problem to be performed	When explaining new knowledge, use photos, videos, and other multimedia tools to show students examples	20 min (22%)
Application	The opportunity to practice the whole problem and receive feedback	Students work in groups to discuss how to solve the problems posed by the teacher and come up with solutions	30 min (33%)
Integration	Include reflecting on, presenting, and thinking about how to use knowledge	Students present the results of the group discussion and summarize what they have learned and are asked to think about how they could apply what they had learned in their future lives	20 min (22%)

Table 2 Explanation of learning activities

	Content of knowledge	Task
Activity 1	<p><i>Self-disclosure</i>: A process of passing on information about yourself to someone else, whether people intend to or not. For example, confiding in others about various aspects of oneself, such as one's attitudes, opinions, interests, work, and personality</p> <p><i>Johari Window</i>: proposed by Joseph Luft and Harrington Ingham in 1955, this is a technique designed to help people understand their relationship with themselves and others. It can divide the size of one's self-disclosure</p>	Use the worksheet and discuss with group members to create students' own Johari window
Activity 2	<p><i>Stereotype</i>: a set idea about what a particular type of people</p> <p><i>Prejudice</i>: a stereotype accompanied by negative feelings, which means the opinion is not based on reason or actual experience</p> <p><i>Discrimination</i>: a prejudice further combined with behavior, which means treating a person or particular group of people differently in a worse way</p>	Using the worksheet and discussion with the group members, explore students' own inherent impressions of the Japanese people and the reasons for such impressions

demonstrates how the principles were implemented in the course for the experimental condition.

Participants

The subjects for this study were 16 third-year students in a course for Japanese language majors at a university in China. The students who participated in this course were between the ages of 20 and 21, with 12 females and 4 males. Their general Japanese language ability was at the N2- or N3-level of the Japanese Language Proficiency Test. The N2-level refers to students who "The ability to understand Japanese used in everyday situations, and in a variety of circumstances to a certain degree." and the N3-level refers to students who "The ability to understand Japanese used in everyday situations to a certain degree." (Japan Foundation & Japan Educational Exchanges and Services, 2012, p. 78). Before the class, the researcher introduced to the students the purpose of this study, the collection and processing methods of data, and obtained the students' approval. Classes were conducted three hours a day for five days, and the content was intercultural education. All lesson content was taught in Japanese.

Data collection

Before starting the course, participants were asked to complete a pretest in the data collection. After the class on the last day, a posttest and postquestionnaire were conducted. The post-questionnaire was used to evaluate whether the elements of FPI were recognized. The FPI questionnaire was based on the Academic Learning Time Questions (TALQ) by Frick et al. (2009). For example, the question about the center of the problem was, "I solved authentic problems or completed authentic tasks during this course"; the question about the activation was, "In this course, I was able to recall, describe, or apply

my past experiences to help me connect to what I was expected to learn.”; the question about the demonstration was, “Teacher demonstrated the skills I was expected to learn during this course.”; the question about the application was, “I had opportunities to practice or try out what I learned during this course.”; the question about the integration was, “I see how I can apply what I have learned in this course to real-life situations.” The FPI questionnaire used a 5-point Likert scale, with 1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Not sure*, 4 = *Agree*, and 5 = *Strongly agree*. Among the questionnaire, there are 4 questions under the problem-centered principle, 5 questions under the activation principle, 4 questions under the demonstration principle, 4 questions under the application principle, and 5 questions under the integration principle. The questions and reliability are provided in Appendix 2.

The test consisted of three parts. The first part was the Japanese Language Proficiency Test (JLPT), including N1- and N2-level questions, which consisted of six reading comprehension questions and two listening comprehension questions. The N1 questions were worth 2 points, and the N2 questions were worth 1 point, comprising a total of 12 points. The test on content was an essay question. Students were asked to answer this question: Please define and explain the difference between stereotypes, prejudice, and discrimination. This is a question that has a correct answer. The reference answer is as follows. Stereotype is meant that a categorized image one has a group of people. Prejudice is meant that a stereotype accompanied by negative feelings (although some argue that it is not necessarily negative). Discrimination is meant that prejudice is further associated with behavior. These correct answers are explained in class. For each concept, one point is given to a description of the concept and one point is given for an explanation of the difference between the concepts. Each concept is worth 2 points, for a total of 6 points. The third part of the test required students to write about 300 words. The theme was, “Write your own thoughts on how to use the aforementioned theories when communicating cross-culturally.” For the evaluation standard, we referred to the “JF Japanese Education Standard” (JF standard for Japanese-language education, 2010) and made an evaluation standard that matched the Japanese ability of the students. The evaluation was based on three items: “content,” “grammar/vocabulary,” and “composition.” The maximum score for each evaluation item was 4, the minimum score was 1, and thus, the maximum total score was 12. For example, when grading content, a score of 4 is given if the main content is explained in detail so that the reader can understand the central idea. A score of 3 is given for lack of explanation and inability to fully understand the content. A score of 2 is given if the reader can vaguely understand the subject matter. 1 point for lack of explanation that makes content difficult to understand. When grammar/vocabulary are scored, a score of 4 is given if the grammar and expressions of words related to the topic are accurately applied. A score of 3 is given if there are some word grammatical errors but the sentence can be understood. A score of 2 is given if some information is not conveyed due to grammatical errors. A score of 1 is given if the sentence is incoherent and does not convey the message due to grammatical errors. When grading the composition of the constitution, a score of 4 is given for paragraphs and sentences that are connected with appropriate words and phrases and have a clear paragraph composition. A score of 3 was given for having some sentences that were difficult to understand in relation to each other. A score of 2 was given for a simple list of

ideas with no connection. A score of 1 was given for fragmented sentences and words with no sentence formation. Specific evaluation criteria are detailed in Appendix 1.

In order to increase the opportunities for students to use what they have learned in class, teacher assigns tasks to students based on class content. Students are called upon to solve the tasks through group discussions. Activity 1 is a learning activity of self-disclosure. Activity 2 is a learning activity about stereotype and prejudice and discrimination. Details about the 2 activities are described in Table 2. In order to capture the relationship among the 4Cs of CLIL that the students focused on when solving the problem, and learning outcomes, and the 5 principles of the FPI designed, the students' group discussions were recorded and coded for analysis.

The worksheets that the students created during the activity are shown in Figs. 2 and 3.

Data analysis methods

First, the data were verified as having a normal distribution using the Statistical Package for Social Science (SPSS) version 26 software program. Since the data were not normally distributed, the Wilcoxon signed-rank sum test was used to analyze the pretest and post-test, and Spearman's rank correlation coefficient method was used to analyze students' learning outcomes. In addition to the mean, the median is also stated in the result. The total score of the basic test was 12, the content test was 6, and 12 for writing. The mean and standard deviation of students' awareness of the designed FPI elements were analyzed. On the other hand, the qualitative data which were collected through group discussion were analyzed through MAXQDA 2020.

