Content-Based Instruction (CBI) for Japanese High School and College English Classrooms: Students' Writing Performances on CBI-Related Tasks, Task Perceptions and Course Perceptions

A dissertation submitted in partial fulfilment of the requirements

for the degree of Doctor of Education

in the Graduate School of Education

Noriko Suzuki

Waseda University

2023

ACKNOWLEDGEMENTS

ABSTRACT

With a view to creating meaningful language use opportunities (Cammarata, 2016), the implementation of Content-Based Instruction (CBI) and Content and Language Integrated Learning (CLIL), especially language-focused theme-based instruction, has been conducted in language classrooms in Japan (Ikeda, 2013; Kashiwagi & Kobayashi, 2019). Although its application has been increasing and so have practical report publications, empirical research which examines the effect of CBI/CLIL on the learners' language performance and development has been scarcely conducted in Japanese contexts, thus needing further exploration. Moreover, to understand how the learners' second language (L2) output skills, which is writing in this dissertation, is affected by the inclusion of content learning to language learning, it is fruitful to refer to the findings in the related research fields, such as research in TBLT (García Mayo, 2015; Moore & Lorenzo, 2015) and second language writing research (Manchón, 2020). In other words, learning in CBI/CLIL can be explored by a unit of task, and the interdisciplinary investigation of CBI/CLIL, TBLT, and second language writing is necessary.

Based on the above discussion, two CBI courses were created in this dissertation, and the effect of two writing task variables dealt with in these two courses (i.e., topic familiarity and integrated writing) on Japanese high school and college students' writing were evaluated to consider the possible impact of CBI learning on learner language performance and development. The comparison of writing performances on different task variables was conducted in terms of complexity, accuracy, and fluency (CAF) and rating. Simultaneously, students' perceptions of these classes were assessed through the analysis of reflection sheets and interviews. The overall research question for this dissertation was as follows: *How does theme-based instruction affect English as a Foreign Language (EFL) students' writing performance and development and their class perceptions?*

This dissertation consists of five chapters. Chapter 1 describes the purpose of this study, its background, and its academic contributions. In Chapter 2, literature of three related research fields, that is, CBI/CLIL, TBLT, and second language writing, was reviewed to understand students' writing performance and development in the CBI/CLIL context. The review starts with a description of the historical emergence of CBI and CLIL (Brinton & Snow, 2017; Coyle et al., 2010), theoretical underpinnings of these approaches for language learning from the interaction approach (Gass & Mackey, 2015), and cognitive, affective, and social theories (Fitzsimmons-Doolan et al., 2017). Then, the major course and lesson planning frameworks referred to in this dissertation [Mohan's (1986) Knowledge Framework, the SIOP model (Echevarría et al., 2017), the six Ts model (Stoller & Grabe, 2017) and Lyster (2011)'s counter balanced approach] are introduced. Next, the review of CBI/CLIL studies conducted in Japan revealed that practical reports are the dominant research type, and the number of empirical studies investigating language performance and development is limited. Extant research in Japan shows CBI/CLIL's positive impact on vocabulary learning (Goya, 2018; Yamano, 2013), grammar learning (Kashiwagi & Kobayashi, 2019), and writing (Ikeda, 2013; Suzuki, 2022), presumably due to repeated contact with the target language items learned in a contextualized manner in the CBI/CLIL.

The focus was narrowed down to written language development, which is a crucial skill not only for content learning in CBI/CLIL but also as a tool for language learning. Writing analysis in terms of CAF and rating in EFL contexts shows CBI/CLIL's favorable effects, especially on vocabulary, accuracy, and rating scores though there are some inconsistencies (Pérez-Vidal & Roquet, 2015; Roquet & Pérez-Vidal, 2017). Studies conducted in Japan also demonstrate primary evidence of writing development (vocabulary, fluency, and rating, Ikeda, 2013; Shibata, 2021). To further examine the instructional characteristics of CBI/CLIL for language performance and development, a close examination of writing task topics and

types related to the content learning (writing topic familiarity, integrated writing) is valuable in addition to using writing tasks unrelated to the content dealt with in CBI/CLIL classrooms, which is typically the case in CLIL studies. Since the content is newly learned in CBI/CLIL, topic familiarity is expected to be low, but it may increase after learning about the content. Writing with reference to a source reading is also a prominent task type in CBI/CLIL (Plakans, 2015).

To review how these task variables affect writing, empirical studies in the field of TBLT and second language writing were referred to. Studies that compare different levels of topic familiarity generally suggest that having or gaining topical knowledge has positive effects on subsequent writing for both linguistic (i.e., CAF, Abrams, 2019) and holistic (i.e., rating, Vandommele et al., 2017) aspects, but there seems to be only a small effect on accuracy. Therefore, it is assumed that writing for familiar topics and writing after gaining topical knowledge in CBI/CLIL may elicit higher performances from the learners. Research findings on the other task variable, which is integrated writing, show that it elicits higher lexical complexity than independent writing, and mixed results are found for syntactic complexity and ratings (Cumming et al., 2005; Frear & Bitchener, 2015). Qualitative investigation of processes and perceptions of these two tasks highlight their different sources of cognitive complexity and language learning opportunities. Despite these findings in writing task research, they have not been incorporated into CBI/CLIL studies to interpret the writing outcomes.

Based on the above discussion, Chapter 3 presents the first study. In this study, one of the CBI-related writing task variables, namely the content familiarity variable, was evaluated by comparing 36 Japanese secondary school students' writing performances on the content-specific power generation topic and the everyday, general topic. A theme-based instruction unit based on the power generation topic was created for the study by referring to several

existing CBI lesson planning models (Echevarría et al., 2017; Mohan, 1986), and writing performance before and after the instruction was compared by repeated measures Analysis of Variance (ANOVA). Findings indicated the existence of topic effects. For instance, the content-specific topic writing, which was rated as being more challenging by the learners, gained higher scores both in the pre- and post-tests in terms of syntactic complexity (mean length of T-unit), fluency, task requirements, and coherence/cohesion. In contrast, the general topic writing's lexical diversity was higher both in the pre- and post-tests, and accuracy was higher in the pre-test. These results suggest that using a writing prompt that requires content knowledge and is perceived to have a certain level of difficulty may elicit higher performance for some linguistic measures and rating. As for the impact of the theme-based instruction, various aspects of writing changed for the two topics after the instruction; task requirements and accuracy were higher in the post-test for the content-specific topic whereas comprehensibility and coherence/cohesion were higher for the general topic. There were additional statistical differences and higher effect sizes for the topic difference variable than the instruction variable; thus, the influence of the instruction was less pronounced compared to the topic difference variable, possibly due to the limited amount of instruction time. However, it has been shown that not only content-specific topic writing but also general topic writing was positively affected by the theme-based instruction, which may lend support to the introduction of CBI/CLIL in general English class for Japanese high schools. Analysis of the reflection sheets revealed that students experienced both positive and negative feelings toward the classes. The positive comments were mainly about the target content and language items described with words, such as "understood, was able to ~, want to review ~," and the negative comments were related to their understanding level and language skills (especially vocabulary and listening) described with words, such as "difficult, was not able to ~, did not understand." It can be said that the students perceived the learning of content and the target

language forms (e.g., compare/contrast language) favorably, but simultaneously, they negatively reflected on the higher input language levels for content learning and their lack of L2 skills. Limitations of this study include the short instructional period with a limited number of participants and the operationalization of topic familiarity, which has also been problematized in previous studies (He & Shi, 2012; Kessler et al., 2022).

Chapter 4 presents the second study, which explored the other writing task variable (with or without the requirement of source integration) in a content-based writing instruction classroom conducted for 27 Japanese college students. More specifically, the with integration task required students to use certain information in a source text while the without integration task was also accompanied by a related source text, but they did not have to integrate its information. The effect of the task variables on the students' writing quality was investigated through linguistic analysis and rating, and their perceptions of the task differences were explored through interviews with 13 participants. The first quantitative analysis obtained by paired t-tests and Wilcoxon signed-rank tests revealed that the with source integration writing obtained significantly higher fluency and marginally higher lexical sophistication scores. There was no difference in terms of the syntactic complexity, accuracy, and rating by human raters. Thematic analysis of the students' task perception interviews highlighted the crucial role of the understanding of a reading material and the support that it can offer for writing and language learning for the with integration task. In contrast, for the without integration task, learners perceived both supportive and difficult aspects of this task type in terms of organization and content generation. As for the interview results of the students' CBI course perceptions, they suggest that the challenges and benefits of having content focus are the two dominant themes. More specifically, perceived difficulty arose from novel content learning, language used for content learning, and the combination of these two aspects. Conversely, students also experienced the positive effects of having content focus for their genuine

interest and value toward the content learning, and they mentioned its effects on their language learning and development. For some students, the challenges caused by the content and language integration gave them a hurdle to overcome, and when they overcame it, they felt a sense of accomplishment in the course. Limitations of this study were that the limited number of participants may have lowered the generalizability of the findings, and the students were allowed to use writing aids, such as a dictionary and translation tools, which may have affected their writing performance.

Chapter 5 presents a general discussion and the pedagogical implications and concludes the study. Results of the two studies are discussed in terms of the task variable differences and students' course perceptions. Regarding the task variable differences, comparisons of content-specific topic vs. general topic writing tasks and with vs. without source integration tasks were made. The content-specific topic gained higher performance in terms of the syntactic complexity, fluency, task requirements, and coherence/cohesion in both pre- and post-test occasions compared to the general topic. While the content-specific topic was rated as being more difficult, it did not necessarily elicit lower performances from the students. In fact, they tried to produce more written output with longer T-unit and meet the task requirements with higher coherence/cohesion. This result is contrary to those of TBLT and second language writing studies, which show higher writing performance for the familiar topic (Abdi Tabari et al., 2021; Yoon, 2017). This finding was probably due to the students' attempts to further explain the topic of power generation than the general topic (Yang & Kim, 2020), thereby suggesting that performance may not necessarily be lower even if the topic was found to be more challenging by the students. Thus, the current result may provide empirical support for the use of content-specific writing tasks in high school classrooms that are compatible with the new Courses of Study that encourages the use of intellectually enticing content. Moreover, different aspects of writing for both content-specific (accuracy

and task requirements) and general topic (comprehensibility, coherence/cohesion) changed after the current CBI unit, implying that the positive influence of CBI/CLIL may not only be limited to the content dealt with in the unit but also for writing with general topics. For the with and without integration task variable, the with integration task elicited higher fluency and marginally higher lexical sophistication, presumably supported by the integration of a reading material. From this study, teachers can predict which writing aspects may be affected by this task manipulation and how it is perceived by students. When they want the students to process a reading material in depth but wish to impose less cognitive burden for idea generation, the with integration writing can be selected whereas when they want the learners to write about the topic related to the reading material read in CBI/CLIL in a more independent manner, then the without source integration task can be chosen.

As for the students' course perceptions, the results in study 1 suggest that they perceived the learning of content and target language forms (e.g., compare/contrast language) favorably, but at the same time, they negatively reflected on the higher input language levels for content learning and their lack of L2 skills. Commonalities can be observed in the results of study 2 in which the students' course perceptions were more deeply explored in the interviews. There were two themes demonstrating the difficult and positive sides of the current course. The challenges were due to the learning of new content, higher L2 skills necessary for learning, and the combination of content and language learning difficulties. In contrast, positive effects were also noticed, such as a supportive role that content can play for L2 writing, value of learning content, and inducing a sense of achievement. Both studies point out the difficulty of L2 input and output caused by content learning and the positive feelings toward learning about the content itself. From the results, it can be summarized that the focus on content learning in Japanese language classrooms can bring about students' positive affect arising from a feeling of content value and language development as well as

negative affect caused by difficulty and inability in L2. Based on both the quantitative and qualitative results, for teachers to know what kind of performance and student perception a task can elicit will help their lesson planning and ultimately support the students' language learning.

The study was concluded by restating its value for the teachers by showing them students' different task performances elicited by CBI-related writing tasks and their perceptions about the facilitative and difficult aspects of this type of learning. In addition, this study tried to understand students' language performance in CBI/CLIL by referring to TBLT and second language writing research findings, which is a necessary direction in CBI/CLIL research (García Mayo, 2015). Future research in Japan should include the investigation of other language skills and a more in-depth qualitative investigation of the students' class-by-class affect change and focus on the teachers' pedagogical decisions in the CBI/CLIL course planning.

TABLE OF CONTENTS

Appendix A: Questionnaire about the Difficulty of the Two Prompts Used in Study 1

Appendix B: Writing Handout for Study 1

Appendix C: Content/Language Goals and Tasks for Each Lesson in Study 1

Appendix D: Example of Coding for the Reflective Comments in Study 1

Appendix E: Values of Kurtosis, Skewness and Shapiro-Wilk Test for Each Data Set in Study 1

Appendix F: Teaching Materials Used in the First Class in Study 2

Appendix G: The Example Lesson Planning Concept Map Based on the Six Ts (Grabe & Stoller, 2017)

Appendix H: Writing Handout Used in Study 2

Appendix I: Interview Sheet Used in Study 2

Appendix J: Values of Kurtosis, Skewness and Shapiro-Wilk Test for Each Data Set in Study 2

Appendix K: Themes for With/Without Integration Tasks with English Translation

Appendix L: Themes for Students' Perceived Learning Experiences in the Current CBI Course with English Translation

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

It is widely accepted that language learning is facilitated in a content-rich learning environment where meaningful input, output, and interaction occur (Krashen, 1981, 1994; Swain, 1985; Long, 1996) rather than one that focuses solely on conducting decontextualized language practices to manipulate the rules. To create such a learning environment not only for English as a Second Language (ESL) learners but also for English as a Foreign Language (EFL) learners, the integration of content and language learning, commonly represented by the term Content-Based Instruction (CBI) or Content and Language Integrated Learning (CLIL), is gaining momentum in the field of foreign language education (Cammarata, 2016). The key feature of CBI/CLIL is the dual focus on both content and language (Stoller, 2008) where language learning can be contextualized in meaningful content learning with rich language input and output opportunities. In the CBI and CLIL literature, research confirms that learning through sustained content benefits the learners' written output (Ikeda, 2013; Kong, 2015), spoken output (Lialikhova, 2021), pronunciation (Hanzawa, 2019), motivation (Lasagabaster, 2011), and quality of classroom interaction (Lo & Macaro, 2015).

To enrich language classrooms in Japan with meaningful language use opportunities, the new Courses of Study (MEXT, 2018) for Japanese high schools that have gone into effect in 2022 specifically mentions the kind of content to be dealt with in English class (familiar topics and socially-oriented topics) and implementation of skill integration (e.g., reading into writing, listening into speaking) to facilitate language development. Since these reforms can successfully be reflected in the framework of CBI/CLIL, several authorized textbooks, such as FLEX (Ikeda, 2022), have started to offer thought-provoking content learning along with knowledge utilization output tasks (e.g., essay writing and presentation). Furthermore, it is predicted that the reform for the high school courses of study may be likely to affect college

English courses and lead the students to focus on language skills development through content learning. Although increasing CBI/CLIL implementation in Japan has been reported (Ikeda, 2013; Izumi, 2019), few studies have examined its potential influence on the students' language performance and development (for exception, see Ikeda, 2013; Suzuki, 2022). Therefore, supportive roles that content can play in language learning need to be fully explored in studies conducted in EFL contexts, such as in Japan.

Although becoming widespread, CBI/CLIL is merely a framework and not a method with predetermined teaching methodology (Stoller, 2008). Thus, what constitutes each class are the tasks for content and language learning, and the students' language development is influenced by the tasks used in the CBI/CLIL classrooms (García Mayo, 2015; Moore & Lorenzo, 2015). Therefore, the interface between CBI/CLIL and task-based language teaching (TBLT) has started to be actively investigated (García Mayo, 2015; Lyster, 2017). In CBI/CLIL, students perform tasks related to certain content and gain knowledge of it; thus, it may be possible to research the effect of CBI/CLIL by focusing on the tasks that reflect content integration variables, such as topic familiarity and integrated writing, in reference to the TBLT and second language writing research findings. By referring to these studies that have the potential to explain the possible and specific learning outcomes in CBI/CLIL, the evidence for conducting this kind of instruction and effectiveness of the new Courses of Study may be obtained. In addition, these task variables need to be explored with the data obtained from not only laboratory settings but also classroom settings to gain ecological validity.

In terms of the language skills to be acquired in Japanese classrooms, the survey on high school students and English teachers (MEXT, 2018) constantly indicates the difficulty of developing and assessing the output skills, namely speaking and writing. Although these abilities are difficult to develop even for immersion students with access to ample input, EFL

students can still increase the chance to practice these capabilities in the process of content learning in CBI/CLIL classrooms, such as through discussions, writing reports, and making presentations. Within the two output modalities, it is indicated that the writing modality is suitable as a language learning tool for those who are still building on their lexico-grammatical knowledge and consolidating it (Gilabert et al., 2016) due to its slower pace of processing and its visibility and permanence of production (Kuiken & Vedder, 2012). However, the use of writing tasks in the classroom is still largely restricted to controlled tasks (e.g., gap filling, translation), and freer form of writing is performed less in Japan (Kobayakawa, 2011). An increasing number of second language writing research tries to explore the potential of writing for language learning (Manchón, 2020), and in a similar vein, the intersection of second language writing research and TBLT research has started to be analyzed (Manchón, 2020), both aiming to support language learning through the use of writing tasks. With the accumulated insights in these two research fields (second language writing and TBLT), students' language performance and development in CBI/CLIL through writing tasks in Japanese classroom settings can be investigated. How to conduct writing activities and use writing to evaluate CBI/CLIL learning needs further accumulation of empirical studies.

Based on the above-mentioned background and the researcher's experience and awareness of issues as a high school/college English teacher, this dissertation aims to examine the writing task variables related to CBI/CLIL, especially the effect of the topic (i.e., content-specific and general topic), task type (i.e., two types of integrated writing task), and students' task perceptions by drawing on previous research in three interrelated areas: TBLT, second language writing, and CBI/CLIL. Two CBI classes were developed and implemented by the researcher as a teacher for Japanese high school and college students, and their writing was analyzed in terms of linguistic measures (i.e., complexity, accuracy, and fluency) and

rating to evaluate the learners' language performance and possible development in the CBI classrooms. In addition, their perspectives on learning in the current research context were obtained through interviews and questionnaire surveys. The overall research question of this dissertation is as follows: *How does theme-based instruction affect EFL students' writing performance and development and their class perceptions?* This research question will be explored through two empirical classroom-based studies.

The potential contribution of this dissertation includes concrete examples of CBI lessons conducted with beginner to intermediate-level Japanese EFL students, empirical investigation on their written output, and theoretical investigation of writing from three different research fields, namely CBI/CLIL, TBLT, and second language writing. Content-based language teaching has become increasingly important, thus reflecting the recent revision of the Courses of Study (MEXT, 2018), which places great emphasis on meaningful L2 input and output. This thesis demonstrates the actual implementation of CBI lessons and investigated learners' writing performances, whose investigation is lacking in previous studies, especially in the EFL contexts. As previous research on EFL classroom instruction has predominantly focused on teaching language aspects, this study tried to focus on both content and language teaching and demonstrate the significant role of content on writing performance. Moreover, it tries to investigate the building blocks of CBI lessons, namely tasks, by referring to TBLT and second language writing research and present a detailed picture of student learning in the CBI context. It is hoped that this study will contribute to teachers' understanding of the effect of their choice of instruction and types of tasks on the students' learning and writing performance.

This dissertation consists of five chapters with two related studies following this introduction. In Chapter 2, to understand students' writing performance and development in the CBI context, the literature of three related research fields, namely CBI/CLIL, TBLT, and

second language writing is reviewed. The review starts with a description of the historical emergence of CBI and CLIL, the theoretical underpinnings of these approaches for language learning, and major course and lesson planning frameworks. Next, reflecting on the current situation where the implementation of CBI/CLIL has been increasing in Japan, studies conducted in Japan were classified into three types (practical reports, course evaluation survey studies, empirical studies investigating language development) and representable studies were reviewed. The focus was narrowed down to writing development in the CBI/CLIL contexts, and the researcher points out the need to pay attention to writing task features related to CBI/CLIL (writing topic familiarity, integrated writing) to better understand learners' writing performances and development in such context. Empirical studies in the field of TBLT and second language writing, which examine these writing task variables, were then integrated into the review. Based on the discussion, Chapter 3 presents the first study, which investigated the writing performance differences according to the topic familiarity variables (topic related to subject content and general topic) and effect of theme-based instruction on writing. Chapter 4 presents the second study, which explored another writing variable (i.e., two types of integrated writing tasks). Tasks were contextualized in a CBI writing course, and the students' task and course perceptions were also determined through interviews. Chapter 5 presents a general discussion on how the writing task variables scrutinized in the two CBI courses conducted in the studies support EFL learners' writing performance based on the findings. It also concludes the study and presents the pedagogical implications for classroom instruction.

Chapter 2. Literature Reviews

2.1 Implementation and Research of CBI/CLIL

2.1.1 Brief History of CBI/CLIL

CBI is defined as an "umbrella term referring to instructional approaches that make a dual, though not necessarily equal, commitment to language and content-learning objectives" (Stoller, 2008, p. 59). The key feature here is "the integration of language and content" (Brinton & Snow, 2017, p. 3), and it is sometimes difficult to balance both in EFL settings where the major focus tends to be on language (Mohan, 1986; Lyster 2007). Historically, it started to take its form under the influence of Communicative Language Teaching (CLT) that aims to equip students with "communicative competence" (Hymes, 1971) for effective communication in a social context using a second language (L2). CBI is broadly considered as one form of CLT since it offers contexts for developing communicative competence (Duff, 2014). To facilitate communication in the academic context, CBI draws a distinction between the kind of everyday language (i.e., basic interpersonal communication skills, BICS) and the academic language (i.e., cognitive academic language proficiency, CALP; Cummins, 1980) and tries to raise awareness and promote the acquisition of the latter. Along with the advent of the French immersion program in Canada for majority-language speakers and ESL programs in the United States for minority-language speakers, various forms of CBI started to be implemented to meet the needs of an increasing number of immigrant students whose first language (L1) is not English.