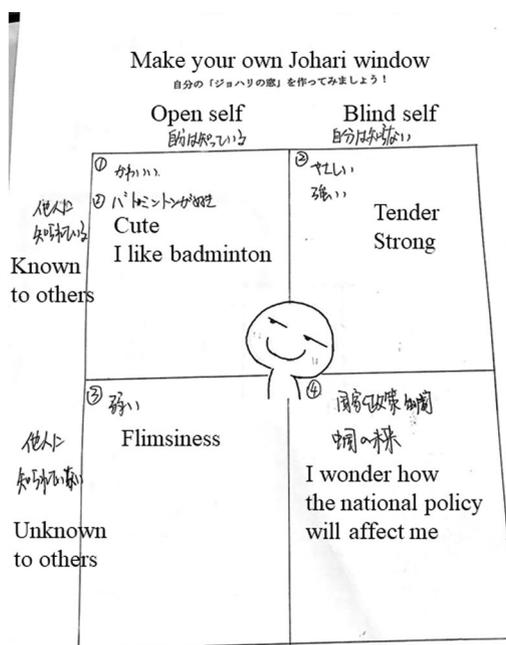


Fig. 2 An example of the worksheet used in Learning Activity 1



Fig. 3 An example of the worksheet used in Learning Activity 2

Result and discussion

Research question 1: how effective is the learning outcome of CLIL courses designed using FPI?

Wilcoxon signed-rank test was used to measure the significance of changes in the pre-post learning outcomes. The results are shown in Tables 3, 4 and 6. As the results shown, the students' learning outcomes have been effectively improved through this CLIL course.

Table 3 indicates that the students' basic Japanese proficiency test mean value improved from 6.13 (out of a full score of 12, $SD=2.09$) to 10.00 ($SD=1.73$). Additionally, a significant difference ($p=0.001$) was noted in the distribution of basic Japanese scores. From this result, we can deduce that the students' basic Japanese knowledge improved during this CLIL course.

Table 4 indicates that the students' content knowledge test mean value improved from 2.00 (out of a full score of 6, $SD=1.51$) to 4.62 ($SD=1.50$). Additionally, a significant difference ($p=0.001$) was noted in the distribution of content knowledge scores. The content question is to explain the following 3 concepts; stereotype, prejudice and discrimination. These concepts have corrected answers which have presented to the students in class. The scoring criteria were determined after discussions with Japanese teachers based on the JF Standard. Each concept is worth 2 points, for a total of 6 points. The minimum score is 0 and the maximum score is 6. The example of student pre-post answer on the learning content section is illustrated in Table 5. From this result, we can understand that the students' content knowledge improved during this CLIL course.

Table 6 indicates that students' writing test mean value improved from 4.19 (out of a full score of 12, $SD=1.00$) to 9.38 ($SD=1.64$). Additionally, a significant difference ($p<0.001$) was noted in the distribution of content knowledge scores. The students' writing was graded in 3 dimensions: Content, Vocabulary/grammar, and Constitution, according to the <Japan Foundation standard for Japanese-language education 2010>. Each dimension is scored on a scale of 1–4. Therefore, the maximum score is 12 and

Table 3 Wilcoxon signed-rank test results of pre-post Basic Japanese Proficiency Test Score ($N = 16$)

Test	Mean	Median	SD	z	p value
Pre-basic	6.13	6.00	2.09	3.305	0.001
Post-basic	10.00	10.00	1.73		

* $p < 0.05$; ** $p < 0.01$ **Table 4** Wilcoxon signed-rank test results of pre-post Content Knowledge Test Score ($N = 16$)

Test	Mean	Median	SD	z	p value
Pre-content	2.00	2.00	1.51	3.424	0.001
Post-content	4.62	5.00	1.50		

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

the minimum score is 3. The example of pre-post answer of writing test is illustrated in Table 7. From this result, we can understand that the students' writing skills improved during this CLIL course.

In sum, it can be understood that the CLIL course designed by FPI has positive effects on language knowledge, writing skills, and content knowledge. In line with this research finding, previous studies also showed that a course designed using FPI has positive impacts on students' learning outcomes. Lo et al. (2018) found that by using FPI in a flipped course, students develop their scholastic ability and Chinese language ability, and learn how to generate and organize ideas.

Because CLIL is characterized by using a foreign language to teach specialized knowledge, the goal is to improve both the foreign language and the specialized knowledge, which Coyle et al. (2010) notes not easily accomplished. Ennis (2015) also notes that CLIL courses require explaining what students do not know in a language they do not understand, which is a very challenging task. In addition, Agustín-Llach (2016) points out that there were no significant differences in student achievement between the CLIL and non-CLIL courses. It can thus be seen that if the language and content are combined, there is a possibility that the desired learning goals will not be achieved. However, this CLIL course designed using FPI guarantees both content and language learning outcomes.

In addition, the writing test result proves that the approach of having students use worksheets to organize their thoughts in this course is effective. For example, the worksheet used in Activity 2 was to help the students present a problem in 3 stages. This helped students better understand the three-paragraph writing method. It was also evident from the example that the students' essays were composed of introduction-first-second-summary. In fact, most students used this composition in writing test at this time. The teacher did not emphasize this writing format to the students during the lesson. It can be seen that the students did a good job of internalizing this form of composition on their own through this course. However, regarding the SD in Table 6, the post-SD (1.64) was also higher than the pre-SD (1.00). Students' writing skills were strengthened

Table 5 Example of pre-post responses on the learning content

	Original text and translation	Score and comment
Pre-test	<p>Original text: 差別は差を持って人を扱うこと。 「Secret Superstar」(神秘巨星)という映画は、男女差別のことをあらわした。インドでは、女の地位がすごく低い。男のアクセサリーといっても過言ではない。</p> <p>Translation in English: Discrimination is treating people differently. The movie “Secret Superstar” (Secret Superstar) shows the discrimination between men and women. In India, the status of women is very low. It is no exaggeration to say that they are accessories to men</p>	<p>Score: 1 Comment: Only discrimination is explained. There is no mention of the other two concepts. Nor does it state the difference between them</p>
Post-test	<p>Original text: ステレオタイプというのは、ある集団に対する固定的な態度である。たとえば「山西省の人はみんな全部金持ちだ」。このように、個人的な特徴を不当的に一般化して、集団の全部人の特徴にする。 偏見とは、片面である人や集団、地域を考える。たとえば「女はリーダーになれない」。人を欠点ばかりに見られる。 差別とは、偏見を行動につける。たとえば、就職する時、ある会社は男だけほしいが女はいらない。</p> <p>Translation in English: A stereotype is a fixed attitude toward a group of people, for example, “All the people in Shanxi Province are rich.” In this way, personal characteristics are unfairly generalized to characterize all the people in a group Prejudice is considered a view of a person, group, or region that is one-sided. For example, “Women cannot be leaders.” People are seen only in terms of their shortcomings Discrimination is putting prejudice into action. For example, when getting a job, one company wants only men but not women</p>	<p>Score: 6 Comment: The 3 concepts are explained, and examples are given to illustrate the differences between them</p>

Table 6 Wilcoxon signed-rank test results of pre-post Writing Test Score (N = 16)

Test	Mean	Median	SD	z	p value
Pre-writing	4.19	4.50	1.00	3.535	0.000
Post-writing	9.38	9.50	1.64		

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

through this lesson, but the disparity between students also widened. Although the design for writing is effective for the whole course, the degree of understanding of writing content may differ due to individual differences. Allison et al. (1998) point out that higher education students undertake limited writing exercises in the classroom and

Table 7 Examples of students' pre-post of Writing Test

	Original text and translation	Score and comment
Pre-test	<p>Original text: ステレオタイプ、偏見や差別、以上三つの行為はすべて個別の行動から全体化し、否定的な感情を持ちます。私たちは公正の態度を持ちべきで、正しい判断をします。</p> <p>Translation in English: Stereotyping, prejudice, and discrimination are all three behaviors that can be generalized from individual actions and carry negative feelings. We should have an attitude of fairness and make the right decisions</p>	<p>Score: 3 Comment: Just a brief explanation of the meanings of the three words There is no clear paragraph All of the text are simple sentences, and there is no connection between them. It is impossible to understand what the writer is trying to say</p>
Post-test	<p>Original text: 今回の集中講義の授業で、異文化について勉強しました。その中で、三つのことをこれから大切にしたいです。まず、価値観の事です。この前は異文化の影響で各国の人の価値観が異なるところがあるという点が知っています。昔中国の諺には「酒逢知己千杯醉、話不投机半句多」という話があります。その意味は知己ができたなら、切っても切れない話がたくさんあり、逆に話が合わないと言半句さえも多いです。つまり、価値観のことが指します。そのことわざによると、価値観や志向と合わない人と友達になる必要がないと思いましたが、今は考えを変えました。それは日本人の思いやりのおかげで、包容すべきだという点が解りました。次は偏見や差別です。偏見ということはステレオタイプに否定的な感情を伴ったもので、ただし、必ずしも否定的なものではないという主張もあります。差別は偏見がさらに行動に結びついたものです。すなわち、偏見は人々の意識で、差別は偏見による行動するということです。ステレオタイプを通してではなく、自分自身の目で固まりの人たちや物事をちゃんと見ていく姿勢を養うように、正しい判断を下します。最後はイメージです。国人も社会も、イメージが大切です。</p> <p>Translation in English: In this intensive class, I learned about different cultures. I would like to emphasize three things from now on First, values. Before this, I know that people from different countries have different values due to the influence of different cultures. There is an old Chinese proverb that says, "When I meet a crony, I'd like to drink tons of wine with him, but if there is a nuisance, I'll be reluctant to say just a few words." The meaning of this saying is that if you get to know each other, there are many things you can talk about inseparably, but if you don't talk to each other, there will be many half-words. In other words, it refers to values. According to that proverb, I used to think that there was no need to be friends with someone whose values and orientation did not match mine, but I have now changed my mind. I now understand that it is because of the compassion of the Japanese people that we should be inclusive Next comes prejudice and discrimination. Some argue that prejudice is a stereotype accompanied by negative feelings, but not necessarily negative. Discrimination is prejudice further coupled with behavior. In other words, prejudice is people's awareness and discrimination is their behavior due to prejudice. We make the right decisions so that we develop a proper attitude to see people and things in solidarity through our own eyes and not through stereotypes Finally, there is image. The image of a country, a person, or a society is important</p>	<p>Score: 10 Comment: There are clear paragraphs, and the paragraphs are connected by successive words. Although there are some grammatical errors, this student makes the point clearly. Also, the use of examples makes the content informative</p>

require one-to-one writing instruction. In this course, although the teacher provided a worksheet to help students organize their writing ideas before they wrote, the teacher did not show students specific examples of writing. This may have resulted in students not understanding the use of the worksheet, thus causing an increase in personal differences due to their own comprehension. This is also expressed in the dialogue.

Example 1. Use the worksheet to make the respective Johari window.

Group 1

Scene 1

G1-S2: Have you all finished writing? I have finished writing all of them.

G1-S3: *Do I need to continue writing? Or is it enough to write only one part of it?*

G1-S1: You need to write all of them.

G1-S2: I think I should have drawn this part a little bigger.

(4 seconds)

G1-S3: Well, then my first part should be bigger too.

Scene 2.

G1-S3: *Eh? Which position should this be written in?*

G1-S2: On love, something

G1-S3: I do not know.

G1-S1: (laughs) Love men and women or both.

G1-S3: *Eh? This is not it? It is the second one, right?*

G1-S2: Position 2, the second one is empty.

Among the four students in Group 1, one student was not clear about the use of the worksheet. Moreover, this student asked how to use the worksheet two times during the discussion. The student who asked the question finally completed the worksheet after constantly checking with the other group members. This is consistent with the results in Table 6. Students experienced some differences in understanding when using the worksheet.

Group 2

G2-S2: I want to write myself やさしい (kind).

G2-S4: That I am かわいい (lovely).

G2-S2: *This should be written in the part that you know that others do not know, right? Is it written in position 3?*

G2-S1: Yes, here, you know, but others do not know.

One student in Group 2 also had a question about using the worksheet. The group members gave clear answers. Again, this group's discussion is consistent with the results in Table 4 in that there are individual differences among students.

Group 3

G3-S4: Been on a diet.

(laughing)

G3-S3: But you are not really doing the exercise to lose weight either.

G3-S1: Yeah, I did not.

G3-S3: Been studying English.

G3-S1: Already written.

(3 seconds)

G3-S3: *Well, we are at an impasse.*

(...)

G3-S2: *So what should be written in the third position?*

G3-S4: It does not have to be a bad aspect. It can be a good one too, right?

G3-S2: But the good aspects are already known to others.

(8 seconds)

G3-S3: *OK. Once again, there is an impasse.*

It can be seen from the conversation that one student in Group 3 also raised questions about using the worksheet. However, the problem raised was not solved in the end. The discussion in Group 3 also showed consistency with the results in Table 6. In particular, asking questions that are not answered can exacerbate personal differences among students.

Group 4

G4-S1: *The second position is to write what others know and what I do not know, right?*

(5 seconds)

G4-S3: Yeah, the second position is what others know and what you do not know yourself.

G4-S1: This is the Johari window in ... others know that I do not know. OK, then a little more specific is...

G4-S4: It is your interest and what you like to do, and so on.

G4-S1: Like habits! It is one of those little habits that you know but others do not. And then there is the kind of mantra you have, you do not realize it, but others do.

G4-S4: Yeah, verbal words must be known to others because you always say them.

G4-S2: *Oh, I see, some little gestures and habits!*

G4-S1: Yeah, Bad habits are something that others know about you, but you do not know about yourself.

G4-S4: Okay, so write about some bad habits in this position.

A student in Group 4 also asked a question about the worksheet. The group members responded positively, and the student who asked the question eventually solved the problem. It is also evident from the conversations in Group 4 that there were personal differences among the students, consistent with the results in Table 6.