There are three prototypical models of CBI, that is, sheltered instruction, adjunct instruction, and theme-based instruction (Brinton et al., 1989). Sheltered instruction is mainly for ESL learners in English-speaking countries, and L2 students are sheltered in a content class with substantial language guides to satisfy the content and language learning objectives

(Brinton & Snow, 2017). Related to this instruction, the Sheltered Instruction Observation Protocol (SIOP) was developed (Echevarría et al., 2017) to effectively make content comprehensible for L2 students and facilitate their language learning, which also has significant implications for CBI for EFL students (for details, see section 2.1.3). With regard to the adjunct instruction, which is also mainly conducted in the ESL setting, a mainstream content course and a language course are paired, and instructors for these courses collaborate to support the students' content and language learning. Typically, only L2 students take the language course, but they learn with native-speaking peers in the mainstream content course (Brinton & Snow, 2017).

Of particular interest for this dissertation is theme-based instruction, which can be applied to classes for all proficiency levels and is widely seen in commercially available textbooks (Brinton & Snow, 2017). A main theme that matches the students' interests is chosen, serving as a springboard for teachers to choose the relevant topics and language forms. These topics can be studied over several lessons, one semester, or one year and can be flexibly accommodated in each teaching context. Some concrete examples of theme-based instruction include Hauschild et al. (2012), and Cumming and Lyster (2016) on environmental education. For Japanese authorized textbooks[1], the idea of theme-based instruction can actually be observed, and attempts have been made to use these textbooks in line with the CBI/CLIL principles (Izumi, 2019); however, due to each teacher's instructional style, they can easily be used as mere language practice materials for translation, attentive reading, and read aloud without meaningful language exchange being intended. Therefore, even if they are using the same theme-based textbooks, it is essential to know how to use them, especially how effectively content can be dealt with in language instruction, to enable

---

[1] In Japan, the textbooks used in public elementary, junior high, and high schools need to be authorized by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT).

communicative language teaching to happen. In this sense, there appear to be numerous benefits for language teachers in EFL settings to learn from CBI/CLIL principles and actual classroom implementation examples.

CBI assumes that content and language emphasis is on a continuum (Brinton & Snow, 2017; Snow, 2014), and several variant forms of CBI, such as English-Medium Instruction (EMI) and Sustained Content and Language Teaching[2] (Pally, 1997, 2001), can be mapped onto this continuum as indicated in Figure 2.1. Among them, the major variant that is gaining much popularity in the Japanese context is CLIL: an approach that has developed in Europe (Coyle et al., 2010). This approach aims to increase the number of multilingual users in the global society through subject teaching in L2, mainly in English, for elementary and secondary school students who can serve as international citizens in the future and tighten the bond of the European Union (Coyle et al., 2010; Snow, 2014). Within the limited class time available for language input, this method successfully increased language input opportunities through L2 subject teaching for students who generally have a common L1 background. CLIL is conducted all over Europe, for instance, in Finland, Italy, Bulgaria, and Spain (Pérez-Cañado, 2012).

Figure 2. 1

Variations of CBI

---
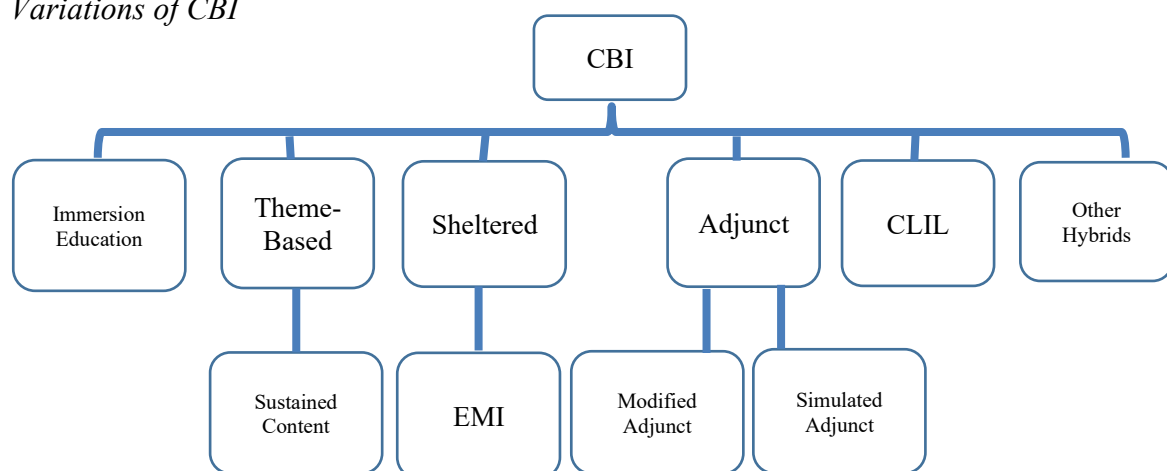
[2] In sustained content and language teaching, a subject area is learned for a continual period of time (e.g., a half to full semester) to advance the learners' critical thinking skills.

CLIL has a commonality with CBI in that both have the primary aim of teaching content and language in an integrated manner and have variations in the amount of focus on content and language. CLIL's variation includes content-driven Hard-CLIL and language-driven Soft-CLIL (Bentley, 2010). The former is similar to immersion and sheltered instruction in that both primarily aim to teach subject content exclusively to L2 learners along with the acquisition of relevant language forms (Brinton & Snow, 2017). In contrast, soft-CLIL shares instructional principles with more language-focused, theme-based CBI. The 4Cs, which are "content (subject matter), communication (language learning and using), cognition (learning and thinking processes), and culture (developing intercultural understanding and global citizenship)" (Coyle et al., 2010, p. 41) are common principles in both kinds of CLIL, and they serve as important cornerstones when constructing CLIL lessons. Another important notion in CLIL is the Language Triptych (Coyle et al., 2010, p. 37), which conceptualizes the types of language used in CLIL and their relationships: language of learning (e.g., content-specific languages, such as coal, LNG, and geothermal, to learn about energy production), language for learning (e.g., language for participating in activities and interacting with teachers and peers, such as what do you think? And I agree with~.), and language through learning (e.g., language necessary for learning that emerges from each learning opportunity from each individual. This kind of language cannot always be made into a list in advance but rather the necessity occurs in situ). It is necessary that teachers should know these three language perspectives that emerge in CLIL classrooms and support students to form form–meaning mapping and connect the language objectives and content objectives. Owing to the presence of these comprehensible pillars of instruction, CLIL has been imported to Japan with a fresh perspective on integrated instruction and has spread widely from elementary school to university levels (Izumi et al., 2012; Sasajima, 2011). Although research has been accumulated under two different names (i.e., CBI and CLIL),

Cenoz (2015) claimed that "CBI/CLIL programs share the same essential properties and are not pedagogically different from each other (p. 8)." Therefore, benefits for sharing research results from these two fields exist, which will be enabled by making the following information public for comparison: research context (ESL or EFL, school level, learner's L1) and instructional goals (content course preparation in ESL, content learning in L2 in EFL or foreign language instruction with a focus on the content added). It can promote research synthesis and provide implications for wider teaching contexts (Cenoz, 2015). Therefore, although the two courses created in this dissertation are based on theme-based instruction in the CBI framework, CBI and CLIL are treated as comparable teaching frameworks in this dissertation, and empirical studies from both fields will be explored.

2.1.2 Theoretical Background of CBI/CLIL

The effect of CBI/CLIL on language learning can be underpinned through second language acquisition (SLA) theories. The most fundamental one, as conceptualized by Gass and Mackey (2015), is the interaction approach, which consists of three interrelated hypotheses: the Input Hypothesis (Krashen, 1982), Interaction Hypothesis (Long, 1996), and Output Hypothesis (Swain, 2005). CBI/CLIL can provide abundant input for content learning by means of subject matter readings and lectures, which can be made further comprehensible for the learners by the instructors' thoughtful scaffolding. The input is then further negotiated through teacher–learner and learner–learner interactions and becomes more comprehensible. Swain (1993) asserted that by observing immersion students, it was obvious that they attained a high level of receptive skills but still had difficulty in producing language with certain language forms (e.g., second-person pronouns, Lyster, 2017) remaining non-target like. This observation indicates that learners also need to be pushed to produce output to discern the gap between their productive and receptive skills, reflect on their language use, and explicitly

learn grammar. Through CBI/CLIL, creating language output naturally occurs when talking and writing about the content for learning and assessment. Although these three hypotheses can hold a strong case regardless of instructional contexts, one can point out that these hypotheses arose from the observation of learners in ESL contexts where the amount of input, interaction, and output inside and outside the class is considerably available than for those learners in the EFL contexts. Therefore, when EFL contexts are being considered, explicit language instruction may be necessary to compensate for the lack of an L2 environment and promote language learning efficiently.

Other benefits of CBI/CLIL can be supported by various cognitive, affective, and social theories of SLA (Fitzsimmons-Doolan et al., 2017). From the cognitive perspective, research from neuroscience broadly supports CBI because "the tendency for the brain to consider the entire experience and to search for meaningful patterns calls for thematic, content-based interdisciplinary language instruction at all levels" (Heyden, 2001). From a more specific approach, transfer appropriate processing (TAP; Roediger & Guynn, 1996) underpins that those language items learned in a CBI/CLIL setting are highly transferable to a similar language use context that involves similar cognitive processes. TAP may support the benefits of CBI/CLIL over isolated item-based learning. Another support from a cognitive perspective is the depth of processing (DOP; Craik & Lockhart, 1972), which is defined as "the relative amount of cognitive effort, level of analysis, and elaboration of intake, together with the usage of prior knowledge, hypothesis testing, and rule formation employed in decoding and encoding same grammatical or lexical item in the input" (Leow, 2015, p. 204). DOP assumes that deeper processing of information enabled by cognitively demanding content learning forms associations between incoming new information and prior knowledge and strengthens the memory of that new data. It is expected that the need for processing both content and language coherently and meaningfully presented in the CBI/CLIL classroom creates high

cognitive load and induces deeper cognitive processing, and certain information can be retained well (Fitzsimmons-Doolan et al., 2017). As for the affective perspective, it can be assumed that the integration of content that is relevant to the learners' interests and is perceived as worthwhile to learn may motivate them to learn the target language (Doiz et al., 2014). However, it is also demonstrated that various aspects of CBI/CLIL (e.g., subject, learners' proficiency levels, social and parental needs toward L2) affect the level of motivation among learners (Lasagabaster, 2017). Additionally, CBI/CLIL can be viewed from the perspective of the sociocultural theory (Vygotsky, 1978) since tasks that facilitate the co-construction of knowledge through scaffolding in the Zone of Proximal Development (e.g., discussion, group presentation, and collaborative writing) can be implemented in CBI/CLIL (Walqui, 2006). The above-mentioned support from a wide range of theories suggests that the inclusion of CBI/CLIL in EFL settings may have a positive impact not only on proficient learners but also on learners at all levels.

Reinforced by these theoretical underpinnings, extensive research has been compiled in ESL and EFL contexts worldwide. Briefly, positive results of CBI/CLIL have been reported in comparison to non-CBI/CLIL courses, for example, on comprehension skills (Lindholm-Leary & Genesee, 2014; Tedick & Wesely, 2015; Yang, 2015), writing (Dalton-Puffer, 2011; Kong, 2014), speaking (Dalton-Puffer, 2011), and pronunciation (Hanzawa, 2019). However, mixed results on language outcomes (i.e., no difference between CBI and non-CBI) have also been reported (Graham et al., 2018); therefore, a closer look at the testing materials and classroom instruction is indispensable. In addition, the investigation of CBI/CLIL in the EFL settings is still a research niche to be filled (Fitzsimmons-Doolan et al., 2017; Pérez-Cañado, 2012). As Lyster (2007) mentioned, simply focusing on content teaching in L2 does not imply that language learning occurs automatically, and careful curriculum development and lesson planning is crucial to maximize the benefits of CBI/CLIL, especially in EFL settings.

2.1.3 CBI Lesson Planning Models

For the effective integration of content and language, teachers need to know various teaching strategies and apply them in their daily teaching. In the classroom, CBI/CLIL is most effectively carried out by combining the use of various language skills, such as speaking, listening, reading, and writing (Garner & Borg, 2005). It is also vital to set content and language learning goals, write them down, and share them with the learners (Snow, 2014) so that both teachers and students share common grounds and goals for learning. Besides, Snow (2014) introduced the following five general instructional strategies for teachers: modifying input (e.g., using easier common words), using contextual cues (e.g., gestures and multimodal input), checking for understanding (e.g., asking comprehension questions), designing effective lessons, and designing language- and discourse-rich activities (pp. 448–449). Related to these five principles, several models that help to effectively combine content and language teaching and support lesson planning for teachers have been proposed. These are Mohan's (1986) knowledge framework, SIOP model (Echevarría et al., 2017), Six Ts model (Stoller & Grabe, 2017), and Lyster (2011)'s counter-balanced approach.

In his book entitled *Language and Content*, Mohan (1986) tried to unveil the relationship between thinking skills and language that is used to enable learning. When teachers choose content, organize the learning activities, and select a language focus, the knowledge structure presented in the book can be a useful model. It explains that the structure of any knowledge can be divided into specific, practical action situations (description, sequence, and choice) and general, theoretical background knowledge (classification, principles, and evaluation). Each action situation is paired with one of the theoretical backgrounds (description–classification, sequence–principles, choice–evaluation). These two aspects (i.e., action situations and background knowledge) comprise the knowledge

framework. Figure 2.2 shows the example of chess analyzed in this framework. A beginner

first needs to identify each chess piece (practical action situation) and learn to classify them

according to their functions (theoretical background information). As shown in the figure,

any learning activity, such as playing chess, can be broken down into practical action

situations and theoretical backgrounds. Mohan (1986) proposed that children first learn a

language in a specific and concrete action situation; however, L2 learners typically lack such

practical action situation, and hence, their learning tends to be decontextualized (pp. 56–57).

Therefore, contextualized learning enabled by CBI is promising for language development.

Figure 2. 2

Example Analysis of Chess in the Knowledge Framework (Mohan, 1986, p. 43)

| | ACTIVITY: CHESS | | |
	Action Situation (Practical)	Background Knowledge (Theoretical)	
Description	Identify chess pieces	Classify chess pieces	Classification
Sequence	Sequence moves	Understand the rules for moves	Principles
Choice	Choose appropriate moves	Evaluate moves according to strategies	Evaluation

To drive learning, certain kinds of thinking skills, visual aids, and language are

necessary, which are summarized in Table 2.1 (Slater & Beckett, 2019). Through this

framework, cognitive processes for content and language learning become explicit for

teachers, thereby enhancing the quality of lessons and students' learning. Previous studies

(Kong, 2015; Moore & Lorenzo, 2015) have adopted the knowledge framework as a guiding

principle for CBI/CLIL curriculum and material development. Related to the six dimensions of the knowledge structure, Bloom's taxonomy (Bloom, 1956; Anderson et al., 2001) also identified six levels of the cognitive process, which are further divided into lower-order processing (remembering, understanding, applying) and higher-order processing (analyzing, evaluating, creating). In CLIL literature, this taxonomy is often referred to for creating a thinking curriculum to support students' development in thinking and problem-solving skills (Coyle et al. 2010, p. 30). It can be said that the practical action situation in the knowledge structure corresponds to Bloom's taxonomy's lower-order processing while the theoretical background knowledge corresponds to the higher-order processing, and both remind teachers that the practice of various thinking skills (not just dealing with lower-order thinking skills) are necessary for successful content and language learning in CBI/CLIL.

Table 2.1

Key Thinking Skills, Visuals and Language in Knowledge Structure in Slater and Beckett (2019) Based on Early (1990) and Mohan (1986)

Knowledge structure	Thinking skills	Key visuals	Language
Classification	Classify Group Sort Define Part/whole	Tree Web Table	General reference Relational verbs (e.g., *be, have*) Additive conjunction (e.g., *and*) Taxonomic, part/whole lexis (e.g., nouns: *types, classes, kinds, categories, ways*; verbs: *classify, sort, group, organize, categorize, divide, comprise*) Passives (e.g., *are classified, are grouped*)
Principles	Explain Predict Draw conclusions Apply rules, causes, effects, means, ends Formulate, test, and establish hypotheses	Cycles Line graphs Cause/effect chains	General reference Action verbs Consequential conjunction and adverbials (e.g., *since, due to, in order to, consequently, because, thus, if-clauses*) Cause‑effect lexis (e.g., nouns: *cause, effect, result*; verbs: *cause, produce, bring about*) Passives + agency (e.g., *is caused by, are*

	Interpret the data		*produced by*)
Evaluation	Evaluate Rank Judge Criticize	Grid Rating Chart	General reference Thinking verbs (e.g., *believe, think, value, consider, rank, judge*) Comparative conjunction (e.g., *likewise, however, while*) Evaluative lexis (e.g., nouns: *best, worst*; adjectives: *good, bad, right, wrong, boring, acceptable*; verbs: *rank, approve, value, like*)
Description	Identify Label Describe Compare Contrast Locate	Picture Map Diagram Drawing Venn Pie chart	General or specific reference Relational verbs (e.g., *be, have*) Existential verbs (e.g., *there is/are*) Additive conjunction (e.g., *and*) Attributive lexis (e.g., *adjectives of color and size*) Language of comparison and contrast (e.g., *the same as, similar to, different from*)
Sequence	Arrange events in order Note changes over time Processes Follow directions	Time line Action strip Flowchart	Specific reference Action verbs Temporal conjunction and adverbials (e.g., *after, since, as, initially, firstly, finally, when-clauses, as-clauses*) Sequential lexis (e.g., nouns: *beginning, end*; verbs: *start, conclude, continue, summarize*)
Choice	Select Make decisions Propose alternatives Solve problems Form opinions	Decision tree	Specific reference Sensing verbs (e.g., *like, want*) Alternative conjunction (e.g., *or*) Appositional choice lexis (e.g., nouns: *choice, option, which + noun*; verbs: *choose, opt, select, prefer*)

From a more holistic and strategic point of view on lesson planning, the SIOP model (Echevarría et al., 2017) presents eight crucial components in making a sheltered class more comprehensible for ESL students in the U.S. while improving all four language skills (i.e., reading, listening, speaking, and writing). The eight components are lesson preparation, building background, comprehensible input, strategies, interaction, practice and application, lesson delivery, and review and assessment. Each component has sub-features, which are summarized in Table 2.2. For example, the first component lesson preparation includes three sub-components, namely setting the language and content objectives, choice of appropriate

content, and adaptation of materials. The model explains the theoretical reasons why it is crucial to share the language and content objective with learners and shows example lesson planning formats. As can be seen in the other components in the table, the model makes the good practices of CBI teachers visible and accessible, backing them up with educational research and language teaching research evidence. It provides the overall framework to plan, conduct, and assess CBI courses.

Table 2.2

Key Components and Sub Features of the SIOP Model in Echevarría et al. (2017)

Components	Sub features
1. Lesson preparation	Define, display, and review content and language objectives
	Choose content that is age-appropriate and matches the students' educational background
	Prepare supplemental materials, adapted content materials, and meaningful activities
2. Building background	Connect to the students' background experience and past learning
	Support vocabulary learning
3. Comprehensible input	Adjust teacher talk to the students' proficiency levels
	Explain academic tasks clearly
	Make content comprehensible by using various strategies (gestures, pictures, multimedia, showing a model of task procedure, repeating important words and concepts)
4. Strategies	Give opportunities for students to use learning strategies
	Scaffold the students' understanding
	Promote higher-order thinking through questions and tasks
5. Interaction	Give ample opportunities for interaction
	Have students work in various grouping configurations (whole class, small group, pair)
	Provide sufficient wait time for response

	Allow students to clarify the content in L1
6. Practice & application	Use hands-on materials and/or manipulatives
	Apply content and language knowledge
	Use activities to promote all four language skills' development
7. Lesson delivery	Deliver lessons to achieve content and language learning objectives
	Enhance student engagement
	Control the pacing of the lesson
8. Review & assessment	Review vocabulary and content concepts
	Feedback on student output
	Assess student comprehension and learning

Although the model targets ESL secondary school learners in the U.S., the principles are equally useful for lesson planning in the EFL contexts (Short et al., 2011). Research evidence suggests that the degree of the SIOP model application and students' achievement on science language assessments on non-essay and essay components are slightly positively correlated (Short et al., 2011), but statistical significance was not found when comparing the SIOP model implemented group and the control group (Echevarría et al., 2011; Short et al., 2011). The results may imply the difficulty in assessing the effects of the entire model that covers various aspects of the lessons (e.g., lesson delivery, variety of content subjects, tasks used, the level of scaffolding) in addition to various external factors present (e.g., teachers' fidelity to the model implementation, number of participants, implementation period). However, the SIOP model is definitely useful for conducting comprehensible lessons and contributes to content and language learning not only for ESL but also for EFL students.

Another framework for planning CBI courses, especially theme-based courses, is the Six Ts framework (Stoller & Grabe, 2017). This approach attempts to create coherence in the theme-based course by making connections among the six Ts, which are *themes, topics, texts,*

tasks, transitions, and *threads*. Coherence is an especially important notion in planning CBI since learners can make greater gains in learning through connections between past learning and new learning. *Themes* are the central ideas that become the basis of the other five Ts and are relevant for the target learners. For instance, themes can be abstract (e.g., liberty) or concrete (e.g., Japan's peacekeeping activities), and the period for dealing with one theme can be flexible (e.g., one semester, one year) according to each teaching context. *Topics* are the sub-units of a theme, which are the small and related topics to describe a theme. Using the above example, personnel contributions, financial contributions, and intellectual contributions can be the topics under the theme of Japan's peacekeeping activities. With the same theme, a different set of topics can be used (e.g., Japan's peacekeeping activities in Congo, Sudan, and Mali). *Texts* are the materials, both written and aural, which facilitate content and language teaching. Various types of texts (e.g., reading materials, videos, guest speakers, worksheets) are necessary to enable the tasks to unfold. *Tasks* are the basic units of CBI lessons in a classroom and realize content learning, language learning, and strategy skill learning. *Transitions* make explicit connections between topics and between texts and tasks to establish coherence in the course by connecting the new with known information. Relationships among a theme, topics, tasks, and transitions are shown in a figure by Stoller and Grabe (2017). Lastly, the *thread* is the linkage among the different themes covered in the CBI course. In many cases, a theme-based textbook has diverse themes (e.g., an anime hero, the development of phones, the Olympics' history) that seem to be unrelated, but if an overarching thread is set (e.g., things that bring happiness to everyone's lives) teachers can develop greater coherence into their whole course. By paying attention to the coherence detailed in the Six Ts model, they can offer well-linked content and language course (Osman et al., 2009).