From the conversations, it can be seen that all four groups of learners had problems with not knowing how to use the worksheet or what kind of content should be recorded in which position in the worksheet. Furthermore, it can be seen that each group addressed the issues raised by the members in different ways. Groups 1, 2, and 4 all solved this problem through discussion, but Group 3 did not solve these problems in the end. This is consistent with the results in Table 6. Students experienced some differences in understanding when using the worksheet, and problems may arise that cannot be resolved in the group, as demonstrated by the conversations in Group 3. This can also cause individual differences between students to grow. Students who already understand how to use worksheets can organize their writing ideas better through the use of the worksheet, while students who do not understand and cannot get answers from the group members become more confused. When designing a course to solve this problem, it is necessary to understand the learning situation of each student and provide support to the individual. In addition, the teacher should evaluate the ease of use of the worksheets before the formal course to improve the effectiveness of the worksheets.

Research question 2: how aware are students of the designed FPI elements?

Through this course, the students have a stronger awareness of the FPI principles followed in the design of the course. Table 8 indicates the students’ awareness of the designed FPI elements. In the results for the questionnaire items for the awareness of FPI elements, all elements received high ratings of more than 4 points, showing that the course was designed to allow students to experience the FPI elements. The highest score is integration element (mean = 4.53, SD = 0.27), followed by activation element (mean = 4.38, SD = 0.32), demonstration element (mean = 4.33, SD = 0.58), and application element (mean = 4.33, SD = 0.27), and the lowest score is for problem-centered element (mean = 4.09, SD = 0.55).

Notably, students did not show much variation in their awareness of the elements in Table 8. However, as described in the course design section, the length of the course’s application element was increased to provide students with more active learning opportunities. The application element accounted for 33% of the course time, while the integration element accounted for 22% of the course time. While the integration element received the highest score, the application element scored relatively low.

Borg (2006) points out that effective language teaching requires rich classroom diversity and an environment where students can increase their participation and the quality of their involvement. This course was designed to ensure that students participated in

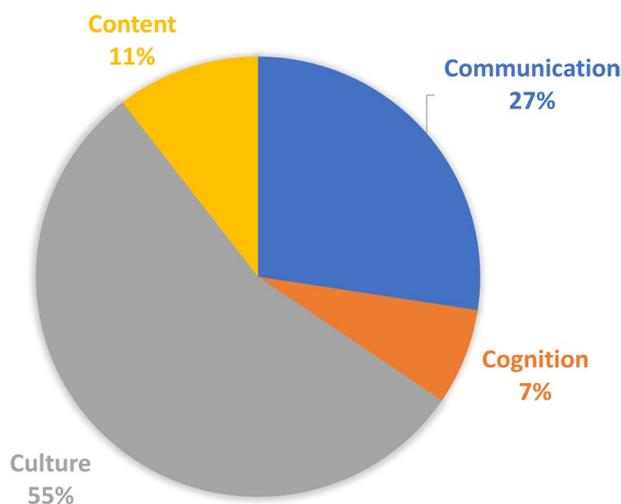


Fig. 4 Proportion of statements related to the 4Cs in group conversations

Table 8 Descriptive statistics of students’ awareness of the FPI elements (N = 16)

	Min	Max	Mean	SD
Problem-Centered	3.25	5.00	4.09	0.55
Activation	3.80	4.80	4.38	0.32
Demonstration	3.00	5.00	4.33	0.58
Application	3.75	4.75	4.33	0.27
Integration	3.80	4.80	4.53	0.27

the activities, and the dialogue content during group discussions was analyzed to capture the quality of learning during participation time. The conversations related to the course content and Japanese language learning in the activity were coded. Of the 983 dialogues, 59 were related to lesson content and 155 were related to Japanese language learning.

The students' conversations were coded into categories according to CLIL's definition of the 4Cs. The coding criteria were edited by first author in CLIL based on CLIL's 4C's principles. Then, the coding criteria were discussed by experts in Japanese language education and experts in ID (co-authors). After the coding criteria were unified, the students' conversations were coded. In addition to the statements related to content and Japanese language learning, there were also statements about cognition and culture. Content was defined as explanations and discussions about what was being learned. For example, in Activity 2, students used a worksheet to summarize their changing impressions of Japanese people, so words related to impressions were judged to be related to learning content in this activity. The definition of communication is that the target language is used, or that discussion of the words and grammar of the target language. The definition of cognition is the ability to summarize what has been learned in one's own words. Or students can give their own opinion based on what they have learned by thinking about themselves. So statements about expressing one's own ideas are considered to be related to cognition. Culture is defined as collaborative work between groups or thinking about internationalization. Therefore, statements that help promote group collaboration, or statements about cultural exchange, are considered as culture in this study.

In the 2 group activities, there were 983 conversations, of which 565 were related to 4Cs. 155 were related to communication. 40 were related to cognition. 311 were related to culture. 59 were related to content (Fig. 4).

Example 2. Group discussion about contents and Japanese.

Group 1

G1-S1: I wrote “同じ”(same). I think it's all the same.

G1-S4: The teacher is asking you when.

G1-S3: *You see, what the teacher's saying is originally you thought the Japanese were like this, and then it changed.*

G1-S4: Yeah, that means when.

G1-S2: *どんなときですか (When is it?)*

G1-S1: No change. It's the same as I thought it would be.

G1-S3: No, the teacher asked when you had changed.

G1-S1: That is, the teacher was asking when it had changed, but I don't think it has.

G1-S4: That is when it has or has not changed.

G1-S1: It hasn't changed.

G1-S4: You just write about one thing that inspired you.

G1-S1: *I am now writing about “同じ”(same).*

(Writing and painting)

Group 1 had 443 conversations in the two learning activities, including 19 conversations that included knowledge content and 30 conversations about Japanese. Moreover, as demonstrated in this dialogue, most of the Japanese used consisted of single

words, and even when the sentences were in Japanese, they repeated the topics of the questions presented. These conversations corroborate the results in Table 2. Students in Group 1 mostly use Japanese to confirm basic words, which can explain that the scores on the basic Japanese test have improved.

Group 2

Scene 1

G2-S2: What about this one?

G2-S3: Let's write another one

G2-S1: *The third is to ask, at first you think so, and then after contact with them, what has changed, right?*

G2-S2: *Yeah. Just think of them for “真面目” (Seriously).*

G2-S1: *どんなときで` (At what time). At what time was it? OK, then next.*

G2-S2: Can this space be written down?

(5 seconds)

G2-S2: Just write keywords?

G2-S1: Yeah.

G2-S3: Right.

Scene 2

G2-S4: Eh, “見た”, is there a small “つ”?

G2-S3: No, there is no “つ.”

Group 2 had a total of 582 conversations during the two learning activities. Of these, 26 conversations were related to the knowledge content, and 77 conversations were related to Japanese. As shown in the example given for Group 2, as in Group 1, most of the conversations in Japanese consisted of sentences confirming the question topic, and most of the other conversations in Japanese were in words. However, there were many instances of confirmation and discussion of Japanese words in Group 2. These conversations corroborate the results in Tables 2 and 5. Although the students did not have the majority of conversations about knowledge content and Japanese in the group discussions during the application time, it was meaningful to confirm and consolidate the words learned during the group discussions.

Group 3

G3-S4: In fact, I would have liked to write about “曖昧”(ambiguous).

G3-S3: My theme is tenderness. 雲のような日本人(Cloudy Japanese). “雲” means cloud.

G3-S1: This is OK.

G3-S2: *What about the specific things? Like they do something that makes you think they are like clouds.*

G3-S3: *That is, the characteristic is 優しい (gentle). Then they are polite. Specifically, the first impression was obtained from the Japanese drama. Then actual contact with the Japanese, the impression has not changed. なし(none).*

G3-S2: OK.

Group 3 had a total of 347 conversations in the two learning activities. Eight of these were about knowledge content and 34 about Japanese. As with the conversations in Groups 1 and 2, most of the Japanese that appeared consisted of single words

rather than complete sentences. This is consistent with the results in Table 2 that students' scores on the basic Japanese test improved significantly.

Group 4

Scene 1

G4-S1: 最近読んだ本は何ですか?(What books have you read recently?)

G4-S2: あのう、これです。< 倉央嘉措诗传全集 > (Um, here it is. *The Complete Poetic Biography of Kurama Gyatso*.)

G4-S1: この本はどうだと思えますか?(What do you think of this book?)

G4-S2: いいです。でも、ちょっと難しい。(It's good. But a little difficult.)

Scene 2

G4-S2: *What is your title?*

G4-S1: *My title is Japanese people.*

(Writing and painting)

G4-S3: What are the characteristics of Japanese people?

(Writing and painting)

G4-S2: *What's your title?*

G4-S3: *I wrote "Japan in Green."*

G4-S2: Oh, OK. Pretty good.

Group 4 had a total of 195 conversations over the two learning activities. There were 6 conversations about content and 14 conversations about Japanese. Although the overall amount of conversations about content and Japanese in Group 4 was less than in the other three groups, members of Group 4 would try to have conversations in Japanese. In this CLIL class, considering the students' Japanese language level, they were not required to use Japanese to communicate in the group discussions. However, the students in Group 4 would still try to communicate in Japanese independently. This could explain the results in Table 2 that the students' Japanese language scores improved in a meaningful way.

From the conversations in the four groups, it can be seen that the students did not spend all their time discussing the knowledge content and language learning during the application time. Some of the conversations were about how to complete and confirm information about the worksheet. Although the results in Tables 2 and 3 show a significant improvement in Japanese language scores and course content scores, it is also clear from the group discussions that students spent limited time on course content discussions and the use of Japanese. Furthermore, the use of Japanese during group discussions was mostly focused on the confirmation of the problem topics and Japanese words. Only the members of Group 4 tried to communicate entirely in Japanese during the discussion. Therefore, in future CLIL design, attention should also be paid to supporting group discussions. Make group members more active and interactive. Allow students to discuss the course content more.

Research question 3: what elements of FPI have a positive impact on CLIL courses?

In order to determine factors that might affect the cultivation of learning outcomes through CLIL, Spearman's Rank Correlation Coefficient was used to analyze the correlations among learning outcomes (pre-post basic test score, pre-post content test score,

pre-post writing test score), 5 principles of FPI (post FPI Questionnaire). The calculation is reproduced below. First, the pre-post difference in basic Japanese proficiency test score, content knowledge test score, writing skills test score, were calculated. Then, the sum of the score of the questionnaires under the each of the 5 principles in the FPI questionnaire was calculated. Finally, because of the small sample size, the Spearman's Rank correlation was used to calculate the correlation between the difference between the score of basic, content, and writing and the total score for each of the 5 principles.

The results in Table 9 show that there is a significant correlation between the problem-centered principle and the integration principle with all learning outcomes; there is a significant correlation between the activation principle and the demonstration principle only with the learning outcomes of basic knowledge; and there is a significant correlation between the application principle with the learning outcomes of content knowledge and writing.

According to the correlation results in Table 9, some elements of FPI awareness reached a significant level of correlation with learning outcomes.

Correlations among the variables are presented in Table 9. The problem-centered element has a moderate positive correlation with the basic Japanese proficiency test score (0.554**, $p < 0.01$), a strong positive correlation with the content test score (0.602*, $p < 0.05$), and a strong positive correlation with the writing test score (0.813**, $p < 0.01$). As described in the course design phase, this course prepared a life-related problem for each class and had student peers learn the course content to solve the problem. The significant positive correlation that emerged between the problem-centered elements and the learning outcomes in this course is consistent with the findings noted by Gardner (2011).

The activation element has a strong positive correlation between basic Japanese proficiency test scores (0.605*, $p < 0.05$) and no significant correlation between content test scores and writing test scores. It is presumed that there is no obvious storyline that appears in each lesson throughout the course, which may be the cause of this result. For example, the problem in the first lesson was "How do you and your friends self-explain to each other?" while the problem in the second lesson was, "What is your impression of Japanese people in real life, and how can you make better friends with foreigners?" Such a question setting may cause students to have a weak sense of the connection between the lessons, leaving them unable to clearly appreciate the connection between what they learned in the previous and the current lessons. Merrill (2020) states that courses should be designed so that students solve problems relevant to real life and have strong connections between problems to evoke better what students have already learned and

Table 9 Spearman's rank correlation coefficient between FPI, Writing Test, and Basic Japanese Proficiency Test Scores

	Problem-Centered	Activation	Demonstration	Application	Integration
Basic	.554*	.605*	.520*	.387	.641**
Content	.602*	.278	.303	.761**	.581*
Writing	.813**	.131	.193	.675**	.624**

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

what they have experienced. The problem in the second lesson should be changed to read, “Looking back at the results of yourself-explain, what can you do to become better friends with Japanese students?” Tying the problems of each lesson into a complete storyline might better mobilize what students have already learned.

The demonstration element has a moderate positive correlation with the basic Japanese proficiency test scores (0.520^* , $p < 0.05$) but no significant correlation with the content or writing test scores. Although Merrill (2002) demonstrates that examples of FPI (use of media, worksheets, guidance, etc.) can make more effective use of the lesson by rationally utilizing these elements that students are used to using in other scenarios, the task of effectively integrating multimedia with content and activities is still challenging. Lo et al. (2018) point out that students can only learn relatively simple knowledge by watching multimedia videos, and they still need guidance and help from teachers when learning complex knowledge. At present, using multimedia in class has become a common trend, but the efficient use of multimedia is still a rather urgent problem to be discussed and solved.

The application element has a strong positive correlation with the content test scores (0.761^{**} , $p < 0.01$) and a strong positive correlation with the writing test scores (0.675^{**} , $p < 0.01$), but no significant correlation with the basic Japanese proficiency test scores. When designing the course, there was expected to be a strong relationship between the FPI application element and the learning outcomes. As a result, although there were robust correlations of the application element with content knowledge and writing skills, there was no significant correlation with basic Japanese proficiency. Although the teacher used Japanese when teaching the content, when the students presented their results, they also used Japanese. The analysis of the conversations in the student group discussions showed that the students used their native language most of the time. When they were in a group discussion, they only used Japanese to confirm the topic problems and write the speeches: During the group discussion activities, a total of 1567 conversations took place, but of these, only 155 (9.89%) were in Japanese.

Example 3. Group discussion about the Japanese language.

Group 1

G1-S3: The formation of impressions ...

G1-S4: *Formative from ドラマ (Drama).*

G1-S1: I'm better at writing that way.