With more explicit attention to language learning, Lyster (2007), based on his

investigations of the students' language development in French immersion schools in Canada, proposed the counterbalanced hypothesis in planning CBI courses. This hypothesis predicts that "interlanguage restructuring is triggered by instructional interventions that orient learners in the direction opposite to that which their target language learning environment has accustomed them" (Lyster, 2007, p. 126). In immersion classrooms, students are exposed to meaningful input, and their receptive skills develop more than their output skills. Therefore, according to the hypothesis, less salient grammatical features, such as grammatical gender in French, need to be explicitly taught through form-focused instruction. In contrast, in EFL contexts, especially in Japan, the instruction is predominantly focused on explicit grammar teaching; therefore, meaning-oriented content learning in L2 rather than an extensive focus on forms may facilitate interlanguage development.

To integrate language instruction, both reactive (e.g., providing corrective feedback, such as recast in response to errors in learners' oral production) and proactive (e.g., pre-planned language tasks) language instruction can be given. One proactive way of systematically focusing the students' attention on language is proposed by Lyster (2017). As shown in Figure 2.3 (Lyster, 2017, p. 119), a CBI unit can begin with attention to content, move on to the noticing and awareness stage of a language feature that is salient in the content, provide practice of the feature in guided and autonomous ways using tasks related to the content, and then focus on the content again. In this overall instructional sequence, learners can contextualize their language learning in certain content and raise awareness about and practice otherwise unnoticed grammatical forms systematically without compromising on content learning.

A comparison of the four frameworks explained above reveals that each one of them has a different focus and is useful for constructing CBI/CLIL courses in not only ESL but also EFL settings. The focal points are cognitive functions necessary for learning about

Figure 2. 3

Lesson Sequence Suggested in Lyster (2017)

Focus on content

Noticing activity

Awareness activity

Guided practice

Predominant focus on language

Focus on content

Autonomous practice
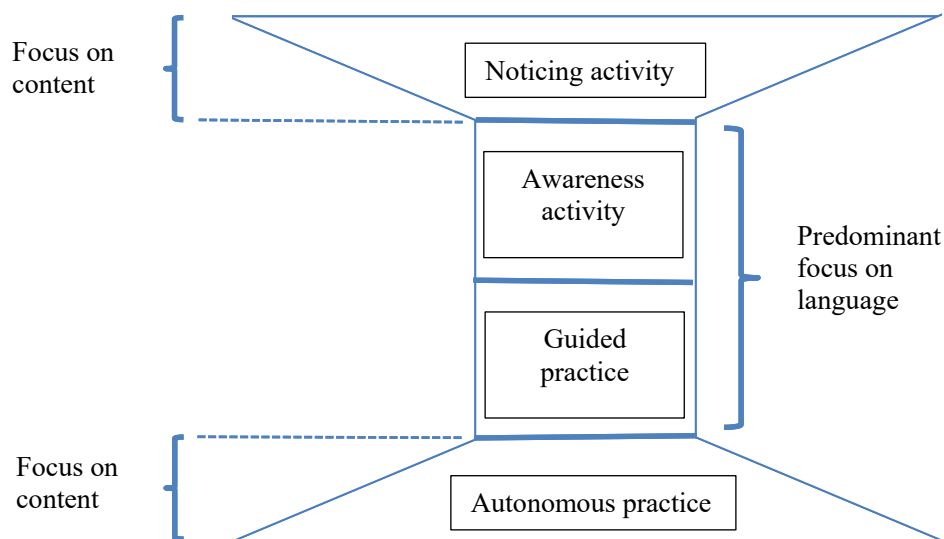
content and language (Mohan, 1986), offering educational support to L2 students (Echevarría et al., 2017), building coherence in CBI (Stoller & Grabe, 2017), and providing language-focused instruction in CBI (Lyster, 2007, 2017).

Considering the development of CBI classes in EFL contexts, the process may take the opposite direction to ESL contexts; in ESL contexts, meaningful L2 input, interaction, and output are naturally present inside and outside of the classes, and form-focused instruction needs to be intentionally integrated into content learning (Lyster, 2007) whereas in the EFL context, such as in Japan, form-focused instruction is the baseline, and the integration of meaningful content has gained importance in just the past few decades. Even so, form-focused instruction is essential since the Japanese language is structurally distinct from English, and English input is scarce in Japan. In particular, low-level learners need language support, and proactive and reactive focus on form is empirically shown to benefit their language and content learning in CBI (Grim, 2008; Valeo, 2013). If classroom instruction in Japan, which is the only source of L2 input for most Japanese people, could include meaningful tasks for content learning in addition to form-focused instruction, they may be

able to gain practical communicative skills along with accurate linguistic skills (Lyster, 2007).

In addition, as suggested in Dalton-Puffer (2013) and Mohan (1986), focusing on the functional aspects of language, such as comparing and reasoning, instead of focus-on-forms (rule-based learning of pedagogic grammar) may positively shift our focus to meaning oriented language use, which is not as common in EFL contexts. The CBI/CLIL framework may help teachers and learners become aware of the functional aspects of language.

Furthermore, all models necessitate meticulous planning about the various aspects of lessons, such as the choice of content and language objectives, preparation of texts and tasks, teachers' comprehensible lesson delivery, provision of feedback, and building coherence in the entire course. Integration of content and language requires preparation for both, and especially, language teachers need to study the content to be able to teach it in L2. In addition, teaching content and language in an integrated manner is different from teaching them separately as the models show that integration is something that is intentionally made. Although practical reports on CBI/CLIL implementation in Japan have been increasing as discussed in the following section, further research accumulation in reference to these models is necessary.

2.1.4 Application of CBI/CLIL in Japanese Classroom Settings

Various forms of CBI/CLIL have been implemented not only in the U.S. and Europe but also in Asian contexts, ranging from content-focused immersion and EMI programs (Kong, 2015; Yang, 2015) to language-focused theme-based programs (Chapple & Curtis, 2000; Ikeda, 2013; Kashiwagi & Kobayashi, 2019; Suzuki, 2022) because CBI can be considered as "a curricular and instructional approach well adapted to transform the FL classroom into the site where intellectually stimulating explorations can become the norm

rather than the exception" (Cammarata, 2016, xiii). Especially in Japan, it is not rare to see classes heavily focused on mechanical grammar drills and exam preparation (e.g., for Test of English for International Communication, TOEIC) in which meaningful communication is limited even though the Courses of Study and university English classes aim to enhance students' communicative competence. Therefore, "there is also a need to continue exploring ways of integrating CBLT (Content-Based Language Teaching) in language-driven classrooms as a means of enriching classroom discourse and increasing opportunities for purposeful communication" (Lyster, 2017, p. 103). Creating enriched language learning opportunities in EFL contexts is especially critical.

Although researchers and practitioners are aware of the benefits of CBI/CLIL in foreign language contexts (Watanabe et al., 2011), several challenges can be pointed out (Butler, 2011). From the teachers' perspective, they need to be trained to be able to handle content and language teaching in the target language, which is especially difficult in EFL settings where the teachers' L2 proficiency is not necessarily high, and they may not be confident enough to teach in L2 (Butler, 2005; Murphey, 1997). Currently, such teacher training is only available to teachers who already know about CBI/CLIL and are motivated enough to participate in such training. Moreover, they often need to modify, add, and create their original materials, which may add to the difficulty and consume their time. Additionally, to create materials, they need to become knowledgeable about the content chosen. In contrast, students may encounter difficulties due to the increased focus on content and higher cognitive complexity imposed by higher-order thinking tasks. Overwhelmed learners may feel at a loss in the class (Suzuki, 2013), and their motivation for L2 learning may suffer. To avoid these pitfalls and bring out the best of CBI/CLIL, the difficulty levels should be controlled through careful planning and extensive understanding of each task characteristic to be used, and empirical evidence should be compiled.

2.1.4.1 Empirical CBI/CLIL Studies in Japan

Most of the CBI/CLIL studies conducted in Japan are practical reports or conceptual papers that greatly inform teachers who implement CBI/CLIL in their own classrooms. However, empirical investigation is still rather scarce (Yanagawa, 2017), and there are much fewer studies on learner language development (Ikeda, 2013). Existing CBI/CLIL studies in Japan mainly try to share the course syllabus and teaching materials that may be useful for other teachers (practical reports), investigate students' course perceptions through surveys (survey studies), and measure students' linguistic outcomes (empirical studies). The studies included in this review were conducted in classrooms, excluding the studies in the immersion, ESP, and EMI contexts.

A wide variety of content is chosen, namely home economics (Clark, 2013: Takagaki & Tanabe, 2007), anthropology (Santos, 2013), intercultural communication (Kavanagh, 2019), animals (Yamano, 2013), international relations (Underwood, 2019), manufacturing (Ichimura et al., 2021), global issues (Ikeda, 2013; Yanagawa, 2017), power generation issues (Suzuki, 2022), linguistics (Izumi, 2012), and physics (refraction; Kashiwagi & Kobayashi, 2019) with various student populations (e.g., elementary schools, Yamano, 2013; junior high schools, Clark, 2013; high schools, Ikeda, 2013; universities, Santos, 2013; graduate schools, Ichimura et al., 2021) and different proficiency levels (e.g., beginners, Yamano, 2013; low–intermediate, Kashiwagi & Kobayashi, 2019; advanced, Izumi, 2012). As for practical reports, especially after the import of CLIL in the last few decades, the number of research presentations and practical reports has increased drastically, thus driving the shift from focus-on-forms to more meaning-oriented classroom practice. By simply reading a few examples (Clark, 2013; Santos, 2013; Takagaki & Tanabe, 2007), one can see the potential of the inclusion of content for enriching classroom communication.

2.1.4.2 Survey Studies Conducted in CBI/CLIL Classrooms in Japan

Along with sharing course syllabuses and materials, some studies conducted questionnaire surveys, interviews, and classroom observations to investigate the students' class perceptions experimentally. For instance, Ichimura et al. (2021) investigated graduate students' perception of the five CLIL manufacturing lessons conducted as part of an English class through questionnaires and interviews. The results confirmed their enhanced motivation and perceived development of speaking skills owing to situated communicative tasks.

Some studies have applied CBI/CLIL to a university writing course. Kavanagh (2019) presented a detailed course description of an academic writing course combined with intercultural content learning, and the survey indicated that they experienced an improvement in writing and speaking skills and showed positive attitudes toward the CLIL course. Similarly in a writing course, Underwood (2019) conducted a freshman CLIL writing course for CEFR A2+ level university students with international relations as the content. Students' writing assignments received more than 80 points, thereby indicating successful learning. They positively perceived their learning experience in CLIL in terms of the content, materials, and activities used. Conversely, they endured difficulties in content-specific vocabulary and idea generation for completing the writing task. Thus, it can be concluded from these studies that conducting a writing class in the CBI/CLIL framework can be successful and perceived favorably in a Japanese context.

The following two studies (Izumi, 2012; Yanagawa, 2017) placed emphasis on the importance of tasks as the building blocks of CLIL lessons, thereby suggesting that investigating the students' task performance can be one way to explore the effect of CBI/CLIL on language learning. Yanagawa (2017) conducted a four-skill development course for university students on the topic of global issues by utilizing tasks as enabling tools for content and language learning. According to a questionnaire survey, it was found that all

students perceived the class as meaningful, which may be positively influenced by the teacher being a Japanese, non-native speaker of English and using L2 English to teach. The author assumes that the students may have been encouraged to use English themselves to participate in the class by seeing a Japanese-speaking teacher who is not necessarily fluent in English teaching content, and thus, their perceived meaningfulness of the class became heightened. Students' written class reflection analysis revealed that they thought that the class provided enjoyment, usefulness, capability, and novelty that may have been brought on by the addition of thought-provoking content and tasks that facilitate cooperative learning. Izumi (2012) conducted a CLIL class with English linguistics as the content for non-English major first-year university students with high English proficiency using pedagogic tasks to enable content learning and cooperative learning. The questionnaire survey revealed that they positively reflected on the class due to the meaningful content and tasks as in Yanagawa (2017), and also noticed an improvement in their English skills.

In summary, the studies reviewed show that in Japanese contexts with a wide variety of participants in terms of school types (elementary school to university) and English proficiency (beginner to advanced), the integration of content learning into English classes was generally perceived positively by the students and teachers in the studies reviewed. However, the feeling of difficulty observed by Underwood (2019) may have the possibility to negatively influence the students' affect and learning, and the learners in Aoyagi et al. (2016) pointed out a lack of language instruction besides content focus. As seen in all the CBI/CLIL studies reviewed, tremendous care was taken for class preparation (e.g., choice of content and language goals, choice of level-appropriate tasks and task sequence, learning about content, preparation of level-appropriate materials, assessment) and delivery (e.g., adjustment of teacher talk to be comprehensible, use of L1, taking the appropriate amount of time for task completion) so that the class would not overwhelm the students in terms of the novelty of

content and level of input language; otherwise, the difficulty of learning in CBI/CLIL may negatively affect the learners, causing anxiety (Suzuki, 2013) and a feeling of marginalization (Kiyota, 2022). In other words, if teachers plan well, the inclusion of content learning in EFL classrooms can act as a strong support for language learning.

2.1.4.3 Empirical Studies Measuring Students' Linguistic Outcomes in Japan

Besides survey studies, a few studies have tried to examine the impact of CBI/CLIL on learner language development, such as vocabulary, grammar, and writing, in classroom contexts in Japan. For example, Yamano (2013) compared three 5th-grade elementary school CLIL and non-CLIL lessons and conducted a student questionnaire survey. Survey results indicated that the students experienced a variety of emotions (e.g., enjoyment, sadness, sympathy, satisfaction) in the CLIL class, and a close investigation of the recorded student talk revealed that they used the target vocabulary more than their non-CLIL counterparts. The data shows CLIL's language-enriching effect and positive impact on the elementary learners' vocabulary learning. Kashiwagi and Kobayashi (2019) investigated grammar learning in the CLIL lessons at a high school. They conducted eight CLIL lessons on physics (i.e., refraction) with the language focus on the 1st and 3rd conditionals (e.g., If I ~, I will ~. vs. If I had X-ed ~, I could have X-ed ~.), which were used by the teacher and students in conducting refraction experiments. In contrast, in the non-CLIL lessons, learners were explicitly taught the rules and completed the regular textbook activities. Students in both groups wrote journals in English five times after the lessons, reflecting on what they had learned in class. Findings of the grammatical judgement tests revealed that the CLIL group outperformed the non-CLIL group, possibly due to the repeated input of the target grammar authentically used in a meaningful context (e.g., experiments). Moreover, the analysis of the students' reflection writing suggests that CLIL students tended to improve their writing

fluency and use more words with cognitive discourse functions (CDFs; Dalton-Puffer et al., 2018). CDF "consists of a seven-fold categorization of verbalizations which express acts of thinking about the subject matter in the classroom (Classify, Define, Describe, Evaluate, Explain, Explore, Report)" (Dalton-Puffer et al., 2018, p. 5). This categorization is thought to help content teachers gain insight into the language functions in content learning, and Kashiwagi and Kobayashi (2019) confirmed the use of these verbs in the students' written output.

The following three studies enquired into students' writing to see the evidence of language learning in CBI/CLIL. Goya (2018) examined vocabulary development by conducting a university writing class with a focus on exploring local issues (e.g., Okinawa's globalization) as the content and investigating the lexical changes in the students' in-class TOEFL writing. The writing analysis using an online analysis tool (the Complete Lexical Tutor) suggests that students used easier words (K1 vocabulary) and avoided using difficult words (K3 vocabulary) and academic words in the later writing (Week 13) compared to the earlier writing (Week 5). Although the author admits that it may be challenging to observe the connection between the results in lexical changes in writing and the effect of the course without a control group, it can be predicted that students tended to use vocabulary that they have easier access to as they gained experience in writing. It may be fruitful to explore the relationship between the topic of the writing prompts and the content in the CBI/CLIL course to draw clearer implications about learning transfer in terms of the lexis.

Ikeda (2013) created a year-long language-driven soft CLIL course for high school students with topics related to global issues and conducted a questionnaire survey and analysis of their writing samples as a course evaluation. The survey results showed that the course was positively perceived as being denser than other normal English courses owing to the integration of the 4Cs, and the learners' writing was better in the post-test in terms of the

holistic rating by criterion, fluency, and lexical complexity, but the accuracy deteriorated. The author mentioned that teachers did not correct the errors in the in-class writing activities, and this treatment may have affected the accuracy of the result. It is also speculated that the students prioritized meaning over accuracy in the post-test writing task through receiving meaning-oriented CLIL class. Although the improvement in writing cannot be solely attributed to the CLIL course since the students were also taking other English courses, the study concludes that the soft version of CLIL is feasible in Japanese high schools and can have favorable effects on writing.

Suzuki (2022) conducted a CBI class that dealt with the issues of power generation methods with a focus on compare/contrast language. A non-CBI class that focused on the same language targets with cognitively less dense everyday topics was also conducted. Students' writing performance did not differ between these two classes, but the learners in the CBI class made considerable improvement in the power generation writing topic in the post-test, and those in the non-CBI class made greater improvement in the general writing topic, suggesting that their learning is highly influenced by the content dealt with in the classroom. This is thought to be especially true in the EFL context where the amount and varieties of students' language input is limited. Therefore, in investigating the learning outcomes of CBI in EFL contexts, the writing topic should be cautiously chosen.

From the review of the above empirical studies conducted in Japan so far, a major issue emerged. In spite of knowledge about CBI/CLIL implementation in Japan being accumulated in the form of practical reports, a relatively fewer number of studies have empirically examined the learning outcomes of students. Investigation into the students' English use and development as well as the impact of CBI/CLIL on their affect (e.g., class perception, motivation, anxiety) should be explored so that CBI/CLIL can establish a firm root in Japanese English education as an evidenced teaching framework. Since there are several

ways of CBI/CLIL implementation, it may not be straightforward to measure the general effect of CBI/CLIL. However, empirical evidence is especially valuable for language education policymakers and teachers to decide on teaching methodologies with confidence. To guide this line of research, some studies (Izumi, 2012; Yanagawa, 2017) focused on tasks. One promising way to effectively investigate the influence of CBI/CLIL on learner language is to break down the unit of analysis from the entire course to each learning task for the content and language learning. How the manipulation of tasks affects language performance and learning has been increasingly investigated in the field of TBLT, and TBLT research findings can greatly inform the learning outcomes of CBI/CLIL (García Mayo, 2015).

2.1.5 Empirical Investigation of Writing Development in CBI/CLIL in EFL Countries

2.1.5.1 Writing as a Language Development Measure in CBI/CLIL

Among the different ways of assessing learner language, writing can be one of the most relevant language skills to assess the students' language outcomes in CBI/CLIL. For example, it is an important academic task for mainstream content learning; thus, it also needs to be practiced in CBI/CLIL (Jexenflicker & Dalton-Puffer, 2010; Weigle & Jensen, 1997). This is especially true in university preparatory CBI courses for ESL students to successfully transfer to mainstream content classes; however, writing is also an essential skill for EFL students since it clearly reflects their understanding of the content and their use of L2. Due to the crucial role that it plays in content learning, integrating the CBI/CLIL approach into an academic writing course is widespread (Heyden, 2001; Kavanagh, 2019; Shih, 1986; Underwood, 2019). Moreover, writing enables skill integration; especially, reading and writing integration is an authentic and vital task for content and language learning (Pally, 2001; Plakans, 2015; Shih, 1986, 1999). Students obtain content knowledge and

simultaneously increase their language resources through reading, and they consolidate both types of knowledge by writing. Despite the close relationship between reading and writing in CBI/CLIL, the interface has not yet been fully explored in previous CBI/CLIL research, especially for EFL students.

Possible effects of CBI/CLIL on various aspects of writing development have been investigated in previous studies with diverse writing tasks, which are analyzed in terms of complexity, accuracy, and fluency (CAF) and rating (Artieda et al., 2020; Bulté & Housen, 2019; Gené-Gil et al., 2015a,b; Ikeda, 2013; Jexenflicker & Dalton-Puffer, 2010; Jiménez Catalán & Agustín Llach, 2017; Lahuerta, 2020; Lasagabaster, 2008; J. Lee, 2020; Lo & Murphy, 2010; Pérez-Vidal & Roquet, 2015; Roquet & Pérez-Vidal, 2017; Ruiz de Zarobe, 2010; Tejada-Sánchez & Pérez-Vidal, 2018). Majority of the research reviewed in this section is CLIL research in Europe where comparison research (CLIL and non-CLIL) is extensively conducted.

2.1.5.2 Input Time Unmatched CLIL Writing Studies

Increased amount of input is one of the biggest benefits of introducing CLIL; however, it may be beneficial to control the input amount in empirical CLIL studies to probe the impact of CLIL instruction. A large body of research compares CLIL and non-CLIL students without taking into account the amount of total English input. For instance, Lahuerta (2020) compared Spanish secondary school third and fourth-year CLIL and non-CLIL students' argumentative English writing cross-sectionally in terms of accuracy (i.e., error-free sentences) and concluded that there was evidence of progress only for the CLIL group. However, the amount of English input for the two groups was not mentioned. In contrast, Lasagabaster's (2008) cross-sectional analysis confirmed better ratings of letter writing performance in English among Spanish secondary school third-year CLIL students (three

hours of formal English instruction per week + four hours of CLIL classes per week), compared to their fourth-year non-CLIL counterparts who were a year older (three hours of formal English instruction per week). However, again, the total amount of English input up to these grades is unclear in this study.

A similar comparison was conducted in Asian secondary schools, which pointed out the benefit of increased input enabled by CLIL introduction. J. Lee (2020) compared the writing of 11th-grade Korean CLIL and non-CLIL students with the same English proficiency level (advanced) and confirmed the CLIL students' better performance on personal narrative writing in terms of syntactic complexity, lexical complexity, and writing quality (e.g., content, organization, vocabulary, and language use). Since their language proficiency was controlled, the author claims that increased input through the addition of CLIL and language experience in content learning may have positively affected the CLIL students' writing. This study supports the introduction of CLIL in Asian contexts where the language system is considerably different from English. Moreover, Lo and Murphy (2010) focused on vocabulary learning, which forms a firm basis for writing, and compared grade 7 and 9 late immersion (20 hours of immersion classes per week) and EFL (eight hours of English classes) students' passive and controlled active vocabulary through the Vocabulary Levels Test in Hong Kong. Findings suggested that the immersion students had a more passive and active vocabulary for both grades. The positive results may have arisen from a wider variety of vocabulary that immersion students are exposed to and higher chances to use the vocabulary in the immersion classes with an increased amount of input. From these studies, it can be concluded that additional input through CLIL may have a positive impact on writing in terms of CAF and rating. However, the type of input beneficial for written language development also warrants research; is any input conducive to writing improvement or CLIL-type input beneficial for better performance? The improved performance may simply be due

to more contact hours with English for the CLIL group (Dalton-Puffer, 2011).

Even with an increased amount of input enabled by CLIL introduction, it seems that not all aspects of writing are equally influenced. Tejada-Sánchez and Pérez-Vidal (2018) explored the impact of different amounts of CLIL input (8760 vs. 7002 hours) on Colombian secondary school students' narrative retellings and confirmed the positive effect of increased input on the accuracy, but there was a negative effect on the syntactic complexity and fluency and no effect on the lexical complexity and ratings. The authors, however, pointed out that both groups had already had ample CLIL input before the research was conducted, which may have obscured the impact of the different input amounts. Focusing exclusively on complexity (syntactic, lexical, and morphological), Bulté and Housen (2019) investigated secondary school students' argumentative, descriptive, and narrative writing produced over a 19-month period in the Netherlands. They showed that 15 hours of CLIL added to English classes (975 hours of total English input) as opposed to three hours of non-CLIL English classes (130 hours of total English input) did not differentiate the rate of complexity development, thereby pointing out the difficulty of developing complexity. Thus, it can be inferred that a mere increase of CLIL input may not necessarily mean better writing, but rather as the authors mentioned, the CLIL course's didactic approach and task effect may have influenced the results. Therefore, CLIL research needs to pay attention to these aspects as research variables to further scrutinize the possible CLIL impacts on learning outcomes.

Another confined variable in CLIL research is the selective nature of CLIL programs; students in CLIL may have higher proficiency from the beginning, and their better performance may not be due to the CLIL programs. Jexenflicker and Dalton-Puffer (2010) compared Austrian secondary school science CLIL and non-CLIL students' letter writing, and the CLIL group was rated higher in all four dimensions (i.e., task fulfilment, organization and structure, grammar, and vocabulary), particularly due to higher English proficiency and

awareness of the pragmatic needs of the task. However, without any information on the students' proficiency level at the beginning of the CLIL program, it cannot be confirmed that CLIL instruction leads to better writing performance.

2.1.5.3 Input Time Matched CLIL Writing Studies

To account for the inaccuracy of the amount of CLIL input, the accumulated amount has been reported in CLIL studies. Although slightly inconsistent, an impact of CLIL instruction on writing emerged when the input time was matched. As for vocabulary acquisition, Jiménez Catalán and Agustín Llach (2017) focused on productive vocabulary by time-matched (1189 hours of instruction) Spanish 8th grade CLIL (science and history) and 10th grade non-CLIL students using a lexical availability task (10 prompts, two minutes each). The CLIL group outperformed the other group by the number of word types retrieved, but the frequency and sophistication level of the retrieved words were the same across the groups. The authors speculate that meaningful language use contexts provided by CLIL may have facilitated incidental vocabulary acquisition and increased the students' lexical diversity. However, no difference in sophistication and frequency may be because, as the authors discuss, their secondary school participants are still at the stage of acquiring basic vocabulary rather than acquiring lower-frequency sophisticated words, regardless of the instructional types.

As for CAF and rating measures, there seems to be almost no CLIL effect on syntactic complexity and fluency, and inconsistent results are found for accuracy and ratings. For instance, Gené-Gilet et al. (2015a) found no differences in any of the CAF dimensions between the input time-matched (210 hours) Spanish secondary school grade 8 CLIL and grade 9 non-CLIL groups in descriptive writing. Another research by the same research team (Gené-Gil et al., 2015b) compared input time-matched (210 hours) Spanish 13-year-old CLIL

students and 14-year-old non-CLIL students' letter writing. Again, no difference was found between the groups in terms of CAF, but the CLIL group's rating scores on organization, language use, and total score were higher than the other group. It seems that CLIL's meaning-oriented class in comparison to form-oriented EFL class may have a positive impact on the overall quality of writing rated by human raters rather than on mechanical scores assigned for the CAF.

Pérez-Vidal and Roquet (2015) compared Spanish secondary school input time-matched CLIL (1330 hours at the time of a pre-test, formal English instruction + science CLIL class) and non-CLIL groups' (1260 hours, only formal instruction) written language production on a picture description task unrelated to the content in the CLIL class. The gain scores for accuracy and rating scores (e.g., task fulfillment, organization, grammar, and vocabulary) were higher for the CLIL group. However syntactic complexity, lexical complexity, and fluency did not differ between the groups. Ruiz de Zarobe (2010) compared Spanish secondary school CLIL and non-CLIL learners' letter writing with the total English input almost held constant (CLIL 910 hours vs. non-CLIL 990 hours). Similar to Pérez-Vidal and Roquet (2015), the results indicated that the CLIL students' writing gain scores were higher for all analytic measures (content, organization, vocabulary, language, and mechanics), suggesting the positive impact of CLIL combined with conventional EFL lessons for applying language knowledge in a meaningful communication situation. This aspect of the benefits of CLIL should be pronounced in EFL contexts, including Japan where fewer L2 use opportunities are available.

Some studies do not show an advantage of CLIL on CAF or rating scores. Roquet and Pérez-Vidal (2017) focused on Spanish secondary school students' picture description writing over a year, with the total amount of English exposure for the CLIL students being 1,330 hours at the onset and 1,540 hours after a year and for the non-CLIL students being 1,260

hours and 1,400 hours, respectively. They found that CLIL learners' syntactic and lexical complexity, fluency, and analytic rating scores (i.e., task fulfillment, organization, grammar, and vocabulary) showed no advancement over those of non-CLIL students in spite of more instructional time, except for accuracy. The authors suggest that CLIL classes may have provided meaningful practice opportunities of language forms dealt with in the regular English class. Artieda et al. (2020) investigated picture description writing performances of input time-matched 12- to 13-year-old CLIL students and 14- to 15-year-old non-CLIL students, and the findings suggested that most of the linguistic and rating scores did not differ between the groups, but the non-CLIL group outperformed the other group in accuracy and coordination index. The author points out that cognitive maturity of the older students may have supported the better results for the non-CLIL group. Therefore, age rather than opting for the CLIL module or not may be the more important factor for early secondary school students.

The above input time-matched CLIL studies showed CLIL's mixed effects on vocabulary development (positive effect: Jiménez Catalán & Agustín Llach, 2017; no effects: Roquet & Pérez-Vidal, 2017; Pérez-Vidal & Roquet, 2015), on accuracy (positive effect: Pérez-Vidal & Roquet, 2015; Roquet & Pérez-Vidal, 2017; no effects: Gené-Gilet et al., 2015a, b), rating scores (positive effect: Gené-Gilet et al., 2015a, b; Pérez-Vidal & Roquet, 2015; Ruiz de Zarobe, 2010; no effects: Artieda et al., 2020; Roquet & Pérez-Vidal, 2017), and no effect on syntactic complexity and fluency (Gené-Gilet et al., 2015a, b; Pérez-Vidal & Roquet, 2015; Roquet & Pérez-Vidal, 2017). It can be concluded that by learning a subject content through L2, the learners are exposed to various words, which may support their vocabulary learning; however, receiving such instruction may not necessarily equip them with the skill to use more complex syntactic structures, which may be something to be learned with an explicit instruction for EFL learners. This difficulty of acquiring syntactic

complexity is also observed in total immersion settings, such as French and Spanish total immersion programs, with a higher level of L2 input (e.g., Day & Shapson, 1991; Lyster, 2004). Moreover, although not consistently, the results suggest that CLIL might have improved accuracy and writing quality assessed by ratings. These are both evaluated by human raters; therefore, CLIL's meaning-oriented instruction may have positively affected the qualitative aspects of writing rather than the aspects, such as the syntactic complexity and number of words (i.e., fluency), which are measured mechanically.

So far, CLIL studies mentioned above measure writing development using general writing topics (e.g., picture description task, letter writing) that are unrelated to the content learned in the CLIL classroom, which may be suitable to investigate the impact of CLIL on general writing. However, learning in CLIL is naturally confined to the topic or subject studied; therefore, the dimension to be developed may be greatly influenced by the actual classroom content, and the improvement may not have been captured in general writing tasks. To account for this potential impact of writing topics, the following studies used tasks which cover CLIL topics.

For example, Whittaker et al. (2011) investigated Spanish secondary school CLIL students' history writing over four years in terms of coherence and noun phrase modification. The writing topics were related to the same topics covered in the history class at the time of data collection, and there was no explicit instruction on how to increase coherence. Findings indicated that the coherence improved over four years by utilizing fewer new nouns, increasing the explanation of the introduced nouns, and using more direct references. As for noun phrase modification, the use of pronouns and unmodified nouns decreased, and higher usage of nouns with pre- and post-modifiers appeared. Although there was no non-CLIL counterpart, the authors discuss that the exposure to English in the CLIL class along with the students' cognitive maturity may have supported their development in building coherence.

2.1.5.4 Short-term CBI/CLIL Writing Studies in Asian Contexts

When shifting the focus to Asian contexts, more controlled and short-term CBI/CLIL studies attempting to assess L2 English writing gains related to a subject content have been conducted. Students' L1 grammar and writing systems in these Asian contexts are remarkably different from English, unlike the learners in European settings. Regarding the language-driven Soft CLIL, Ikeda (2013) showed primary evidence for development (i.e., holistic rating, lexical complexity, and fluency) of argumentative writing on a content-related topic (e.g., global affairs) over a year of CLIL instruction at a Japanese high school, which is possibly attributable to the addition of a CLIL-type course to regular English classes. However, the learners' accuracy deteriorated in the post-test, thus implying a possible trade-off effect between complexity/fluency and accuracy. Another study conducted in Japan by Shibata (2021) created a theme-based course for Japanese university students and analyzed their writing in terms of lexical diversity over the course of an academic year. Their writing and reflective comments about the course revealed that their writing became lexically more diverse as they learned about the topic and revised their writing. In a different study by Shibata (2019), a content-based instruction course on writing for Japanese high school freshmen was conducted, and their class perceptions and writing development were analyzed. Results showed that their perceived usefulness of CBI for writing development increased steadily over the year, and the fluency almost doubled in the last writing assignment, which may have been achieved by the inclusion of reasons and examples learned in the CBI in their writing.

A study conducted in Hong Kong (Kong, 2014, 2015) showed detailed CLIL history lesson plans for late English immersion secondary school students based on a knowledge framework (Mohan, 1986) and text structure (Derewianka, 1996), and language and content teachers collaboratively planned history writing tasks. A close examination of the

instructional process over one semester confirmed the learners' better application of the target language features, such as compare/contrast and cause-effect languages, in their in-class writing.

The studies reviewed (Ikeda, 2013; Kong, 2015; Shibata, 2019, 2021; Whittaker et al., 2011) aim to assess development through the use of in-class writing tasks related to the content covered in CBI/CLIL lessens, which is not thoroughly discussed in existing CLIL studies and requires further research attention (Dalton-Puffer & Smit, 2013). To summarize the results, the positive effect of instruction on the writing whose topic is covered in CBI/CLIL can be seen in lexical complexity (Ikeda, 2013; Shibata, 2021), fluency (Ikeda, 2013; Shibata, 2019), holistic rating (Ikeda, 2013), coherence (Whittaker et al., 2011), and use of target language functions (Kong, 2015). This line of research is necessary to further investigate the potential effect of CBI/CLIL on EFL students. Moreover, while a holistic view of language development measured by a standardized English test in a CBI/CLIL setting is of crucial importance for a course evaluation, it may be fruitful to consider the specific effects of content learning in a CBI/CLIL classroom on the students' writing performance and development (García Mayo, 2015; Ortega, 2015). It is difficult to examine the effect of this instructional type without focusing on the individual task employed in actual CBI/CLIL classrooms. Certain types of writing tasks that are relevant to CBI/CLIL, such as explaining a newly learned concept (Kong, 2015) and writing with a source text (Plakans, 2015), have been closely investigated in second language writing and TBLT studies, and insights from these fields should help to more precisely understand the area of language gains in CBI/CLIL.

2.1.6 Section Summary

In this section, the theoretical background of CBI and other related instructional types

(e.g., CLIL) were explored from the interactionist (i.e., input, output, and interaction hypothesis), cognitive, and sociocultural perspectives, and existing CBI/CLIL course planning frameworks were reviewed to determine how the integration of content and language can be realized in class. A comprehensive review on CBI/CLIL research in Japan revealed that a number of practical reports have been produced while empirical investigation on students' language outcomes is still lacking. Narrowing down the focus on written language development, which is a crucial skill not only in subject learning but also for language development, CLIL studies show CLIL's favorable effects, especially on vocabulary, accuracy, and rating scores, though there are some inconsistent results. In addition, studies conducted in Japan reveal the primary evidence of writing development (vocabulary, fluency, rating); however, a close examination of writing tasks in terms of the topics and task types is valuable to view the learners' language development contextualized in content learning.

In the next section, an attempt to understand writing development in CBI/CLIL from different research fields, namely task-based language teaching and second language writing research, will be made.

2.2 CBI/CLIL Writing Tasks Investigated from Different Perspectives: TBLT and Second Language Writing Research

As described in the earlier section (section 2.1.3) on CBI/CLIL lesson planning, tasks are the essential building blocks of actual CBI/CLIL classroom implementation (Moore & Lorenzo, 2015). As Skehan (1998) pointed out, CBI is a "particular pedagogic manifestation of the task-based approach (p. 276)." Ellis (2003) also considered CBI as "a kind of task-based approach (p. 234)." However, CBI/CLIL and TBLT have rarely been discussed in relation to each other (Llinares & Dalton-Puffer, 2015). TBLT research focuses on different

task conditions (e.g., planning time and repetition) and task types (e.g., picture description task and dictogloss) to explore how these differences may affect the learners' language performance and development mostly from the cognitive perspective. Therefore, by referring to the TBLT research results that examine learner language outcomes for certain tasks related to CBI/CLIL, the possible effects of CBI/CLIL on learner language development may be more precisely explained. Specifically, the potential effects of the inclusion of content learning in language learning could be captured by the investigation of the task variables related to CBI/CLIL implementation, such as content provision variable and source text integration variable. Moreover, it will be possible to examine the pedagogic effect of CBI/CLIL from task perspectives, not solely from the amount of instructional hours as currently mentioned as a reason for language development, especially in CLIL research. The following sections will review the interface between CBI/CLIL, TBLT, and second language writing research, outcome measures widely used in the TBLT and second language writing research, and results of the empirical investigation of task variables related to CBI/CLIL contexts.

2.2.1 Background of TBLT and a Definition of Task

Since it has been shown that learners follow their own internal developmental paths and do not acquire grammatical structures as they are taught (Corder, 1967), researchers and teachers started to focus their attention on holistic language learning experiences that tasks can provide. Simultaneously, owing to the need to foster communicative competence among language learners, TBLT has gained popularity in not only ESL but also EFL countries, including Japan, where the synthetic approach to language teaching is prevalent (Butler, 2011) since TBLT provides holistic learning opportunities by employing goal-oriented and meaning-focused pedagogic tasks and enables experiential learning (Samuda & Bygate,

2008).

In the TBLT literature, several definitions of a task exist. One widely recognized

definition is by Ellis (2003, 2009), which addresses the four task features that increase the

pedagogical potential of a task.

1. The primary focus should be on "meaning" (which implies that learners should be

mainly concerned with processing the semantic and pragmatic meaning of utterances).

2. There should be some kind of "gap" (i.e., a need to convey information, express an

opinion, or infer meaning).

3. Learners should largely have to rely on their own resources (linguistic and non-

linguistic) to complete the activity.

4. There is a clearly defined outcome other than the use of language (i.e., the language

serves as a means for achieving the outcome and not as an end in its own right) (Ellis,

2009, p. 223).

Tasks with these characteristics can be naturally built in CBI/CLIL lessons (Llinares &

Dalton-Puffer, 2015), and learners may process language for task completion deeply along

with the learning of a cognitively engaging content over a certain instructional period (Pally,

2001). For instance, writing a report on a science experiment is a meaning-oriented task,

which has a communicative intent to demonstrate the understanding and report the results of

an experiment. There is a "gap" to be filled between the student's understanding and the

teacher's teaching goals. Students must rely on their own language and content resources to

complete the report, and the outcome is the completed report. This is just one example, and

there are different task types that are typical of CBI/CLIL (e.g., role plays, whole class

discussions, presentations); thus, researchers have pointed out the positive synergy of relating

CBI/CLIL and TBLT research outcomes (García Mayo, 2015; Moore & Lorenzo, 2015; Ortega, 2015).

In TBLT instruction, the following instructional cycle is proposed in the literature (Ellis, 2003; Ellis et al., 2020). The first is the task cycle. A task is implemented cyclically, typically in the pre-task (introduction to a topic and a task), during task (carrying out a task), and post-task phases (focus on language used in the task, Ellis, 2003; Willis, 1996). Considering the task cycle encourages learners to complete the task and focus on language forms that are useful in the task. Another key element in TBLT is the types of tasks, and attempts have been made to classify them. For example, Willis and Willis (2007) identified listing, sorting, classifying, matching, comparing, problem-solving, projects, and storytelling as a possible task classification. Teachers can use preexisting tasks or create their own tasks according to these task types. By referring to the task definition, task cycle, and task types relevant to their learners, practitioners can construct their own task-based lessons.

In the wealth of the empirical TBLT research, students' learning outcomes by conducting different types of task (e.g., oral or written) and the effect of the manipulation of a single task feature (e.g., planning time or no planning time) have been examined in reference to the notion of cognitive task complexity. These features are predominantly investigated with oral tasks, but recently they have begun to be explored in writing tasks. In TBLT research, experiments are conducted to observe the performance change in a laboratory or classroom setting, and data collection is completed in one-shot or in a short-term period (Ellis et al., 2020), which is the stark difference between TBLT and CBI/CLIL research. The latter typically investigates the program-level effect in a longitudinal design; therefore, the investigation of a task-level effect may add to the understanding of the program-level language development in CBI/CLIL.

2.2.2 Writing in TBLT Research

TBLT research has dealt with speaking more than writing due to cognitive SLA research's emphasis on online language processing, which is assumed to be visible through the investigation of spontaneous speech samples. A need to explore writing in TBLT research arises since it has equal importance and varied potential for language learning in TBLT (Gilabert et al., 2016). In Gilabert et al. (2016), the attributes of writing, as opposed to the speaking modality, are reviewed, such as the slower speed of production, visible output, and absence of an immediate audience. These characteristics contribute to the use of both explicit and implicit knowledge, deeper analysis and processing of language, and elimination of the time pressure to respond. Based on the differences, the potential benefits of writing over speaking can be pointed out, especially for lower-level learners. They can process language and feedback on their written language slowly and with visual support, thereby leading to a deeper processing of language and contributing to discernment and consolidation of knowledge (Shintani, 2019). Students have opportunities to draw on their explicit knowledge of language repeatedly through writing, and the chance to automatize the explicit knowledge may increase (Manchón & Williams, 2016). Writing modality is preferred not only from cognitive processing perspectives but also from affective perspectives. Cho (2018) showed that Korean students prefer written modality from an affective perspective since there is no need to embarrass themselves by speaking in L2 in front of other students without sufficient preparation. Given these benefits of writing, it is fruitful to incorporate the investigation of writing tasks in TBLT as well as in CBI/CLIL in EFL settings. Concerns have been raised about what kind of writing task is appropriate for low- to intermediate-level EFL students (McDonough & Crawford, 2020), and the need for a context-appropriate writing-to-learn task (McDonough et al., 2014) is also mentioned.

Until recently, second language writing research has prominently explored the

compositions of ESL writers in academic settings, and writing research that focuses on the writers' language development through writing has been hardly conducted, though it is gradually increasing in number (Manchón & Williams, 2016). Recently, Manchón (2011, 2020) proposed a framework of L2 writing research that includes both literacy development and language development in writing.

As Figure 2.4 shows, the learning-to-write studies cover research that explores the composition of L2 learners in ESL settings and investigates the quality of rhetoric, which has been dominant in second language writing research (Manchón, 2011). In other words, this strand of research focuses on how to write rhetorically appropriate and sophisticated compositions in L2. However, there is another way of looking at L2 writing, that is, as a tool to learn something. In other words, this aspect "can contribute to development in areas other than writing itself (Manchón, 2011, p. 3)." For the learning-to-write content dimension, writing contributes to content learning while for the learning-to-write language dimension, it contributes to the development of language knowledge and skills. The writing-to-learn aspect is as important as the learning-to-write one; however, systematic investigation has started rather recently. Reflecting on the CBI/CLIL context, writing is used for both content and language learning. Thus, as Manchón and Williams (2016) pointed out, "How language and writing develop in content-based instruction and CLIL programs is another area in need of investigation in future research on SLA-L2 writing interfaces (p. 580)."

In her latest model of writing and language learning, Manchón (2020) featured an SLA-oriented body of empirical writing research as indicated in Figure 2.5. Empirical writing-to-learn research is classified into two broad categories: writing-related variables and written corrective feedback. The first category includes task-related variables, such as complexity, modality, and repetition while the second concerns the writing conditions (e.g., individual or collaborative writing), and the third investigates the writing processes of learners.

Figure 2. 4

Classification of Second Language Writing Research by Manchón (2011, 2020)

-	The learning-to-write
-	The writing-to-learn --- Learning-to-write content
	Learning-to-write language
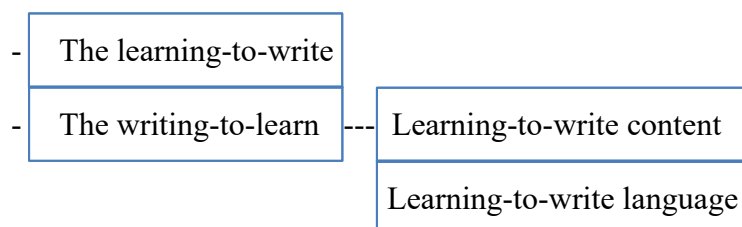
The second broad category focuses on written corrective feedback and its effects on products and processing, which has been extensively researched in L2 writing research. In this model, the writing task variables and task modality are grouped as the major factors that influence language learning experience and development as indicated in the left end in Figure 2.5. This new framework (Manchón, 2020) of writing-to-learn research favorably expands the scope of second language writing research and serves as a theoretical basis for interface studies between TBLT and second language writing research.