(Writing and painting)

Group 1 had a total of 443 conversations in the two learning activities. There were 30 (6.77%) conversations that included Japanese. As this example shows, in the second learning activity, out of 124 conversations in the group discussion, only 5 appeared Japanese, and in all cases, they were only single Japanese words. This explains the results in Table 6 that the application elements only showed significant correlations with the scores on the Japanese writing test and no significant correlations with the scores on the basic Japanese test, for students spent most of their time in group discussions conceptualizing the composition of their writing rather than studying Japanese.

Group 2

G2-S2: Eh, *how should 厳しい (strict) be written?*

G2-S3: seriousness, 厳肅 (solemnly)

G2-S1: *You can't write 厳肅(solemnly), 厳肅(solemnly) means the atmosphere is serious.*

G2-S2: *OK, then write 厳しい(strict)*

20 seconds later

G2-S3: *I'll write another one, はじめに(firstly), this one can be written in third.*

G2-S2: *厳しい(strict).*

5 seconds later

G2-S3: *OK, what else?*

G2-S4: *根性(guts)*

G2-S3: *根性強い(Strong guts). It's okay, right?*

G2-S4: *Yeah.*

Group 2 had 583 conversations in the two learning activities, including 77 conversations in Japanese (13.21%). Just as in Group 1, only Japanese words appeared in most of these conversations. Nevertheless, as the example shows, a characteristic of Group 2 is that the students thought about the deeper meanings of words and discussed the appropriateness of the words used. This explains the results presented in Table 6: While no significant correlation was shown between the application element and the basic Japanese test scores, a significantly strong correlation appeared between the application element and the writing score. When Group 2 students discussed their assignment, they did not deal with the basic words taught but considered what words they could use to better express their ideas in writing.

Group 3

G3-S2: *Hey, Japanese people who love to give gifts, in Japanese it should be that “プレゼントを上げるのが好きな日本人”*

G3-S3: *Japanese people who like to give gifts, right? プレゼントを贈ることが好き. Like to send gifts, right?*

G3-S2: *Right, right.*

(Writing and painting)

G3-S4: *What is this theme of yours?*

G3-S2: *My theme is Japanese people who like to give gifts.*

G3-S3: *What about specific? Specific.*

G3-S2: *Do I say it in Chinese?*

G3-S3: *OK*

G3-S2: *It was that at first, I just thought that Japanese people were giving to Japanese people, but then I found out that they invited people from other countries too, and then my impression of them changed, and I thought they were warm. Then every time we met, they always sent me a gift.*

Group 3 had 347 conversations in the two learning activities, of which 34 included Japanese (9.80%). Again, however, most of the dialogues only contained single Japanese words. However, it is worth noting that, as the example shows, the members of Group 3 corrected each other's incorrect grammar usage, which involved a usage that was not taught in this course. This explains the significantly stronger correlation in Table 6 between the integration elements and the basic Japanese scores. Such grammar correction is not only effective for this instance of learning but will lead students to think about the situation that was corrected this time and then use grammar correctly when using Japanese in the future.

Group 4

G4-S2: This write the 環境守る (Protecting the environment) or write the 方法 (method).

G4-S3: En, 方法 (method).

(10 seconds)

G4-S2: 保護 (Protection)

G4-S3: Yes, the pronunciation is ぽご (Protection).

(Writing and painting)

Group 4 had 195 conversations in the two learning activities, including 14 conversations in Japanese (7.18%). As in the other three groups, most of the Japanese that appeared in the conversations in Group 4 were words. This is consistent with the results shown in Table 9. Although the usage and pronunciation of words were discussed during the activities, overall, the use of Japanese was the least frequent in Group 4.

As the group discussion examples show, the students used their first language most of the time. Moreover, almost all of the Japanese appeared as single words, not as fully expressed sentences. Therefore, students actually used Japanese less than other FPI elements under the application element during class time. Macaro et al. (2020) point out that when students learn complex knowledge, using their foreign language for discussion may cause a heavy cognitive burden. Therefore, this course is designed to allow students to use their native language in their discussions. However, Lo (2015) also points out that excessive use of a native language may pose an obstacle to learners' target language learning. Thus, in future course design, the use of native and foreign languages should be appropriately balanced. In addition, when using FPI for CLIL design, it is necessary not only to ensure that students have enough time for the learning activities but also to improve the quality of the activities.

The integration element has a strong positive correlation with the basic Japanese proficiency test score (0.641^{**} , $p < 0.01$), a moderate positive correlation with the content test score (0.581^* , $p < 0.05$), and a strong positive correlation with the writing test score (0.624^{**} , $p < 0.01$). The correlation results are consistent with the questionnaire results about FPI awareness. Students had a high awareness of the integration elements in the course, and this awareness had a positive impact on students' basic Japanese language knowledge, writing skills, and content knowledge. Merrill (2013) points out that the integration element includes reflection, discussion, and creation. As described in the course design, this course gave students the opportunity to reflect on themselves through group discussions, present what they had learned, and think about how they could creatively apply what they had learned in their future lives. The results show that all these designs have had a positive impact on students' learning outcomes.

Conclusion and implementation

This study applied FPI elements to design a CLIL course to improve the Japanese language knowledge, intercultural communication content knowledge, and writing skills of 16 university students.

First, the test includes basic Japanese proficiency, intercultural communication content, and writing skills, which were administered to assess students' learning outcomes before and after the course. The results indicated a significant improvement in all three domains. However, some personal differences in writing emerged. The analysis of the

group conversations showed that the students showed some confusion in the use of the worksheets. Furthermore, conversations about how to improve writing skills were not addressed at all in the discussion. This indicates that instructions on using worksheets and guiding students to learn from each other during group discussions are essential.

Second, to clarify the FPI elements that affect learning outcomes, an FPI questionnaire was conducted to assess students' awareness of FPI, then the relationships between FPI element awareness and learning outcomes were assessed. The results indicate that all FPI elements had a positive impact on basic Japanese proficiency except for the application element; the problem-centered, application, and integration elements had a positive impact on intercultural communication content and writing skills. FPI as a problem-centered theory was used positively in the design of the CLIL course. In particular, there was a significant positive correlation between the problem-centered elements and the content and language of CLIL in this course. This confirms that the problem-centered course design approach is effective for CLIL, as mentioned in the hypothesis. However, it is also clear from the results that the application elements of FPI did not correlate significantly with the basic language test of CLIL. Therefore, care needs to be taken when designing future CLIL courses using problem-centered theory, and students should also be guided to apply the fundamentals in the activities. Therefore, when designing future CLIL courses using problem-centered theory, students need more guidance in applying the basic language knowledge in the activities.

However, there are some limitations to this study. One limitation is that the number of participants is relatively small and the course length is relatively short, and it has not been tested for many disciplines, which impairs the generalizability of the results. Another limitation is the data type for qualitative analysis. Although this study analyzed the students' group discussions, the students' learning behaviors could not be revealed from the conversations. A better understanding of students' learning behaviors in CLIL courses is needed in order to be able to improve the CLIL course design in a more targeted manner. Finally, because this was a practical course and the students were on a tight schedule, the students were not scheduled for a delayed test in this study. In future course design and experiment, consideration can be given to implementing delay test for learning content that can be affected by short term memory, such as learning content. To better verify the course effect.