So far, this section has explored the link between CBI/CLIL and TBLT as well as between TBLT and second language writing research, both of which certainly share crucial instructional principles and focus, but their interfaces need further research investigation. This dissertation will explore the writing task variables (i.e., topical knowledge and source integration) that reflect the characteristics of CBI/CLIL learning (i.e., dual focus on content and language learning). The written modality was chosen to be investigated because of its benefits on language learning over spoken modality, both from cognitive and affective perspectives, especially for lower level EFL students in the current research context. In investigating their written products, this study draws on empirical research on task variables in TBLT and second language writing. In the next section, the writing production model will

be reviewed to identify the source of cognitive complexity in writing in general, and the two approaches to task complexity dominant in TBLT (Robinson, 2003; Skehan, 1998) will be presented to establish the possible type of task complexity in the writing task in CBI/CLIL. Based on the model and theoretical approaches, empirical studies on related task variables will be reviewed.

Figure 2. 5

Manchón's (2020) Overview of Empirical Research on Writing and Language Learning

Empirical research	
Writing	Written corrective feedback

TASK-RELATED ISSUES Task complexity/ Task modality/ Task repetition	WRITING CONDITIONS Individual/ Collaborative writing	WRITING PROCESSES Nature & temporal distribution/ Learning potential	EFFECTS ON PRODUCTS Feedback-related variables	PROCESSING Levels of processing/ Effects on products

Role of individual differences

2.2.3 Cognitive Task Complexity of Writing Tasks in CBI/CLIL

Writing task in CBI/CLIL, such as writing a report on a newly learned content, necessarily entails the inclusion of content knowledge in writing, thus affecting the cognitive complexity of the task. According to Kellogg's (1996) written language production model, the act of writing can be divided into the following three stages: formulation, execution, and monitoring. This model takes working memory into account, which "makes use of knowledge and experiences stored in long-term memory (p. 57)." It consists of the central executive system, which acts as a multipurpose regulatory system and deals with centralized tasks, namely reasoning and problem solving, and two slave systems, that is, visuo-spatial

sketchpad and phonological loop. The former deals with storage and processing of visual information and the latter stores and processes auditory information. In the formulation stage, which is the most cognitively demanding stage, the idea to be written is developed and translated into L2; in the execution stage, the physical act of writing is conducted which entails the least cognitive complexity (indicated as a dotted line from the central executive in Figure 2.6); and in the monitoring stage, the reading of the written words and editing is conducted. Not only does the stage proceed from left to right but also more than one stage (e.g., formulation and monitoring) can be simultaneously activated. Working memory system

Figure 2.6

Stages of Writing and Related Working Memory Functions in Kellogg (1996)
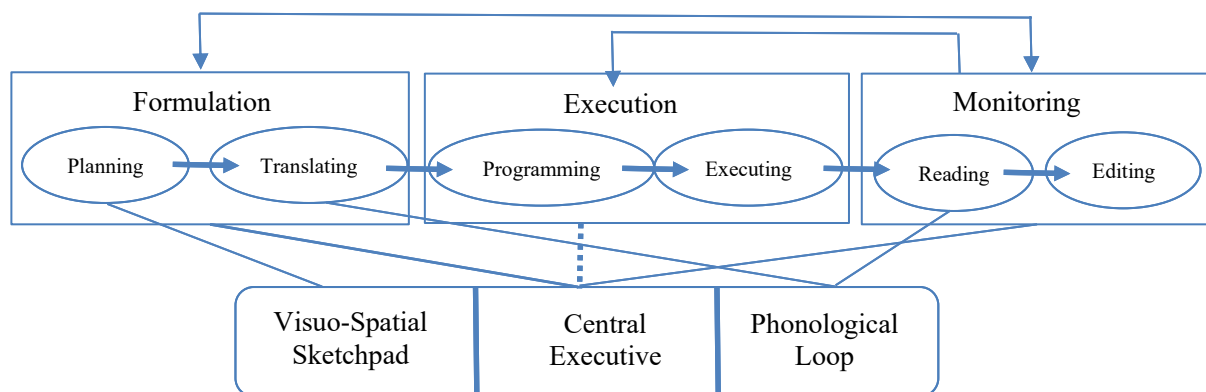
supports the three stages of the writing process differently as indicated with the lines connecting the system and the process in Figure 2.6. Central executive supervises all the three processes while the other two slave systems influence the sub processes: the visuo-spatial sketchpad supports planning while the phonological loop supports translating and reading.

    Relating the Kellogg's model and writing task in CBI/CLIL, learning about specific content in L2 and writing about it may aid the formulation stage in two ways. First, content learning can provide ideas to be used in writing, and it may aid the planning stage and lessen

the cognitive demand. In contrast, even if the content may become familiar to writers through content learning, the reasoning and organizational demand could be heightened in academic writing. Therefore, a cognitive task demand could be either aided or heightened by the inclusion of content learning in CBI/CLIL. This point will be further discussed in relation to the Cognition Hypothesis by Robinson (2005) in the next section. Secondly, the translating stage can be supported by L2 input in CBI/CLIL; writers can draw on their language resource gained by reading content materials and listening to lectures. In contrast, for L2 learners to convey an academic argument may require advanced vocabulary and unfamiliar structures that may not be covered in the input, thereby adding cognitive complexity. Therefore, as can be seen, writing in CBI/CLIL may provide ideas and language that may support better writing, but at the same time, constructing academic arguments may increase the cognitive load both mentally and linguistically.

2.2.4 Task Complexity in TBLT Research

Cognitive complexity has garnered much attention in TBLT research with the aim to investigate each task's potential for learner language development. Since teachers construct each lesson considering the task demand and task sequencing, this line of research directly relates to classroom instruction. There are two major proposals on cognitive task complexity: the Limited Attentional Capacity approach (Skehan, 1998) and the Cognition Hypothesis (Robinson, 2003, 2005). Following a series of studies conducted by Skehan and Foster (1996) and Foster and Skehan (1999), the Limited Attentional Capacity approach suggests that there is a trade-off between complexity and accuracy due to the limited capacity of our working memory, and it may be difficult to improve both by changing the level of cognitive task complexity. For example, in Foster and Skehan (1996), providing planning time improved accuracy, but providing planning time and ideas for writing heightened only complexity but

not accuracy, thus showing the tradeoff effect between accuracy and complexity. In the later discussion (Ellis et al., 2020) of the Limited Attentional Capacity approach, the existence of the tradeoff effect is less emphasized presumably due to the research results based on Robinson's Cognition Hypothesis and their own research (Tavakoli & Foster, 2008), thus showing that both complexity and accuracy could improve by taking into account the combination of task characteristics (e.g., structured/unstructured) and task conditions (e.g., with or without planning time). This approach now resembles Robinson's Cognition Hypothesis in that it also differentiates between the sources of task complexity. All in all, this approach emphasizes that the influence of limitations in working memory capacity and attention should be considered when investigating L2 learners' task performance.

Conversely, Robinson's Cognition Hypothesis suggests that it is possible to improve both complexity and accuracy by manipulating the different task complexity features and task sequencing. In other words, attention can expand to meet the cognitive demand paused by the task. In his triadic componential framework (Robinson, 2005), he summarizes three factors that influence the outcome of task performance: task conditions (interactional factors), task difficulty (learner factors), and task complexity (cognitive factors). Task conditions include participation variables, such as one-way or two-way tasks, and participant variables include pair formation according to each other's familiarity. Task difficulty deals with affective variables, such as motivation and anxiety, and ability variables, such as working memory and aptitude.

With all these three aspects combined, complex classroom learning can be investigated in a manageable way. The most extensively researched component in this framework is task complexity, which is further divided into resource-directing and resource-dispersing aspects. Increasing task complexity along resource-directing dimensions (e.g., ±elements, ±there-and-then, ±reasoning demands) necessitates the use of specific language; thus, it may expand the

learners' language resources. For example, giving reasons requires the use of logical subordinators (e.g., because, therefore, since), and introducing other people's views necessitates cognitive state verbs (e.g., argue, suggest, suppose), both of which complexify a sentence. Increasing the task complexity along resource-directing dimensions heightens attention and supports the use of complex sentences, thus enabling a double increase in accuracy and complexity and decrease in fluency. In contrast, the resource-dispersing aspects of task complexity include ±planning time, ±single task, and ±prior knowledge, and increasing these aspects of complexity strengthens the real-time access to language knowledge but does not expand the learners' language repertoire. In addition, increasing complexity along these dimensions is supposed to decrease all CAF measures.

Although the above-mentioned proposals primarily concern oral language production, attempts to apply them to written production have been made (Yoon, 2021), which will be detailed in the following section. When writing tasks in CBI/CLIL (e.g., writing about an academic topic that students are learning in the classroom) are concerned, it is predicted that the task can be both cognitively complex and less complex in reference to the Cognition Hypothesis; it may increase complexity by adding reasoning demand along resource-directing dimension, which is a typical feature in academic writing, and/or it may decrease complexity in resource-dispersing dimension by enabling learners the use of prior knowledge (i.e., content knowledge). Therefore, an investigation of the writing task that affects both resource-directing and resource-dispersing aspects and learner performance needs to be conducted. Another important aspect in both the Cognition Hypothesis and the Limited Attentional Capacity approach is that the models do not take cognitive complexity induced by a lack of language resources into consideration, which is an important issue in content learning in L2, especially in EFL situations.

In this section, the cognitive writing process and models of cognitive task complexity

were explored. It is useful to refer to these existing models to further understand the possible cognitive complexity that the CBI/CLIL writing task (i.e., the writing topic is related to the content that the students are learning about) may have and how the students' writing output may be affected by task complexity. In the next section, typical writing development measures used in TBLT, second language writing, and CBI/CLIL research will be explained, and empirical studies that examine one of the task variables highlighted in this dissertation, namely content support, will be reviewed.

2.2.5 Measures for Assessing Writing Task Performance

Previous research has utilized both linguistic (complexity, accuracy, fluency; CAF) and rating measures for writing assessment in TBLT and second language writing research (Abrams & Byrd, 2017; Yoon, 2021). Three measures included in CAF are considered to tap into different aspects of L2 performance and are suitable for holistically assessing learner language performance and development. Among CAF measures, complexity can be divided into the following two different aspects: syntactic complexity and lexical complexity. Development of these components means that learners can use the more varied and sophisticated syntactic and lexical knowledge at their disposal (Norris & Ortega, 2009). As for syntactic complexity, various types of indices have been used in previous research, such as the mean length of T-unit and the clause per T-unit ratio (for full variations, see Wolfe-Quintero et al., 1998). It has been pointed out that the selection of indices should cover the diverse aspects of syntactic complexity and reflect the learners' proficiency levels. Norris and Ortega (2009) stated that the measures can be classified into length measures, subordination measures, and phrasal measures. Length measures, such as the mean length of T-unit [i.e., T-unit being "a main clause plus all subordinate clauses and non-clausal structures attached to or embedded in it" (Hunt, 1970, p. 4)] and mean length of clause, are used to measure child

L1 development (Hunt, 1970) and thought to reflect the overall syntactic complexity. Subordination measures (e.g., the number of clauses per T-unit) are any ratio measures in which the number of clauses is divided by a given production unit and is especially reflective of measuring intermediate-level learners' performance, which is the target proficiency level of the students in this dissertation. These length (e.g., mean length of clause) and subordination (e.g., dependent clause per T-unit) measures are shown to best predict the holistic score of writing (Yang et al., 2015). Phrasal measures, such as noun clause length (i.e., grammatical metaphor in systemic functional linguistics; Halliday & Martin, 1993), is particularly suitable for measuring performance by advanced level learners since it has been confirmed that learners improve their syntax by first using coordination, then subordination, and lastly nominalization. Therefore, to fit the current study's participants (Sasayama et al., 2021) and not increase the number of dependent variables by employing redundant measures, the mean length of T-unit and number of dependent clauses per T-unit were adopted. For the measurement of syntactic complexity, automated software, such as the L2 Syntactic Complexity Analyzer (Lu, 2010), has been utilized in research, whose reliability has been confirmed in comparison to human coding (Kessler et al., 2022).

Along with syntactic complexity, lexical complexity is also considered as one of the valid measures to assess writing performance and development. Lexical complexity is often investigated from two aspects: diversity (variety of words) and sophistication (less frequently used words identified in the reference-corpus), the increase of which is considered as a sign of development. The measure of lexical diversity is obtained from automated tools, namely Coh-Metrix (McNamara et al., 2010) and Text Inspector by the name of Measure of Textual Lexical Diversity (MTLD), which is a type-token ratio-based measure controlled for text length and considered to be robust even for short texts (McCarthy & Jarvis, 2010). Lexical sophistication, in contrast, can be obtained from the Lexical Complexity Analyzer (Lu, 2012),

which is calculated by dividing the total number of sophisticated word types (i.e., beyond the most frequent 2,000 words in the British National Corpus) by the total word types indicated by the name of Lexical sophistication II in the system.

As for accuracy, various measures, including counting the number of error-free units, number of errors, and holistic rating (Polio & Shea, 2014), have been used in L2 writing research, and all of them have shown the same levels of reliability (Polio & Shea, 2014). However, to make more subtle inferences about the accuracy improvement of early-stage writers, Foster and Wigglesworth (2016) proposed the weighted clause ratio, whose validity is shown to be especially pronounced for measuring lower-proficiency learners' performance (Evans et al., 2014). To calculate a weighted clause ratio, each written text is first segmented by clause, and each clause is then assigned a rating of the score 1 (error-free), 0.8 (error[s] that did not impede comprehensibility), 0.5 (error[s] that impeded reading but not comprehensibility), or 0.1 (error[s] that impeded comprehensibility; Foster & Wigglesworth, 2016). Scores for each clause are added and divided by the total number of clauses to obtain the final accuracy score. Lastly, in writing research, fluency is often operationalized as the number of words produced within a certain time limit (Abrams & Byrd, 2017; Johnson, 2017; Ong & Zhang, 2010) and shown to be an important predictive variable for the holistic performance of writing of different genres (e.g., argumentative, narrative, expository, and expo-argumentative; Yang, 2014).

Most studies use CAF for performance evaluation, but how each CAF dimension develops over time as learners gain proficiency has not yet been fully investigated (Vercellotti, 2017). However, it is generally agreed that learners are able to produce higher CAF as they gain proficiency (Michel, 2017), but their performance and development are greatly influenced by task features (e.g., topic, genre, and planning), instructional time duration (e.g., short-term or longitudinal design), context (e.g., ESL or EFL), and proficiency

level. For example, fluency tends to increase by the function of time with some phase transitions (Baba & Nitta, 2014), and it is considered as one of the valid measures to assess the learners' proficiency levels. Accuracy development is somewhat unobservable for adult advanced learners (Polat & Kim, 2014; Polio & Shea, 2014), but it is evident for secondary school students with lower proficiency (Lahuerta, 2020). As for complexity, the picture becomes less straightforward because it has two dimensions (i.e., syntactic and lexical complexity), and various measures exist. As for beginner level learners, Kim (2021) investigated the developmental pattern of complexity in a writing corpus collected 10 times over a period of 10 weeks from CEFR A2 level Korean learners of English in various English classes and confirmed no change in syntactic complexity except for only a slight decrease in the ratio of complex nominal. In addition, lexical sophistication increased but lexical diversity did not. As suggested in previous research (Verspoor et al., 2012), the lower level learners may first expand the lexicon and subsequently increase the syntactic complexity. Conversely, using advanced ESL learners' writing corpus, Bulté and Housen (2014) showed that syntactic complexity becomes higher as students received as short as four months of instruction in the ESL context, but the lexical complexity (i.e., lexical diversity and sophistication) did not change. Mazgutova and Kormos' (2015) writing data from ESL university learners after a month of English for Academic Purposes (EAP) instruction also showed noun phrase complexity improvement for the intermediate-level students. Contrary to Bulté and Housen (2014), the participants in this study (both intermediate and upper-intermediate students) also improved lexical diversity. From these results, we should take into account the fact that a change in complexity is dependent on various factors, such as the environment (ESL or EFL), type of instruction that the students received, learner proficiency, and writing task topic and genre. With these general CAF characteristics in mind, the effect of the change in certain task elements on learners' writing performance measured by CAF

will be discussed in section 2.2.6.

Together with the above CAF measures, which measure the linguistic aspects of writing, recent studies have also increasingly begun to investigate the communicability of written work (Michel, 2017; Pallotti, 2009) using a rubric (e.g., functional adequacy scale, Kuiken & Vedder, 2017; a holistic rating scale, Abrams & Byrd, 2017). This is owing to the concern that better linguistic performance does not necessarily imply functionally more appropriate L2 performance, and such successful task fulfillment should be measured separately from CAF dimensions (Pallotti, 2009). One rubric that aims to measure the functional aspects of writing is Kuiken and Vedder's (2017) functional adequacy scale, which was created based on Grice's (1975) maxims of conversation and has been shown to reliably evaluate the four dimensions of functional adequacy. It is a six-point rating scale and includes the following four dimensions: content (number and type of information units and their relevance independent from the specific task requirements); task requirements (fulfillment level of the specific instructions and requirements); comprehensibility (effort required for readers to understand the text); and coherence and cohesion (number of coherence breaks, use of cohesive ties, conjunctions and repetitions). In CBI/CLIL, language is learned in a contextualized manner (i.e., embedded in sustained content learning), and students learn language functions by exposing themselves to contextualized language input. This is one of the strong points of CBI/CLIL since language items are often taught individually and separately from their contexts in a language classroom, leaving learners unable to use the items in context. Therefore, the use of the functional adequacy scale is necessary and suitable for the assessment of CBI/CLIL learning outcomes. Recently, other studies have examined the scale itself and used it for their research (Nuzzo & Bove, 2020; Suzuki, 2022; Xu, 2021), and its reliability is confirmed for L2 writing samples. Based on the above discussion on the importance of considering both the linguistic and functional aspects of L2 assessment, several

studies in TBLT and second language writing research have analyzed both the aspects of writing (Abrams & Byrd, 2017; Bulté & Housen, 2014; Vandommele et al., 2017) to more fully understand the area of language development and performance change caused by independent variables (e.g., time, task type, task manipulation).

This section reviewed the commonly used performance measures (i.e., CAF and functional adequacy scale), how to obtain their scores, and what developmental trend can be generally observed based on previous research. In the next section, empirical studies on the task variables related to CBI/CLIL, which are topic familiarity and integrated writing, will be reviewed and possible written language development in CBI/CLIL in EFL contexts will be discussed.

2.2.6 Empirical Studies of Written Task Variables

In this section, two writing task variables (topic familiarity and source-based writing) that are in need of investigation not only in the field of writing (Weigle, 2002) but also in CBI/CLIL will be analyzed in reference to TBLT and second language writing research. The unique point of CBI/CLIL is that language learning is contextualized in content learning (Brinton & Snow, 2017). In other words, gaining familiarity with certain content or topics will affect the areas of the learners' language performance and development. In addition, it is generally acknowledged that writing tasks supported by reading materials, background knowledge, or experiences may elicit better writing quality (Kroll & Reid, 1994). Topic familiarity is often operationalized as one research variable unrelated to the actual content learned in an instructional context, and the process of gaining content knowledge is not included in the study. Therefore, it is necessary to investigate how this variable influences learner performance when incorporated into CBI/CLIL. Especially in the EFL context where L2 input is typically limited and language is taught item-by-item, the learners' language

outcomes in a holistic language learning environment possibly created by CBI/CLIL is also worth exploring.

Regarding source-based writing, it is most relevant in CBI/CLIL (Plakans, 2015) as learners need to show their understanding of the readings covered in CBI/CLIL and their thoughts on the content in writing. Although this task type is widely used in proficiency tests, such as Test of English as a Foreign Language Internet-based test (TOEFL iBT), the use of integrated writing in the classroom and investigation of its language learning potential have been less focused on in research (Cho & Kim, 2021), especially in Japanese contexts (Kowata, 2018; 2019). Moreover, integrated writing can be operationalized in different ways in a CBI/CLIL classroom (e.g., citation required or not required), and the influence of this task variable on performance has been less explored.

The section below first reviews the studies on content knowledge and subsequently on integrated writing, both of which are from second language writing and TBLT research to obtain insights for conducting these tasks in CBI/CLIL.

2.2.6.1 Content Knowledge as a Task Variable in L2 Writing Performance

Some researchers recognize content knowledge as one of the essential constructs of language proficiency (Banerjee, 2019; He & Shi, 2012), which is the core feature to be learned in CBI/CLIL. Previous TBLT and writing research has also investigated this variable mainly based on Robinson's Cognition Hypothesis. Robinson (2005) situates prior knowledge (i.e., topic familiarity) as one of the variables in the resource-dispersing dimensions and proposes that the task with a lack of topic familiarity is more complex and negatively affects all aspects of CAF. However, increasing complexity along with this variable is thought to strengthen the links among the learners' existing language repertoires and positively support the fulfilment of the given task. As for a CBI/CLIL writing task, it can

58

be predicted that learning of content knowledge in the class may positively affect all three dimensions of CAF, but the investigation of this task variable with the data obtained from the actual CBI/CLIL classroom is scarce. Since topic familiarity studies need to be classified according to their various operationalizations (Yang & Kim, 2020), empirical research in TBLT and second language writing that investigate topic familiarity are classified into three types in this study: providing ideas to be used in the writing, topic familiarity derived from having personal experience or subject knowledge, and gaining topical knowledge through instruction. Possible impact of existing content knowledge on language performance and development will be examined.

2.2.6.1.1 Providing Ideas to be Used in the Writing

To control the level of content knowledge among the participants and investigate the effect of lowered cognitive load enabled by content support, some studies provide ideas to be used in the essay. As for the effect of this task manipulation on linguistic aspects, Révész et al. (2017), as part of a larger study, found that the provision of ideas contributed to higher use of K2-level than K1-level vocabulary and increased the syntactic complexity (mean length of T-unit) for advanced-level ESL learners with various L1 backgrounds. As part of a larger study of Singaporean university students, Ong and Zhang (2010) also showed that the group that was given a topic, ideas, and macro-structure and the group that was given a topic and ideas outperformed the group that was only provided with a topic in lexical diversity but not in fluency of the rewritten text. Yoon (2021) reviewed high intermediate-level ESL learners' writing in terms of two genres (narrative and argumentative) and task complexity ($\pm$ idea support). Results showed that the provision of ideas positively affected one of the syntactic complexity measures (i.e., the number of clauses per T-unit) but negatively affected lexical sophistication. Lastly, Jung (2020) investigated Korean EFL university learners' integrated

reading-writing task performance, and the findings indicated that the + content support condition elicited more K2-level words and higher number of noun phrase modifiers (i.e., higher syntactic complexity).

The findings of the above four studies partly support Robinson's (2005) hypothesis for resource-dispersing aspects and show that scaffolding the content to be written by providing ideas may increase the syntactic complexity (e.g., mean length of T-unit, the number of clauses per T-unit, the number of noun modifiers) and lexical diversity and sophistication (except for Yoon, 2021). Fluency, however, seems to be unaffected according to Ong and Zhang's (2010) research, and accuracy and rating of writing quality were not examined in any of the studies.

Not only linguistic aspects but also writing process and task perception have been evaluated in some research. Ong (2014) showed that students in the topic, ideas, and organization group engaged less in idea generation and organization in pre-writing and during-writing planning, and their writing performance in such a condition was better than that of the students who were only given a writing topic (Ong & Zhang, 2013). Furthermore, Jung (2020) showed that the lack of content support caused advanced level Korean univerisity learners to pause frequently and longer during the writing and insert more words during revision. Therefore, providing students with ideas to be used may free up attentional resources and elicit higher writing performance. As for task perception, using a questionnaire, Yoon's (2021) study showed that the + idea support condition significantly lowered high intermediate-level ESL learners' perceived mental effort and difficulty level for the writing task. In summary, content provision raises syntactic and lexical complexity, lowers mental effort for planning and writing, and improves writing quality.

2.2.6.1.2 Topic Familiarity Operationalized as Personal Experience and Subject Knowledge

Another way to operationalize content knowledge is to focus on the learners' current knowledge, operationalized as $\pm$ personal experience and $\pm$ subject knowledge. McDonough and Crawford (2020) operationalized topic familiarity as $\pm$ personal experience (e.g., writing a proposal for planning a welcome event for incoming first-year university students as a + personal experience since students had just experienced such an event, and writing an application for a 90-day trip by ship with foreign students as a – personal experience since none of them had this experience). Results from beginner-level (i.e., CEFR A1–2) Thai university students showed that + personal experience writing was rated highly in terms of task accomplishment, content, grammar, and vocabulary and high in subordination and target verb use, but there was no difference in accuracy. Authors emphasize the facilitative effect of having topic familiarity for not only advanced learners, on which most of the topic familiarity research has focused, but also for novice learners since it may enable the learners to utilize their language resources effectively. Similarly, Yoon (2017) found that topics with greater relevance to the college students' personal life (e.g., part-time job vs. banning smoking) earned higher syntactic complexity (e.g., length and phrasal measures) and lexical sophistication by using EFL Chinese college students' argumentative writing corpus. Somewhat contradictory, Yang and Kim (2020) compared + familiar topics (Internet's benefits and problems for university students) and – familiar topics (Internet's benefits and problems for people in underdeveloped areas) in writing and found that lexical diversity and sophistication were higher for + familiar topics, but there was no difference in terms of syntactic complexity, accuracy, and fluency. The results for syntactic complexity conflict with McDonough and Crawford (2020) and Yoon (2017), and the authors speculate that this may be due to how topic familiarity is operationalized in this study, where both prompts are

on an everyday life subject matter. Concerns for topic familiarity operationalization are also raised in Ruiz-Funes (2015), which investigated the writing performance of advanced and intermediate-level learners of Spanish in an American university, and found that – task complexity writing (operationalized by the combination of various factors, such as + topic familiarity, – reasoning demand, and + discourse genre familiarity) tended to have lower syntactic complexity and a relatively higher level of accuracy and fluency. However, no statistical tests were performed due to the small sample size, and the results are difficult to interpret since various task complexity factors are included in this study. A subsequent meta-analysis by Johnson (2017) on topic familiarity included three studies (Ruiz-Funes, 2015; Salimi & Fatollahnejad, 2012; Yang, 2014) and confirmed a medium positive effect of having topic familiarity on the lexical complexity ($d = 0.50$).

From the above studies, the topics that the researchers judged to be more familiar tended to elicit better analytic rating scores and higher CAF. However, there are some inconsistent results among the studies reviewed above in terms of the syntactic complexity and accuracy. In addition, topic familiarity is operationalized differently in different studies, and this lack of a framework may have caused inconsistency in the findings. Moreover, the researchers' intuitive operationalization of topic familiarity may not necessarily reflect the students' perception of topic familiarity. Therefore, the following three studies incorporated the students' rating of a set of prompts from the existing test batteries, and the two prompts with higher or lower topic familiarity were compared. He and Shi (2012) compared ESL writers' writing on a familiar topic (what to study at a university) and an unfamiliar topic that requires specific knowledge (federal politics in Canada), and the results indicated that the former obtained higher rating scores for content, organization, and language for all three (basic, intermediate, and advanced) proficiency levels. Post-hoc interviews revealed that the difficulty arose from the lack of content knowledge, lack of vocabulary, and lack of

confidence to write about the topic. As a conceptual replication of He and Shi's (2012) research, Kessler et al. (2022) obtained contradictory results from ESL students' corpus; less familiar topic (electronic cigarettes are/are not safer than regular cigarrettes) yielded higher lexical sophistication and accuracy and a more familiar topic (cellphone use should/should not be banned while driving) yielded higher lexical diversity and syntactic complexity with greater subordination. There was no difference in terms of fluency, and the regression analysis showed that holistic scores are influenced by the degree of topic familiarity. Making the picture more complicated, Abdi Tabari et al. (2021) compared familiar and unfamiliar topic writing (prompts not shown in the paper) by ESL graduate learners of English with advanced proficiency and confirmed higher syntactic complexity (mean length of clause, possessives per nominal, and prepositions per clause) for the familiar topic, but there was no difference in lexical diversity.

From the above three studies (Abdi Tabari et al, 2021; He & Shi, 2012; Kessler et al., 2022), + topic familiarity seems to elicit higher syntactic complexity and holistic rating scores, but disconfirming results are obtained for lexical complexity. As the prompts rated in these studies are all in English test batteries aiming to equally assess the learners' writing proficiency, it is speculated that there may not have been a substantial difference between the two prompts even though they were rated as being more or less familiar. As He and Shi's (2012) minus topic familiarity prompt required specific knowledge of politics and student interviews confirmed higher task complexity of this prompt due to this requirement, the kind of knowledge required in writing (e.g., general world knowledge or subject knowledge) is also a crucial factor in investigating topic familiarity.

One early study that considers subject knowledge as a task variable is by Tedick (1990), which investigated ESL graduate students' writing in two prompt types (one that requires students' field-specific knowledge and a general topic) and revealed that the field-specific

topic writing was better in terms of the holistic rating, syntactic complexity (mean length of T-unit), and fluency (overall length). As with the previously discussed studies, this research also shows that the students' topic familiarity operationalized as + subject knowledge supported their improved writing performance. In addition, Tedick (1990) raised a critical point revealed by the close examination of the accuracy scores of three proficiency groups (beginning, intermediate, and advanced), which had the following tendency:

> If, on the one hand, L2 writers have a limited amount of linguistic knowledge in the L2, their familiarity with the subject matter of a writing task does not provide them with the linguistic knowledge also required to produce quality writing. If, on the other hand, L2 writers are capable of producing syntactically complex utterances with fewer errors, their familiarity with the subject matter allows them to demonstrate this capability. (Tedick, 1990, p. 136)

This observation clearly emphasizes that it is vital for teachers to pay attention to not only the topic (content) familiarity but also the language necessary to write about it when selecting writing tasks for learners, especially for low- to intermediate-level learners. For L1 students, having topic familiarity nearly equals having language knowledge to demonstrate that topical knowledge, but it is not necessarily true for L2 students.

Similar to the above discussion, one aspect that needs to be considered in Robinson's Cognition Hypothesis is whether prior knowledge includes only the concept of that knowledge or also the L2 repertoire to express that knowledge. Especially high school and adult L2 learners have a solid knowledge base in L1 but do not know how to express it in L2. In this case, + prior content knowledge does not necessarily increase the linguistic performance, which is supposed to improve along the resource-dispersing dimension. If

lessons are based on an intellectually diminished content to learn a language with, it may lead to demotivation for L2 learning. Thus, it is crucial to further investigate prior knowledge (i.e., content knowledge) variables on language performance and learning. It may provide vital information for lesson planning so that teachers can be made accountable for intellectually interesting and linguistically appropriate L2 teaching.

2.2.6.1.3 Topical and Language Knowledge Gained through Instruction

In relation to the point raised above, some studies recognize the necessity to include both the process of activating and gaining content and language knowledge within the study (Yang & Kim, 2020, p. 101). The following studies attempted to do this and scrutinized its effect on written output. For example, Vandommele et al. (2017) focused on beginner-level immigrant learners of Dutch in two writing classes: in-school and out-of-school multimodal writing classes in which students learned about Antwerp's tourist information in L2 and created a multimodal website about it. Furthermore, there was a control group with no instruction on the topic. Two types of writing samples (persuasive writing related to the content covered and one new narrative piece) were collected from the three groups after the instruction. Results for the persuasive writing showed that students in the two experimental groups outperformed the control group in communicative effectiveness, content, syntactic complexity, lexical diversity, and text length but not in accuracy. As for the new narrative topic, no major quality change was observed between the pre- and post-test except for the mean length of T-unit for an in-school group with a small effect, suggesting that writing skills practiced with the tourist topic may not have transferred to the new topic. Thus, gaining content knowledge in L2 may positively affect the students' L2 written output when they write on the same topic. Especially for EFL students with negligible L2 input, language input in the classroom may be the sole base for L2 written language development; therefore, the

topic with which EFL students practice writing affects their written language development. Additionally, the findings suggest that writing improvement in the CBI/CLIL topic may not easily transfer to a new topic.

Similarly, Abrams (2019) focused on intermediate-level learners of German and compared content-supported writing based on reading materials and movies and personal topic writing based on the students' own world knowledge. Results confirmed that all CAF dimensions were better in the content-supported writing, and fluency and lexical richness yielded robust results in particular. The author concludes that content learning offered a lexical springboard and facilitated ideational borrowing, thereby contributing to higher fluency.

From a different perspective, Xu (2021) investigated the effect of multimodal writing project intervention, which provided similar learning experiences as CBI/CLIL in that students gain content knowledge from various media throughout the project (e.g., visuals, videos, and readings) and create their own multimodal video presentation. The author compared this group with a control group that underwent a traditional writing class that covered the same topics. Comparison of the pre- and post-test individual writing by Chinese university students revealed that the mean length of T-unit, text length, and three of the four functional adequacy scales (task requirements, content, comprehensibility) were significantly higher for the multimodal writing group than the traditional writing group, but there was no difference in lexical complexity, accuracy, and coherence/cohesion. Therefore, the availability of multimodal resources in instruction may influence the students' writing performance. Importantly, the functional adequacy scale, which is also used in this thesis, may be sensitive to assessing the holistic effect of instruction (multimodal writing projects).

In summary, as in the other two types of topic familiarity studies, it can be said that learning about a topic of writing through instruction, in other words, gaining topical

knowledge, positively affects various aspects of writing, such as complexity and fluency, as well as the holistic aspects of writing rated by human raters; however, accuracy showed contradictory results in the studies reviewed. To investigate the impact of integrated learning of content and language, the inclusion of instruction in the research design may provide ecologically valid information when considering the influence of CBI/CLIL instruction; nevertheless, such studies are limited in number.

So far in this section, an investigation of the possible effects of content learning on writing performance in CBI/CLIL classrooms was attempted by reviewing the literature on topic familiarity in TBLT and second language writing research. Literature was classified into three categories to better understand the influence of different operationalization of topic familiarity: providing ideas to be used in the writing, topic familiarity derived from having personal experience or subject knowledge, and gaining topical knowledge through instruction. In summary, the above-mentioned studies generally suggest that having or gaining topical knowledge has positive effects on subsequent writing, generally for both linguistic aspects (i.e., complexity and fluency) and rating. However, some contradictory results have been obtained for CAF possibly due to differences in task designs and learner proficiency, and the small effect of topic familiarity on accuracy seems to be rather strong (six out of eight studies investigating accuracy had null or negative effects). As discussed in Tedick (1990), having topic familiarity and being able to accurately utilize that knowledge in L2 writing may not necessarily mean the same thing for L2 learners (see also Chapple & Curtis, 2000, p. 428).

The first study of this thesis attempts to investigate a topic familiarity difference (i.e., a power generation topic covered in the CBI class and a new general topic) and the effect of topic learning on Japanese high school students' writing in a CBI classroom and determine which aspects of secondary school EFL learners' writing may be influenced. Another writing

variable that plays an important role in CBI/CLIL is the integration of information from reading into writing (Plakans, 2015). In CBI/CLIL, writing can be used by learners to consolidate the knowledge gained from reading materials and lectures and expand their own thoughts on the topic. In the next section, this second task variable's linguistic characteristics, learners' task perception, and pedagogical potential of this task type for language learning will be reviewed and discussed.

2.2.6.2 Integrated Writing and CBI/CLIL

CBI/CLIL provides holistic learning experiences (Garner & Borg, 2005) with the use of all four language skills in an integrated manner in a meaningful context. Thus, the integration of different skills occurs naturally for learning in CBI/CLIL (Plakans, 2015). In particular, reading and writing can be effectively learned through CBI/CLIL since content works as a solid foundation for synergistically using these two skills, and positive learning outcomes have been empirically confirmed in CBI contexts (Kong, 2015; Pally, 2001; Pessoa et al., 2007; Plakans, 2015; Shih, 1986). As also discussed in literacy studies, reading and writing abilities develop hand in hand (Grabe & Zhang, 2016) and should be taught in an integrated manner (Plakans, 2015). From the TBLT perspective, the integrated writing task is recognized as a versatile task used in everyday life, but research has only recently begun to take root in the TBLT field. Thus, when and how this task type can be introduced to L2 learners still needs empirical investigation (Abrams, 2019). In this section, literature on integrated writing will be analyzed from L2 writing research to understand the uniqueness of this task type in contrast to independent writing and learners' task perception. Moreover, the interface of integrated writing and TBLT task complexity research will be reviewed to better understand the possible influence of the use of such tasks in CBI/CLIL.

2.2.6.2.1 Integrated Writing in Second Language Writing Research

Integrated writing is an important task type in the academic field. Majority of the research on this task type is based on L2 literacy studies (Grabe & Zhang, 2016) and is investigated with ESL adult learners who were studying in university settings (Knoch & Sitajalabhorn, 2013; Grabe & Zhang, 2013). Simultaneously, the CBI/CLIL approach is also adopted for literacy teaching (Grabe & Stoller, 1997); therefore, the compatibility of this task type in CBI/CLIL is high. However, research investigation of this task type contextualized in CBI/CLIL is scarce, and the majority of integrated writing research is conducted in the language assessment field using data from large-scale language tests, such as Test of English as a Foreign Language (TOEFL) (Chan et al., 2015; Knoch & Sitajalabhorn, 2013).

From these backgrounds, integrated writing is recognized as an indispensable skill to survive in mainstream subject classes for ESL learners; however, the characteristics of this task type can and should be examined as one task type in CBI/CLIL and from the language learning perspective. Such research that will provide insights into language learning, specifically the comparison of integrated and independent writing and their differential effects on performance and learning, will be explored in this section. As explained above, integrated writing can be conducted based on the content dealt with in CBI/CLIL, and the comparison with independent writing tasks may highlight the pedagogical potential of this task and thus, the possible effects of CBI/CLIL.

2.2.6.2.2 Definition of Integrated Writing

Various types of integrated writing tasks have been created and used according to each instructional context and purpose: summarization (McDonough et al., 2014), term paper/research paper writing, summarizing the source text and adding a personal reflection (Cho & Kim, 2021), and continuation task (continuing a story with the writer's own ideas

after reading the first part of it, Peng et al., 2020). Knoch and Sitajalabhorn (2013) extracted

six crucial factors for a task to be identified as an integrated writing task:

> Integrated writing tasks are tasks in which test takers are presented with one or more
>
> language-rich source texts and are required to produce written compositions that
>
> require (1) mining the source texts for ideas, (2) selecting ideas, (3) synthesizing ideas
>
> from one or more source texts, (4) transforming the language used in the input, (5)
>
> organizing ideas and (6) using stylistic conventions such as connecting ideas and
>
> acknowledging sources (p. 306).

According to this definition, an integrated writing task needs to have text(s) with a certain

amount of words and requires writers to "integrate" the information from the text(s) into their

own writing using appropriate writing conventions. In contrast, Plakans (2012) identified

another type of integrated writing wherein the reading materials are provided but writers are

not required to use the information in their writing. Both types of integrated writing are

conducted in classrooms, and whether this task manipulation elicits different task

performance and learning experience warrants investigation, especially with EFL learners

because of the dominance of ESL contexts in this research field. In study 2 of this

dissertation, two types of integrated writing tasks [based on definitions by Knoch and

Sitajalabhorn (2013) and Plakans (2012)] were conducted in the CBI writing class, and the

performance and perception differences were assessed.

2.2.6.2.3 Comparison of Integrated Writing and Non-integrated Writing

    To illuminate the differential writing process and performance, past research has

examined the differences in integrated and independent writing performance. For the two

tasks used in this dissertation, both provide learners with reading materials before writing, but the without integration task does not require the use of a source text while the with integration task requires source integration. For both task types, students are required to present their own solutions and suggestions. Thus, the results of independent and integrated task comparison studies may inform the current study.

As for lexical complexity, integrated writing task elicits higher lexical diversity (Cumming et al., 2005; Frear & Bitchener, 2015), sophistication (Kyle & Crossley, 2016; Shin & Kim, 2014), and use of longer words (Cumming et al., 2005). Integration of source texts may enable learners to utilize more diverse and sophisticated lexical items since they can borrow words from the text, and this opportunity to use these words that may be beyond their lexical knowledge may trigger the learning of new words. In addition, a careful reading of a source text in search of usable words may induce deeper processing of the text, thus leading to the learning of language items. As for syntactic complexity, the results have been mixed; independent writing gained higher syntactic complexity in Frear and Bitchener (2015) and Shin and Kim (2014) while there was no difference in the mean length of T-unit in Cumming et al. (2005). As Frear and Bitchener (2015) have proposed, according to the writers' proficiency levels, the cognitive demand of integrated writing (specifically, reasoning demand, and the number of elements) may or may not facilitate the use of more syntactically complex structures. Results obtained for accuracy show no difference according to the task types (Cumming et al., 2005). In addition, cohesion measured by Coh-Metrix was higher for the integrated writing task in Shin and Kim's (2014) study. As for the rating results, Cumming et al. (2005) confirmed better argument structure and quality of claims for independent writing. In contrast, Cho (2019) found that the reading-based prompts condition elicited superior writing quality rated with a rating scale in comparison to independent writing prompts.

As summarized above, integrating reading materials into writing seems to have a positive impact on lexical aspects, but no unanimous results are obtained for syntactic complexity and rating possibly due to the variety in learner proficiency and language use contexts [e.g., intermediate-level ESL learners in Frear and Bitchener (2015), ESL TOEFL test takers' corpus in Cumming et al. (2005), Korean low- to intermediate-level university learners in Cho (2019), and Korean high school students in Shin and Kim (2014)] as well as the actual tasks used (e.g., letter writing, TOEFL integrated writing prompts, expository writing, and argumentative writing). Overall, it can be said that the integration of reading materials elicits different task performances and opportunities for language learning by imposing differential levels of cognitive task complexity. Integration of content learning into language learning is the main characteristic of CBI/CLIL, and the impact of this instructional characteristic may partly be measured using an integrated writing task, which is a commonly used and cognitively challenging task in CBI/CLIL. This task type needs investigation not only from the assessment perspective but also from the language learning perspective.

To further determine why performance differences arise for independent and integrated tasks, previous studies have explored the writing process, source use purposes, and task perception in qualitative manners. Plakans (2008) compared the process of integrated and independent writing via the think-aloud method and interviews. Findings revealed that discourse synthesis was more apparent in the integrated writing task while greater initial planning characterized independent writing. From this result, it can be speculated that the source of cognitive complexity may differ across the two task types (i.e., discourse synthesis or overall planning). Michel et al. (2020) also investigated the learners' cognitive processes in TOEFL integrated and independent writing tasks by using keystroke logging, eye-tracking, and stimulated recall and found that students spent less time on planning for integrated writing due to the support they gained from the reading and listening materials. In general,

72

these process studies highlight the different underlying sources of cognitive complexity in independent and integrated writing, and thus, it is important to understand the task difference from a cognitive perspective to better utilize the tasks in language classrooms.

Previous studies have also explored source use purposes from the writers' perspectives. Plakans and Gebrils' (2012) mixed-method study suggests that the source text provided support for opinion formation, idea generation, language sophistication, and text organization. Similarly, through interviews, Leki and Carson (1997) found that the source texts "serve as a general model, stimulate thinking, and supply many of the resources the students complained of lacking without source texts, such as vocabulary, sentence structures, writing style, organizational patterns, ideas, and information" (p. 51). From these observations, it can be emphasized that the integrated tasks' potential as a language learning task is high: it can provide opportunities for deeper processing of reading materials, expand the variety of vocabulary and sentence structures, and enhance organization. Studies confirm that there is a relationship between the understanding of source texts and the writing quality (Payant et al., 2019), especially for lower-level learners (Plakans & Gebril, 2012), and facilitating text understanding by providing guiding questions led to better writing in terms of the readability, accuracy, and coverage of content (Proske & Kapp, 2013). Therefore, when using an integrated task for lower-level learners, guiding text comprehension may enhance not only the writing product but also language learning through writing.

Lastly, the learners' task perception toward integrated writing has been compared through interviews. In Plakans' (2008) study, nine out of 10 students preferred reading-to-write tasks since they provided them with ideas and points for argumentation in writing. It implies that integrated writing has potential as a pedagogic task as it is in line with the learners' task preferences. Similarly, Neumann et al. (2019) investigated the benefits and challenges of source-based writing. One perceived benefit was that a source text can be a

resource for writing if students understand the text while the challenges included the difficulty in understanding the text, choosing appropriate information, integrating the information into writing and paraphrasing.