Still, the findings of this study can provide some ideas for CLIL course design when teaching Japanese learners of the same level, indicating that the study's findings would have been more representative and convincing if a greater number of participants had been included, more time had been available, and more topics for the course had been used in the study. In addition, this CLIL course is not an exclusively target language learning environment. This course allows students to use their native language during group discussions. In future CLIL course, worksheets can be designed based on the principles of FPI to support CLIL that are entirely in the target language learning environment. As a result, such future investigation will contribute to CLIL course design. The methodology of this design can be applied to other topics in future CLIL courses to obtain more data for a more extensive validation of the feasibility of this course design. Furthermore, future research should also focus on the variability of students' individual opinions, thereby improving the quality of the design.

Appendix 1

Writing Evaluation Criteria (Japan Foundation, 2010).

Content:

4. The information necessary for the theme is explained to an extent so that the reader can understand.

3. There is a lack of explanation; the reader cannot understand without checking.

2. The reader can vaguely understand the subject, but it is difficult to understand what convey overall.

1. It is difficult to understand because the information necessary for the theme is not accurate.

Vocabulary/grammar:

4. The writer accurately uses words, expressions, and syntax related to the topic.

3. There are some vocabulary and grammatical errors, but the sentence can be understood.

2. There are some parts where the message that the writer wants to convey cannot be conveyed.

1. Since the information is written in pieces, it is difficult to understand what they convey.

Constitution: the composition of the text, including subparagraphs, connections between paragraphs, etc.

4. The reader can connect words and phrases using basic, parallel contact expressions.

3. There are some parts where the relationship between sentences is difficult to understand.

2. Only the points that the writer wants to convey are written side by side.

1. Writing sentences and words in pieces; no sentence composition.

Appendix 2

TALQ (Frick et al., 2009).

a. Authentic Problems Scale (Merrill, Principle 1): Cronbach's $\alpha = 0.81$.

- I performed a series of increasingly complex authentic tasks during this course.
- I solved authentic problems or completed authentic tasks during this course.
- I solved a variety of authentic problems that were organized from simple to complex in this course.
- The assignments, tasks, or problems I did in this course are clearly relevant to my professional goals or fields of work.

b. Activation Scale (Merrill, Principle 2): Cronbach's $\alpha = 0.91$.

- I engaged in experiences that subsequently helped me learn ideas or skills that were new and unfamiliar to me.
- In this course, I was able to recall, describe, or apply my past experiences to help me connect to what I was expected to learn.

- My instructor provided a learning structure that helped me mentally organize new knowledge.
 - I was able to connect my past experiences with the new ideas and skills that I learned in this course.
 - I was not able to draw upon my past experience, nor was I able to relate to the new things I was learning in this course. (–)
- c. Demonstration Scale (Merrill, Principle 3): Cronbach's $\alpha = 0.88$.
- My instructor demonstrated the skills I was expected to learn during this course.
 - My instructor gave examples and counter-examples of concepts that I was expected to learn.
 - My instructor did not demonstrate the skills I was expected to learn. (–)
 - My instructor provided alternative ways of understanding the same ideas or skills.
- d. Application Scale (Merrill, Principle 4): Cronbach's $\alpha = 0.74$.
- My instructor detected and corrected the errors I was making while solving problems, doing learning tasks, or completing assignments.
 - My instructor gradually reduced coaching or feedback as my learning or performance improved throughout this course.
 - I had opportunities to practice or try out what I learned during this course.
 - The course instructor gave me personal feedback or appropriate coaching on what I was trying to learn.
- e. Integration Scale (Merrill, Principle 5): Cronbach's $\alpha = 0.81$.
- I had opportunities to explore how I could personally use what I had learned in this course.
 - I see how I can apply what I have learned in this course to real-life situations.
 - I was able to publicly demonstrate what I learned in this course.
 - I was able to reflect, discuss, and defend what I learned in this course.
 - I don't expect to apply what I learned in this course to my profession or field of work. (–)

Items with (–) are negatively worded, and the rating scores are reversed for their analysis.

Abbreviations

CLIL	Content and Language Integrated Learning
FPI	First Principles of Instruction
FL	Foreign Language Learning
ID	Instructional Design
4Cs	The four elements of Content and Language Integrated Learning

Acknowledgements

Dr. Yamada received the research grants for this and other research from Japan Society for the Promotion of Science (JSPS). Prof. Susono received the research grant for other research from JSPS.

Authors' contributions

HH designed the course and performed the data analyses and wrote the main manuscript text. HH and HS performed the practice. MY contributed to write and consider the conception of the study and supervised this research overall. XG and LC contributed to consider the conception of the study. All authors read and approved the final manuscript.

Funding

This research was supported by JSPS KAKENHI [Grant Nos. JP22H00552, JP21K18134, JP21KK0184].

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author on reasonable request.

Declarations**Ethics approval and consent to participate**

Not applicable.

Competing interests

Not applicable.