Qualitative investigation of the integrated tasks reviewed above highlights distinct sources of cognitive complexity between independent and integrated writing (i.e., overall planning and discourse synthesis, respectively) and a supportive role that source texts can play in task completion, which may be conducive to language learning. Although integrated writing is identified as one of the most important authentic task types in the field of literacy studies and language assessment, the use of integrated writing as a pedagogic task in language teaching contexts has only begun to be examined (Abrams, 2019). Several TBLT studies have explored this task type in relation to cognitive task complexity, which will be discussed in the next section.

2.2.6.2.4 Integrated Writing in the TBLT Research

Integrated writing investigated in the TBLT research is still in its infancy, and the affordances of integrated writing tasks for language learning as well as how to employ this task in language classrooms need further research (Abrams, 2019; Cho & Kim, 2021; Golparvar & Rashidi, 2021; Jung, 2020). For example, there is an attempt to examine integrated writing as a task complexity variable and determine whether the use of source text heightens or lowers cognitive task complexity according to Robinson's (2003) and Skehan's (1998) models as detailed in section 2.2.4.

For example, Abrams (2019) found that source-based writing showed higher complexity, fluency, and accuracy than personal topic writing written by university learners of German. In addition, multimodal inputs (e.g., films and reading materials) seemed to have supported content borrowing and reduced the learners' cognitive burden, leading to enhanced

CAF. Discussing these results in reference to Robinson's (2005) model, the author argues that the availability of source text in writing provides ideas and language to be used (creating + prior knowledge condition), thus it may decrease complexity in the resource-dispersing dimension. In contrast, integration of source text information into one's writing adds further elements to the writing task (e.g., constructing opinion by taking source text information into consideration, thus, – few elements and justifying the way in which the particular information is used in the discussion, thus – no reasoning demands), leading to higher complexity in the resource-directing dimension. Therefore, it is assumed that integration of reading into a writing task has possibilities to both heighten and lessen cognitive task complexity, and how this task aspect influences EFL learners' writing performance and task perception warrants further research. As for the linguistic results of Abrams' (2019) study, all CAF measures were higher for the integrated task, which is in line with Skehan (1998) in that – task complexity condition positively affected all CAF dimensions. It is also in line with Robinson (2005) when the integrated task is investigated in the resource-dispersing dimension (e.g., + prior knowledge condition elicits better CAF) and resource-directing dimension (e.g., – few elements and – no reasoning conditions elicit better complexity and accuracy).

Another research that explores TBLT and integrated writing interface suggests a possible effect of having source reading materials on vocabulary learning. Jung (2020) investigated the role of a task complexity variable (i.e., idea provision) in integrated writing and found that learners in + task complexity (i.e., no ideas provided for use) were able to guess which pseudowords were used in the source text (there were a total of 16 pseudowords in the recognition test, of which 8 were used in the source text, and the other 8 new pseudowords were not used in the source text) more than the – task complexity group (i.e., ideas provided). Therefore, higher task complexity induced by no idea provision may have enhanced the involvement in the reading material, leading to more word recognition. In other

words, it can be said that integrated writing which requires careful reading of a source text may have a positive influence on vocabulary learning. This discussion is also suggested by Robinson (2011), claiming that higher task complexity elicits higher attention to task input and results in the learning of language in the input.

Although still limited in number, integrated writing's task complexity has begun to be investigated in relation to TBLT task complexity research as in the above studies (Abrams, 2019; Jung, 2020). When planning a CBI/CLIL course, it is often included as a learning activity (Plakans, 2012), and knowing about the manner in which the task's complexity affects the learners' writing performance and their learning experience provides valuable information to teachers when they plan a CBI/CLIL lesson. In Japan, integrated writing has also been identified as an important task type (Hosogoshi et al., 2016) and also evinced as is in the new Courses of Study for high school English classes in Japan (MEXT, 2018), but its application and investigation in the classroom has been scarce. CBI/CLIL can provide an ideal context for the practice of this task type.

To summarize the integrated writing section, this task is said to be highly compatible with CBI/CLIL classrooms (Plakans, 2015) since establishing one's opinion in relation to external sources is an important skill in content learning. As for task performance, it has been shown that integrated writing elicits different performances from independent writing; more specifically, higher lexical complexity for integrated writing and mixed results were found for syntactic complexity and ratings. The qualitative investigation also highlights the differential effects of the tasks' cognitive complexity and language learning opportunities. However, only a limited number of previous studies (Abrams, 2019; Frear & Bitchener, 2015; Jung, 2020) has investigated the pedagogical effect of the source-based writing tasks and their outcomes in classroom settings even if it is widely used in real classrooms, such as in CBI/CLIL, because it has been mainly used as a language assessment tool (Cumming et

al., 2005; Kyle & Crossley, 2016; Plakans, 2008; Plakans & Gebril, 2012). Since the type of source-based writing and its applicability to low–intermediate students need further investigation (Abrams & Byrd, 2017), this study examines two types of integrated writing performance (i.e., one that requires the use of source text information and the other that does not) elicited in a CBI/CLIL writing classroom in Japan and the learning opportunities they afford by referring to student interviews. Moreover, few studies have explored this task variable in Japanese classroom contexts (Kowata, 2018, 2019), and the investigation of the various types of integrated writing tasks with different writer populations is necessary (Knoch & Sitajalabhorn, 2013; Plakans, 2010).

2.3 Statement of Problems

The literature review section began with an overview of the historical emergence of CBI/CLIL, their theoretical supports, and several lesson planning models proposed in previous research. Although the majority of CBI/CLIL research conducted in Japan is in the form of a practical report, existing empirical studies have shown their positive impact on learner language in terms of vocabulary, grammar, and writing. Conversely, writing is one of the vital skills in content learning for consolidating knowledge and expressing opinions related to the content. Moreover, the potential of writing for language learning has begun to be recognized (Manchón, 2020; Gilabert et al., 2016) for its slower and self-paced production and visibility of output. These characteristics of writing provide learners with time for production and attending to their own output. Therefore, the focus was narrowed down to writing performance in CBI/CLIL settings. Such studies were classified into input time-unmatched (i.e., a total amount of L2 input in the classroom is not equal or not indicated for the CLIL and non-CLIL groups), input time-matched (i.e., a total amount of L2 input in the classroom is nearly equal for the CLIL and non-CLIL groups), and writing topic aligned to

CBI/CLIL content categories. For the input time-unmatched studies, inconsistent results on CAF and rating have been found. Input time-matched studies showed mixed effects on lexical complexity, accuracy, and rating and no effects on syntactic complexity and fluency. When the writing topic is aligned to the CBI/CLIL content, positive effects were seen on fluency, rating, use of target language functions, and lexical complexity. These results are a reflection of the whole CBI/CLIL instruction, and a closer look at the task characteristics typical of CBI/CLIL, such as the third condition, that is, topic aligned to CBI/CLIL content, and its cognitive complexity, may aid our understanding of written language performance and development in CBI/CLIL.

The second part of the literature review summarized past studies on the two task variables most relevant to CBI/CLIL (topic familiarity and integrated writing) in TBLT and second language writing research. Results suggest that having content knowledge supports better writing performance in terms of CAF and ratings though there are some inconsistent results for syntactic complexity and accuracy. Moreover, research on integrated writing shows that it may support the use and learning of vocabulary. Although the findings for syntactic complexity and rating were mixed, a qualitative analysis of the task process and task perception revealed that it has different sources of cognitive complexity from independent writing, and writers perceived supportive roles that a source text can play in their writing. These TBLT and second language writing research results clearly indicate that supporting content knowledge in language classrooms has positive effects on the learners' written output, and this support can be naturally made available in the CBI/CLIL classroom.

Although the application of CBI/CLIL instruction in Japanese classrooms is increasing, the empirical investigation of language performance is still insufficient. Therefore, this dissertation aims to examine Japanese high school and college students' writing performance in terms of CAF and rating on two CBI-related task variables, which are topic familiarity and

integrated writing, contextualized in two theme-based units developed for the study. The overall research question is as follows: *How does theme-based instruction affect EFL students' writing performance and development and their class perceptions?* This dissertation tries to explore the uniqueness of CBI/CLIL, which is the presence of meaningful content learning in language classrooms, by narrowing down the focus on task features. For the research design and measurement of language performance, TBLT and second language writing research studies were consulted, thereby aiming to understand the learning in CBI/CLIL from the findings of multiple research fields.

   To answer this overall research question, the following two studies were conducted. Study 1 examines the performance difference of $\pm$ topic familiarity writing tasks by Japanese high school students. + topic familiarity condition was made by choosing a general topic that can be handled by the current participants by referring to high school learners' world knowledge while – topic familiarity condition utilized a power generation topic, which was yet to be dealt with in L1 subject classes. Furthermore, a three-month CBI unit on the power generation topic was created, and the learners' pre- and post-writings were compared to examine the effect of CBI instruction on the performance of familiar and unfamiliar writing topics. More specifically, the following research questions are addressed:

Study 1-1: *Is there any difference in Japanese high school students' writing performance according to the topic difference (content-specific writing topic and general writing topic)?*

Study 1-2: *How do the students perform in the content-specific writing topic and general writing topic differently before and after the theme-based instruction?*

From the literature reviewed, it can be predicted that the general topic writing (+ topic

familiarity) may elicit higher performance in terms of CAF and rating, but syntactic complexity and accuracy need to be observed carefully since mixed results have been found for these two measures (e.g., Abrams, 2019; Vandommele et al., 2017). As for the second question, the extant CBI/CLIL literature suggests that the instruction has a positive impact on lexical complexity, fluency, and rating if the writing topic is aligned with the CBI/CLIL subject content (Ikeda, 2013; Kong, 2015). To interpret the effect on non-content related writing topic (in this study, the general topic), the majority of CLIL writing studies can be referred to since it employs a topic unrelated to the subject covered (Roquet & Pérez-Vidal, 2017; Ruiz de Zarobe, 2010). The results of these studies comparing CLIL and non-CLIL groups were mixed for CAF and rating. In addition to the past research, the current study can add insights in that it employs both content-specific and general topics to investigate the effect of topic difference and instruction. Existing CBI/CLIL studies only utilize either one of these types of prompts. Moreover, the learner population (i.e., Japanese high school learners) is less focused on in the empirical CBI/CLIL, TBLT, and writing research.

To further examine the learners' class perceptions, their reflective comments after class were thematically coded, and the tone of each statement (e.g., positive, negative, neutral) was also coded. The third research question is as follows:

Study 1-3: *Which aspects of CBI lessons were reflected in the students' reflection sheets and what are their tones (e.g., positive, negative, neutral)?*

The second study's focus is on the reading/writing interface since reading is one of the main content learning tools in CBI/CLIL. Applicability of this task type in CBI/CLIL is mentioned in the past literature (Plakans, 2015); however, empirical investigation in a classroom setting is scarce since it has mainly been examined in the language assessment

literature. Moreover, it has only started to be researched as a language learning task in classroom settings (Abrams, 2019), and scant research has been conducted in language classrooms in Japan (Kowata, 2018, 2019). Therefore, study 2 examines Japanese college students' performance on two types of integrated writing (i.e., source-text integration required or not) to highlight the effect of source integration on performance and possible language learning. As in study 1, the writing was assessed in terms of CAF and rating. Participants were also interviewed to clarify their task perceptions about different sources of task complexity in these two types of tasks and their usefulness in language learning. Additionally, their thoughts on the combined learning of content (e.g., environmental problems highlighted in the sustainable development goals, SDGs) and language were elicited through interviews to evaluate the entire CBI writing course created for the study. The research questions are as follows:

Study 2-1: *Is there any difference between Japanese college students' writing performances on the two types of integrated writing (i.e., source-text integration required or not)?*

Study 2-2: *How did the students perceive the difference between the two writing task types?*

Study 2-3: *How did the students perceive the combined learning of content (SDGs) and English writing?*

As for the first research question in study 2, it is predicted that the performance of source-integration required writing elicits higher lexical complexity (Cumming et al., 2005; Kyle & Crossley, 2016) owing to the support from the source text. For the other measures (syntactic complexity, accuracy, and rating), the results have been mixed (Cumming et al., 2005; Shin & Kim, 2014); therefore, further empirical evidence is needed, especially from EFL students since the major data sources have been ESL university students. For the second research

question, previous studies have identified the different sources of cognitive complexity for integrated and independent writing, such as discourse synthesis and planning for the organization (Neumann et al., 2019; Plakans, 2008) through interviews, questionnaires, and think-aloud method. The current study will add the Japanese college learners' perspectives on the differences between the two types of integrated writing tasks and their perceived usefulness of these tasks for language learning by conducting interviews. Lastly, the third question explores the overall impression of the CBI writing course created for the current study to gain insights for further improvement on CBI/CLIL implementation in Japan.

Chapter 3. Study 1: Japanese High School Students' Writing Performances According to a Topic Difference Variable and their Change after Theme-Based Instruction

3.1 Introduction

The first variable for the CBI/CLIL writing task, namely topic difference, was examined by comparing two writing performances of different topic familiarity (i.e., a power generation topic and a general topic) by Japanese high school students. Moreover, a theme-based unit on the power generation topic was created, and the performances of the two writing topics before and after the instruction were compared to reveal the effect of this instruction. To supplement the understanding of the students' experiences in the current unit, their reflective comments were also analyzed. This study examines the following three research questions.

Study 1-1: *Is there any difference in Japanese high school students' writing performance according to the topic difference (the content-specific writing topic and general writing topic)?*

Study 1-2: *How do the students perform in the content-specific writing topic and general writing topic differently before and after the theme-based instruction?*

Study 1-3: *What aspects of CBI lessons were reflected in the students' reflection and how are their tones (e.g., positive, negative, neutral)?*

The following sections detail the research method with examples of lesson plans and materials used for the unit. Results are discussed in reference to not only CBI/CLIL research but also TBLT and second language writing research.

3.2 Method

3.2.1 Participants and the Research Site

The participants were 36 students (male = 28, female = 8) aged 17 to 18 years majoring in electrical and computer engineering at a five-year science college in rural Japan. At this college, the curriculum for the first three years covers most of the subjects in the Courses of Study for high schools with an increasing amount of specialized science classes added as they proceed to the third year. Most of the classes become specialized classes when they reach the fourth year. Data for this study was obtained in one of the three regular English classes, which was 90 minutes long and conducted twice a week in the semester. Along with the reading-based instruction using a government-authorized textbook, the present unit was taught as part of the regular English class by the current author as a teacher. For the present CBI unit, 11 30-minute lessons were created and conducted over a nine-week period (i.e., 1–2 times a week). Participants' English proficiency was limited (i.e., M_{TOEIC} = 315, SD = 69, MIN = 225, MAX = 485) although this is the typical level of Japanese high school students. More than 60% of the third-year Japanese high school students are found to be at the Common European Framework of Reference (CEFR) A1 level (MEXT, 2018) for reading and listening and more than 80% are at A1 level for speaking and writing. The participants' English usage outside the class was homogeneously limited according to the background questionnaire. Three students responded that they had taken cram school English classes at the time of data collection, and none had studied abroad or traveled to English-speaking countries. Students were informed of the research purpose, their authority to refuse to participate in the study, and the fact that there was no relationship between their grade and their withdrawal from the study. With the explanation and consent form, all students agreed to participate in the study and to the use of writing data.

3.2.2 Writing Prompts

To answer the first research question, two kinds of prompts with different topic familiarity were created. The first type required students to compare and the contrast the benefits and drawbacks of different power generation methods, which was classified as a less familiar topic. Students need to know the characteristics of each power generation method to develop an argument in the writing, and the interview with science teachers before creating the unit confirmed that this topic had not been covered in the specialized science classes. In contrast, the second kind asked them to discuss the advantages and disadvantages of two counterpart ideas which are presumed to be easier for high school students to form an opinion on based on their everyday school life experiences (e.g., video call or email, e-textbooks or paper textbooks, a human or robot clerk).

To check the students' perceptions of these two types of writing prompts, a questionnaire survey related to the difficulty of the prompts (see Table 3.1 and Appendix A) was conducted after the intervention. They were asked to choose one out of three options (e.g., 1. Power generation topic was difficult, 2. General topic was difficult and 3. Both were difficult) and write down the reasons for their choice. A total of 32 out of 36 participants answered the questionnaire survey due to absences.

Table 3. 1

Students' Choice of Writing Prompts in terms of Difficulty and the Number of Students

Topics	# of students (n = 32)
Power generation topic was difficult	19
General topic was difficult	3
Both were difficult	10

The survey responses revealed that the power generation topic was perceived to be more challenging. Nineteen students (60%) considered it to be more cognitively complex, and their

reasons for the choice revealed that the difficulty mainly came from unfamiliar vocabulary (n = 8), a lack of knowledge (n = 5), and difficulty in explaining the content (n = 2). In contrast, only three students (9%) considered the general topic challenging due to lack of interest in the general topic, lack of ideas and vocabulary, and difficulty in reasoning. Ten students (31%) said both were difficult and specified it as the reason for their choice (n = 5), and three students said that they lacked English proficiency (n = 3). Of the ten students, two answered that they did not remember their writing experience. All the prompts are shown in Table 3.2, they were also indicated in Japanese to ensure the students' understanding. Actual writing handout used for data collection is included in Appendix B. Prompts were piloted with six students at the same institution having similar English proficiency with the current participants, and their understanding of the writing prompts, time it took for writing, and writing length and quality were checked through interviews, observation, and reading of the finished writing. After the piloting, it was decided that a Japanese translation of the writing instruction should be included to ensure the students' understanding of it, and the minimum word limit (i.e., 50 words) was indicated in the instruction so that the writing length would be sufficient to use the automated text analysis tool.

Table 3.2

Writing Prompts for General and Power Generation Topics

Power generation topics
1．You are responsible for writing a report on two ways of power generation for your engineering class. Choose two from the list below. Discuss their advantages and disadvantages, and explain which one will be more important for Japan in the near future and why. You must write 50 words or more. ● thermal power generation（火力発電） ● nuclear power generation（原子力発電） ● hydroelectric power generation（水力発電） ● solar power generation（太陽光発電） ● wind power generation（風力発電） ● other（その他）：

２．A newspaper company is hosting a high school student essay contest. The topic is "promising methods of power generation in the next 30 years." All 3rd year Kosen students will participate in this contest. Choose two methods from the list below. Discuss their advantages and disadvantages and explain which you think is more promising and why. You must write 50 words or more.
● thermal power generation（火力発電）
● nuclear power generation（原子力発電）
● hydroelectric power generation（水力発電）
● solar power generation（太陽光発電）
● wind power generation（風力発電）
● other（その他）　:

３．Suppose you work for an electric power company. You are going to submit a report to the city council to suggest which type of power station should be promoted: (a) a solar power plant, or (b) a thermal power plant. Discuss their advantages and disadvantages, and explain which you think should be promoted and why. You must write 50 words or more.

General topics

１．There are many ways to communicate with people. If you communicate with your friends in the U.S. who only speak English, which means would you use, (a) video call or (b) email? Discuss their advantages and disadvantages and explain which means would be better for you and why. You must write 50 words or more.

２．Kosen has started to offer two options for textbooks, (a) e-textbooks in a tablet or (b) paper textbooks. You can choose either of the two. Discuss the advantages and disadvantages of them, and explain the reason for your choice. You must write 50 words or more.

３．These days, some shops and hotels are replacing human workers (e.g., shop clerks and receptionists) with robots. Discuss the advantages and disadvantages of (a) human workers and (b) robots, and explain which you would like to receive service from and why. You must write 50 words or more.

The students wrote two out of three prompts for each kind of prompt, writing four times in total. There were three prompts, but they wrote on two of them because the original research planned to obtain pre-, post-, and delayed post-data, and three prompts were created to counterbalance the prompt effect. However, due to the classroom circumstance, it was not possible to conduct the third data collection. They were given 30 minutes to write without the use of a dictionary, and the prompts were counterbalanced between the learners and the time given. Average word count was 80 words, and the standard deviation was 30 words.

3.2.3 CBI Unit Creation

3.2.3.1 Content and Language Objectives and the Outline of the Unit

In creating the whole unit plan and each lesson, the CBI lesson planning guides reviewed in section 2.1.3 (Echevarría et al., 2017; Dalton-Puffer, 2013; Lyster, 2011; Mohan, 1986; Stoller & Grabe, 2017) were referred to in light of the current EFL college context. The intervention classes were taught by the author, who had six years of English teaching experience with secondary school students at the time. It was her third year teaching the same group of students. For study 1, a theme-based unit with the following content objectives was devised: to compare/contrast the benefits and drawbacks of different power generation methods currently used in Japan and discuss their optimal future combination for use in 2030. The content reflected the participants' major, but the perceived difficulty of this topic was higher according to the questionnaire, and it had not been formally taught in L1 subject classes. In contrast, the language objective was to be able to understand and use the language to compare and contrast, which was necessary for learning the characteristics of each power generation method. Specifically, the language covered included comparative/superlative forms, adverbs (e.g., *however, similarly*), and conjunctions (e.g., *although*). These objectives were shared among the learners and the instructor at the beginning of each class (Echevarría et al., 2017).

3.2.3.2 Materials Development

Teaching materials included various multimodal inputs (e.g., texts, pictures, videos), following the SIOP (Echevarría et al., 2017) feature 4 "Supplementary materials used to a high degree, making the lesson clear and meaningful" and Snow (2014). For more specific examples, the following were used: reading texts adapted from reports by the Ministry of

88

Economy, Trade, and Industry; websites of electricity companies; a video clip about renewable energy; and handouts aimed at facilitating the understanding of these materials. Furthermore, materials were created with the help of a content teacher in the same school who recommended the reference materials and answered clarification questions about the content.

After deciding the content and language objectives and gathering various materials, the possible tasks, learning questions, and language to be used were planned according to Mohan's (1986) knowledge framework to make the practical action situation and theoretical background knowledge clear. Based on this overall plan, each lesson plan with the content/language objectives, learning tasks in reference to Mohan's knowledge structure (see Table 3.3), and detailed materials were created.