Received: 20 April 2022 Accepted: 8 September 2022

Published online: 09 January 2023

References

- Agustín-Llach, M. P. (2016). Age and type of instruction (CLIL vs. traditional EFL) in lexical development. *International Journal of English Studies*, 16(1), 75–96.
- Allison, D., Cooley, L., Lewkowicz, J., & Nunan, D. (1998). Dissertation writing in action: The development of a dissertation writing support program for ESL graduate research students. *English for Specific Purposes*, 17(2), 199–217.
- Ball, P., & Lindsay, D. (2010). Teacher training for CLIL in the Basque Country: The case of the Ikastolas—An expediency model. In Y.-L.T. Ting (Ed.), *CLIL in Spain: Implementation, results and teacher training* (pp. 162–187). Cambridge Scholars Publishing.
- Borg, S. (2006). The distinctive characteristics of foreign language teachers. *Language Teaching Research*, 10(1), 3–31.
- Cenoz, J., Genesee, F., & Gorter, D. (2013). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics*, 35(3), 243–262. <https://doi.org/10.1093/applin/amt011>
- Coyle, D. (2006). Content and language integrated learning—Motivating learners and teachers. *Scottish Language Review*, 13, 1–18.
- Coyle, D. (2007). Content and language integrated learning: Towards a connected research agenda for CLIL pedagogies. *International Journal of Bilingual Education and Bilingualism*, 10(5), 543–562.
- Coyle, D., Hood, P., & Marsh, D. (2010). *Content and language integrated learning*. Ernst Klett Sprachen.
- Crandall, J. (1999). Content-based instruction (CBI). In B. Spolsky (Ed.), *Concise encyclopedia of educational linguistics*. Cambridge University Press.
- Dale, L., & Tanner, R. (2012). *CLIL activities with CD-ROM: A resource for subject and language teachers*. Cambridge University Press.
- Dourda, K., Bratitsis, T., Griva, E., & Papadopoulou, P. (2014). Content and language integrated learning through an online game in primary school: A case study. *Electronic Journal of e-Learning*, 12(3), 243–258.
- Ennis, M. J. (2015). "Do we need to know that for the exam?" Teaching English on the CLIL fault line at a trilingual university. *TESOL Journal*, 6(2), 358–381.
- Evnitskaya, N. (2014). "Do you know Actimel?" The adaptive nature of dialogic teacher-led discussions in the CLIL science classroom: A case study. *The Language Learning Journal*, 42(2), 165–180.
- Filice, S. (2020). CLIL in pharmacology: Enabling student voice. *Latin American Journal of Content & Language Integrated Learning*, 13(2), 313–338.
- Frick, T. W., Chadha, R., Watson, C., Wang, Y., & Green, P. (2009). College student perceptions of teaching and learning quality. *Educational Technology Research and Development*, 57(5), 705–720.
- Gardner, J., & Jeon, T. K. (2009). Creating task-centered instruction for web-based instruction: Obstacles and solutions. *Journal of Educational Technology Systems*, 38(1), 21–34.
- Gardner, J. (2011). How award-winning professors in higher education use Merrill's first principles of instruction. *International Journal of Instructional Technology and Distance Learning*, 8(5), 3–16.
- Hao, H., & Yamada, M. (2021). Review of research on content and language integrated learning classes from the perspective of the first principles of instruction. *Information and Technology in Education and Learning*, 1(1), Rvw-p001.
- Heo, Y. (2006). Content-based instruction. *Hawaii Pacific University TESOL Working Paper Series*, 4(2), 25–31.
- Hernandez, H. P. (2016). OBE EAP-EOP model: A proposed instructional design in English for specific purposes. *Journal on English Language Teaching*, 6(4), 1–12.
- Honebein, P. C. (2019). Exploring the galaxy question: The influence of situation and first principles on designers' judgments about useful instructional methods. *Educational Technology Research and Development*, 67(3), 665–689.
- Hüttner, J., & Rieder-Bünemann, A. (2010). A cross-sectional analysis of oral narratives by children with CLIL and non-CLIL instruction. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 61–80). John Benjamins.
- Japan Foundation. (2010). *JF standard for Japanese-language education 2010* (2nd ed.). Japan Foundation.

- Japan Foundation & Japan Educational Exchanges and Services. (2012). *Japanese-language proficiency test*. Bonjinsha.
- Kanamura, K., & Miyajima, R. (2016). Program design at the Center for Japanese Law Education and Research (CJL): Interpretation with integrated language learning (CLIL): From mission policy, curriculum to community building. *Nagoya University Asian Law Bulletin*, 2, 2–24.
- Langé, G. (2007). Postscript to CLIL 2006 and future action. In D. Marsh & D. Wolff (Eds.), *Diverse contexts, converging goals: CLIL in Europe* (pp. 350–358). Peter Lang.
- Leal, J. P. (2016). Assessment in CLIL: Test development at content and language for teaching natural science in English as a foreign language. *Latin American Journal of Content and Language Integrated Learning*, 9(2), 293–317.
- Lo, C. K., Lie, C. W., & Hew, K. F. (2018). Applying “First Principles of Instruction” as a design theory of the flipped classroom: Findings from a collective study of four secondary school subjects. *Computers & Education*, 118, 150–165.
- Lo, Y. Y. (2015). How much L1 is too much? Teachers' language use in response to students' abilities and classroom interaction in content and language integrated learning. *International Journal of Bilingual Education and Bilingualism*, 18(3), 270–288.
- Macaro, E., Tian, L., & Chu, L. (2020). First and second language use in English medium instruction contexts. *Language Teaching Research*, 24(3), 382–402.
- McGee, P., & Reis, A. (2012). Blended course design: A synthesis of best practices. *Journal of Asynchronous Learning Networks*, 16(4), 7–22.
- Mehisto, P. (2008). CLIL counterweights: Recognising and decreasing disjuncture in CLIL. *International CLIL Research Journal*, 1(1), 93–119.
- Mendenhall, A. M. (2012). Examining the use of first principles of instruction by instructional designers in a short-term, high volume, rapid production of online K–12 teacher professional development modules. Doctoral dissertation, Florida State University.
- Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, 50(3), 43–59.
- Merrill, M. D. (2013). *First principles of instruction: Identifying and designing effective, efficient and engaging instruction*. Pfeiffer.
- Merrill, M. D. (2020). A syllabus review check-list to promote problem-centered instruction. *TechTrends*, 64(1), 105–123.
- Meyer, O. (2010). Introducing the CLIL-pyramid: Key strategies and principles for quality CLIL planning and teaching. In M. Eisenmann & T. Summer (Eds.), *Basic issues in EFL-teaching and learning* (pp. 11–29). Universitätsverlag Winter.
- Navés, T., & Victori, M. (2010). CLIL in Catalonia: An overview of research studies. In Y.-L.T. Ting (Ed.), *CLIL in Spain: Implementation, results and teacher training* (pp. 30–54). Cambridge Scholars Publishing.
- O'Dwyer, F., & de Boer, M. (2015). Approaches to assessment in CLIL classrooms: Two case studies. *Language Learning in Higher Education*, 5(2), 397–421.
- Ouazizi, K. (2016). The effects of CLIL education on the subject matter (Mathematics) and the target language (English). *Latin American Journal of Content and Language Integrated Learning*, 9(1), 110–137.
- Pérez-Vidal, C., & Juan-Garau, M. (2010). To CLIL or not to CLIL? From bilingualism to multilingualism in Catalan/Spanish communities in Spain. In Y.-L.T. Ting (Ed.), *CLIL in Spain: Implementation, results and teacher training* (pp. 115–138). Cambridge Scholars Publishing.
- Pladevall, R. C., Montserrat, I. N. S., & Evnitskaya, N. (2011). Rethink, rewrite, remake or learning to teach science through English. *AICLE-CLIL-EMILE: Educació plurilingüe. Experiències, research & polítiques*, 2, 167–177.
- Reigeluth, C. M. (1999). What is instructional-design theory and how is it changing. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (Vol. 2, pp. 5–29). Lawrence Erlbaum Associates.
- Reiser, R. A. (2001). A history of instructional design and technology: Part II: A history of instructional design. *Educational Technology Research and Development*, 49(2), 57–67.
- Richards, J. C., & Lockhart, C. (1994). *Reflective teaching in second language classrooms*. Cambridge University Press.
- Richards, J., & Rodgers, T. (2001). *Approaches and methods in language teaching*. Cambridge University Press.
- Sabet, M. K., & Sadeh, N. (2012). CLIL European-led projects and their implications for Iranian EFL context. *English Language Teaching*, 5(9), 88–94.
- Seels, B. B., & Richey, R. C. (1994). *Instructional technology: The definition and domains of the field*. Association for Educational Communications and Technology.
- Tu, W., & Snyder, M. M. (2017). Developing conceptual understanding in a statistics course: Merrill's first principles and real data at work. *Educational Technology Research and Development*, 65(3), 579–595.
- van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Educational Technology Publications.
- Yufrizal, H., & Huzairin. (2017). Project-based Content Language Integrated Learning (CLIL) at Mathematics Department Universitas Lampung. *English Language Teaching (Toronto)*, 10(9), 131–139.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.