As shown in Table 3.3, the content and language focus were set for each lesson. The lesson first introduces the different energy sources used in Japan and then moves on to an examination of the change in the use of energy sources over the past several decades. After learning about the change and its reasons, the students focused on learning about the pros and cons of each power generation method in relation to Japan's energy policy, that is, S+3E (Safety, Energy security, Economic efficiency, Environment). Lastly, based on the information learned, they predicted the optimum breakdown of energy sources with reasons. The content was organized to proceed from lower- to higher-order thinking in Mohan's (1986) knowledge framework (i.e., from practical action knowledge to theoretical background knowledge, from description/classification to choice/evaluation). Language targets were derived from the language necessary for content learning. For this unit, the major language function was to compare/contrast the pros and cons of different power generation methods; therefore, such language was presented in the input in a manner that attracts the learners' attention (e.g., bold fonts, underlining) and can be practiced in a contextualized task.

As for the tasks, the content learning (picture–definition matching, classifying, information gaps) and language learning tasks (enhanced input, fill-in-the-blanks, guided sentence writing) were conducted with an aim to utilize all four language skills. Students cooperatively completed the tasks mainly in pairs followed by a whole-class discussion to share what had been discussed in pair work.

Table 3.3

Unit Plan with Content/Language Objectives, Tasks, and Corresponding Knowledge Framework in Suzuki (2019)

Unit	Content	Language	Tasks	Mohan's knowledge framework
1	Students identify and label different energy sources according to their characteristics	Vocabulary (e.g., hydropower, electricity, coal, petroleum) Relative pronouns used to define each energy source	Brainstorm what energy sources there are Picture–description matching task	Description Classification
2–4	Students predict the current breakdown of power generation methods used They explain and give reasons for the changes of power generation methods used at three time points (1973, 2010, 2016)	Vocabulary (e.g., increase, decrease, remain the same, account for) Comparative connectives (e.g., compared to, although, in contrast)	Label a pie chart and give reasons Read a reading material and write a description of some noticeable changes in the use of power generation methods over time Identify the compare/contrast expressions in the input	Description Sequence Principles (cause/effect, explaining)
5–7	Students explain Japan's energy policy (S+3E) They describe the pros and cons of thermal power and nuclear power generation	Comparative connectives (e.g., but, however, while) Comparative and superlative forms	Read and fill in a table describing the characteristics of power generation methods Reorganize the information in the table Write a passage explaining the pros and cons using a sentence	Description

			frame	
8–9	Students describe the pros and cons of other kinds of renewable energy	Comparative connectives (e.g., but, however, while) Comparative and superlative forms	Watch a video clip and take notes on the benefits and downsides of various renewable energy Listen and fill in the listening script with compare/contrast expressions	Description
10–11	Students read graphs to understand the current trend in the use of renewable energy sources Students predict the 2030 energy use	Vocabulary covered and expressions for compare/contrast	Predict and draw a pie chart for 2030 energy source breakdown Summarize the pros and cons of renewable energy in a table Write one solution to achieve the 2030 energy goal	Choice Analyze

For the lesson sequence, the primary focus was firstly on the content, and it switched to language when the language of compare/contrast played an important role in content learning and was practiced in a contextualized manner. In addition, for any language problems that the researcher noticed that the students had, scaffolding was provided (paraphrasing the content, using easier words, and providing the Japanese meaning). In this manner, providing opportunities to focus the learners' attention on the language features is empirically recommended by past CBI/CLIL studies (Pena & Pladevall-ballester, 2020; Lyster, 2011). After the language-focused activity, the focus returned to content learning as suggested by Lyster (2011). The language for instruction was English, and the materials were also created in English. However, Japanese was occasionally used to clarify novel vocabulary and content concepts to ensure the students' understanding and avoid negative reactions for not understanding the content. Students mainly used Japanese in their pair discussions; however, they were encouraged to use English with scaffolding from the researcher in the whole-class discussion.

To illustrate the lesson sequences and teaching procedures, the content/language goals and tasks of each lesson are shown in Appendix C. One illustrative task for student pairs is the information gap task conducted in lesson 3 in which one student compared two pie charts showing the breakdown of the power generation methods used in Japan in 1973 and 2010 while the other was shown similar charts for 2010 and 2016. In addition, each of them was given different reading materials related to the charts and asked to point out the noticeable changes between these two time points and possible reasons for the changes referring to both the charts and a reading. They summarized the changes and reasons in writing, and some useful expressions to compare/contrast the change (e.g., increase, decrease, more, less) were highlighted in the input and orally introduced by the researcher. Finally, they shared what they had written orally in pairs and in a whole-class discussion. After the meaning-focused content learning task, the students' focus was drawn to the compare/contrast language using a fill-in-the-blank passage explaining the same pie charts, and different language functions (e.g., language for comparison, reasoning, and addition) were highlighted. As in the above example, in-class short writing tasks which took around 10–15 minutes to write were conducted to support the cognitive processes for content learning and language internalization, modification, and consolidation (Shintani, 2019). Since the participants' language proficiency was limited, writing was strategically used to ensure opportunities for producing output in a scaffolded manner (i.e., slower pace of production and visual scaffolding provided by written words).

3.2.4 Data Collection Schedule and Procedure

Over the 12-week research period, pre-writing for the two topics was conducted in two separate class times in the first week in their normal classroom. Students were given 30 minutes to write without a dictionary. From weeks 2–10, the CBI unit with 11 lessons was

conducted. After the unit was completed, post-writing was conducted in weeks 11 and 12.

For both the pre- and post-writing task, the general topic was conducted first followed by the

content-specific topic. The overall research schedule is shown in Figure 3.1.

Figure 3.1

Summary of the Research Procedure

Pre-writing 1 & 2 (Week 1)
CBI unit power generation topic + compare/contrast expressions 30 min. * 11 times (Weeks 2–10)
Post-writing 1 (Week 11)
Post-writing 2 (Week 12)

3.2.5 Writing Data Analysis

Hand-written writing was transformed into digital data. Analysis was conducted in

terms of the linguistic aspects (i.e., CAF) and functional aspects (i.e., rating using a rubric).

Regarding the linguistic aspects, automated text analysis tools, such as the L2 Syntactic

Complexity Analyzer (Lu, 2010) and Coh-Metrix (McNamara et al., 2010), were used to

obtain the mean length of T-unit (MLT) and number of dependent clauses per T-unit (DC/T)

for syntactic complexity and textual lexical diversity (MTLD) for lexical complexity. A

dependent clause in DC/T is a finite adverbial, adjective, or nominal clause. MTLD (Measure

of Textual Lexical Diversity) is the mean length of sequential word strings that maintain a

given threshold of type-token ratio in a text (McCarthy & Jarvis, 2010). Both tools are widely

used in applied linguistics research with L2 writing data. As discussed in section 2.2.5,

considering the learners' proficiency levels (i.e., beginner to intermediate-level) and the need

for reduction of the dependent variables, these three complexity measures were selected (Norris & Ortega, 2009).

As for accuracy, the weighted clause ratio (Foster & Wigglesworth, 2016) was used to capture subtle changes in the writing of the current students (Evans et al., 2014). The researcher and a second rater with an M.A. in English education rated 17% of all the data, and reliability using Cronbach's alpha ($\alpha = 0.91$ for the power generation topic, $\alpha = 0.72$ for the general topic) was obtained, which was shown to be acceptable (Larson-Hall, 2010). Due to the research schedule, the evaluation of the general topic was conducted several months after the power generation topic assessment, which may have negatively affected the reliability. The remaining data was rated by the researcher.

Lastly for fluency, the total number of words produced within 30 minutes was used as an indicator of writing fluency according to, for example, Johnson et al. (2012) and Yang and Kim (2020).

For the rating, the functional adequacy rating scale (Kuiken & Vedder, 2017) was used. This scale has the four dimensions: comprehensibility, content, task requirements, and coherence/cohesion. Among these, task requirements were operationalized for both kinds of tasks as the degree to which students argue the benefits and drawbacks, adding reasons for their choice from either of the options. The researcher and a second rater who held an M.A. and had more than two decades of teaching experience rated all the data. For the rating session, the researcher first explained the construct of the rubric face-to-face using the benchmark writings and practiced rating collaboratively. Then, individual ratings were conducted with sample essays obtained from a similar participant population followed by discussions to clarify any concerns. Intra-class correlation coefficients were used to check the inter-rater reliability, and moderate to good reliability was obtained (Koo & Li, 2016, see Table 3.4). The average scores of the two raters were used (Abrams & Byrd, 2017).

Table 3.4

Intra-class Correlation Coefficients for the Functional Adequacy Rating

Components	Intra-class correlation coefficients
Comprehensibility	0.72*
Content	0.85*
Task requirements	0.87*
Coherence and cohesion	0.79*

*$p < .001$

For the statistical analysis, mixed-design repeated measures ANOVA (within-group variable: time, between-group variable: topics, dependent variables: CAF measures and functional adequacy ratings) was performed. Critical alpha level was set to 0.05. Bonferroni correction was used to control the overall alpha level ($\alpha = 0.05/9 = 0.0056$), but the results approaching the significant level were also taken into consideration to make weak but possible inferences about the results. Due to absences at the time of data collection and not meeting the required word limit of 50 words (L2 Lexical Complexity Analyzer's minimum word limit), four students' data was not included, thus leaving a total of 32 participants.

3.2.6 Analysis of Reflection Sheets

Reflection sheets were utilized as supplemental data to understand the students' perceptions of the lessons. They voluntarily wrote reflective comments in Japanese seven times after the class. Due to time constraints, comments were not obtained in all the eleven classes. To elicit the students' main focus of the class from their point of view, the reflection sheet did not have any specific questions but simply asked them to mark their understanding level and write their comments about the class. In order not to take excessive time on reflective comment writing, the comment sheet was kept short and simple. Students wrote

comments 52.5% of the time (136 comments out of all 259 possible comments). The average amount of comments was 22.6 characters (SD=16.2, min=3 [1 word], max=81 [3 sentences]).

For coding, the researcher read through all the comments repeatedly, which were parsed basically at the sentence level, except one topic that was discussed over several sentences, and inductively created codes (Braun & Clarke, 2006). Eleven codes (1. Target content of the unit, 2. Target language of the unit, 3. Reading, 4. Listening, 5. Speaking, 6. Writing, 7. Vocabulary, 8. Grammar, 9. Holistic comments about the class and class delivery, 10. Understanding level, 11. Others) were obtained, and the comments were coded as positive, negative, or neutral. Each parsed comment was classified into one or more categories according to the content. For example, one sentence containing both positive and negative comments was coded as both positive and negative. The code neutral was used for comments that did not connote positive or negative remarks, such as stating a fact. The researcher and a second coder, who was a Ph.D. student in applied linguistics, coded all the data, and the agreement for the first coding was 86% for positive/negative/neutral and 75% for the eleven codes. The main source of disagreement was ambiguous comments, and it was resolved through discussion and by referring to the class syllabus and the students' other comments. The excerpt of the comment coding is indicated in Appendix D.

3.3 Results

3.3.1 Results of the Writing Topic Difference

Before performing the statistical analysis, normality was checked using Shapiro-Wilk tests, and 39% of the data sets showed violation. Values of kurtosis and skewness and results of the Shapiro-Wilk tests for each of the datasets are shown in Appendix E. Although as Larson-Hall (2010) pointed out, statistical power may be lower due to the violation of the assumption; it does not imply that the results are incorrect. Homogeneity of variance was

checked with Mauchly's sphericity test, and it was always met as the repeated measures had only two levels (i.e., pre- and post-tests). Thus, the analysis was run, and the results were interpreted with caution. Effect size benchmarks for ANOVA were the following: small η^2 = .01, medium η^2 = .06, large η^2 = .14 (Mizumoto & Takeuchi, 2010). For the post hoc analysis, the following criteria were used: small d = .40, medium d = .70, and large d = 1.00 (Plonsky & Oswald, 2014). Statistical results with marginal significance were also included to make weak but possible interpretations.

Repeated measures ANOVA was performed to answer the first research question "Is there any difference in Japanese high school students' writing performance according to the topic difference (the content-specific topic and general topic)?" Descriptive statistics for each measure are presented in Tables 3.5 and 3.6. To illustrate each data visually, box plots are shown in Figure 3.2.

Table 3.5

Means and Standard Deviations for the CAF Measures

	Power generation topic		General topic	
	pre-test	post-test	pre-test	post-test
MLT	10.30 (2.97)	9.92 (1.48)	8.84 (1.22)	8.86 (1.92)
DC/T	0.28 (0.23)	0.35 (0.35)	0.35 (0.27)	0.32 (0.23)
MLTD	34.63 (11.48)	35.19 (12.81)	46.55 (18.11)	49.16 (18.29)
Accuracy	0.63 (0.12)	0.67 (0.15)	0.69 (0.12)	0.69 (0.12)
No. of words	86.53 (32.25)	83.63 (26.78)	76.50 (32.34)	74.44 (28.60)

Note. MLT (mean length of T-unit), DC/T (number of dependent clauses per one T-unit), MLTD (measure of textual lexical diversity)

Table 3.6

Means and Standard Deviations for the Functional Adequacy Scale Scores

	Power generation topic		General topic	
	pre-test	post-test	pre-test	post-test
Comprehensibility	3.20 (0.81)	3.44 (0.73)	3.19 (0.84)	3.48 (1.04)
Content	3.06 (1.22)	2.88 (0.92)	2.83 (1.25)	2.69 (1.02)
Task requirements	3.73 (1.23)	4.31 (0.85)	3.33 (1.22)	3.63 (1.23)
Coherence and cohesion	2.61 (0.70)	2.59 (1.03)	2.00 (0.70)	2.33 (0.94)

Figure 3.2

Boxplots of Each Measure
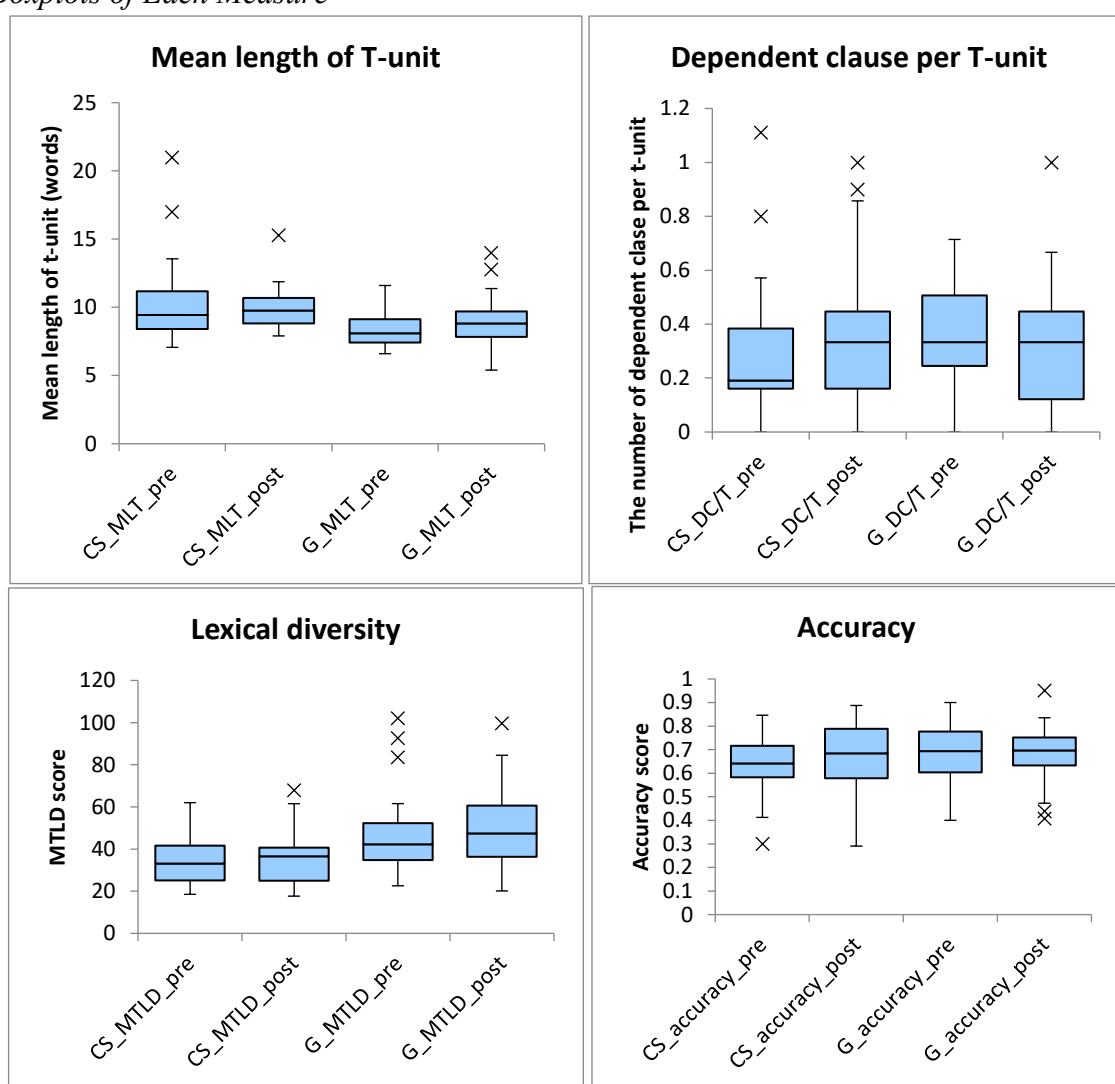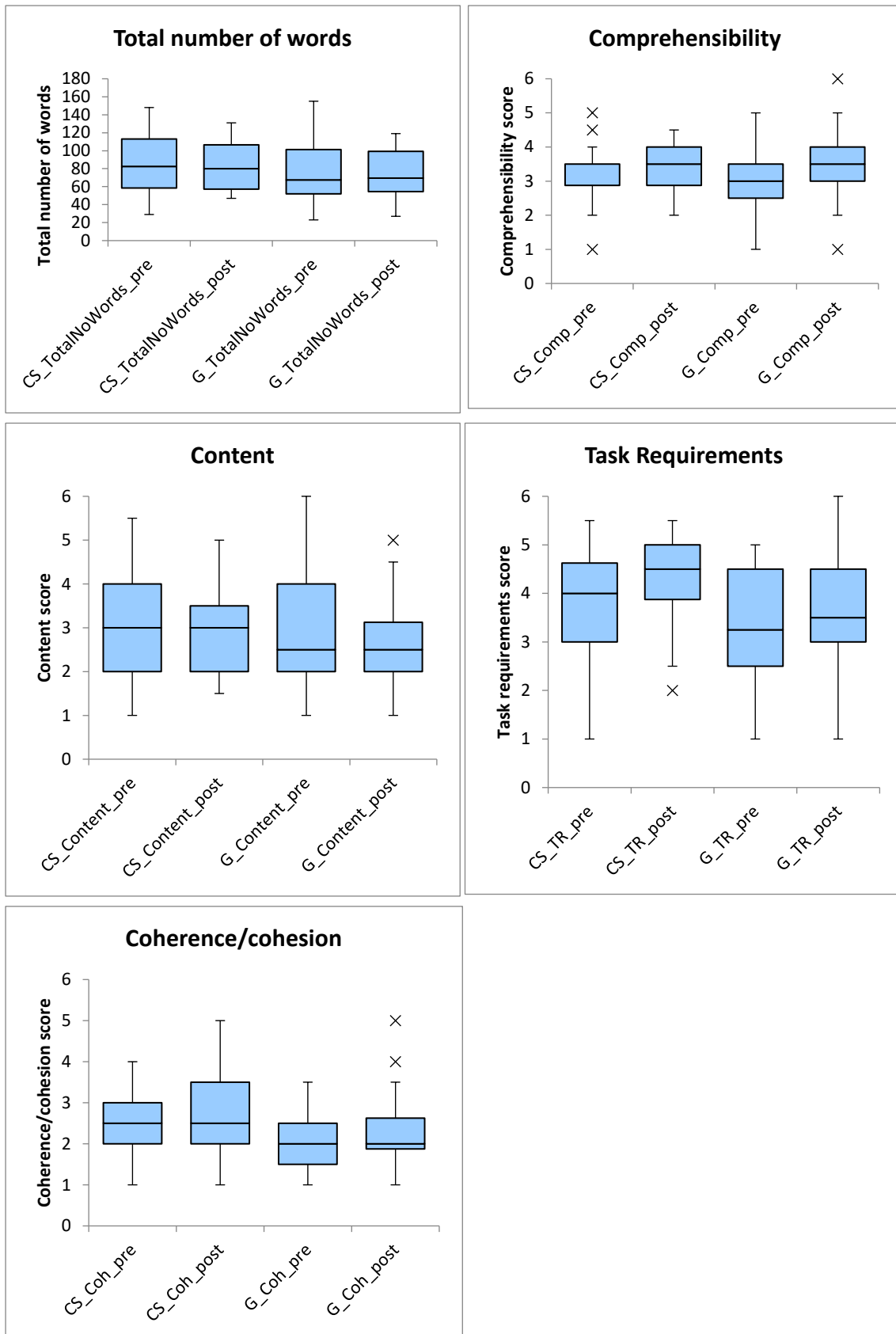
Note. CS stands for the content specific topic (power generation topic), G stands for the

general topic, pre stands for pre-test, and post stands for post-test.

First, as for the main effects of the topic factor, a significant difference was found for fluency (F [1, 31] = 10.6744, p = 0.0027, η^2 = 0.2561), lexical diversity (F [1, 31] = 31.8288, p < 0.001, η^2 = 0.5066), mean length of T-unit (F [1, 31] = 25.5225, p < 0.001, η^2 = 0.4515), task requirements (F [1, 31] = 9.0784, p = 0.0051, η^2 = 0.2265) and coherence/cohesion (F [1, 31] = 21.3944, p < 0.001, η^2 = 0.4083), showing large effect sizes. Although not significant, marginal significance was found for accuracy (F [1, 31] = 4.3792, p = 0.0447, η^2 = 0.1238). Comprehensibility was unaffected in terms of topic familiarity difference. Subsequently, the post hoc analysis for simple main effects was performed, and the significant and marginally significant results are summarized in Table 3.7 with p values and effect sizes.

Table 3.7

Significant Results from the Post hoc Analysis in terms of the Topic Difference

Measures	Time	Higher in general or content specific topic	p values and effect sizes
MLT	Pre	G<CS	p < 0.001, d = 0.816
	Post	G<CS	p = 0.021, d = 0.585
MTLD	Pre	G>CS	p < 0.001, d = 0.783
	Post	G>CS	p < 0.001, d = 0.886
Accuracy	Pre	G>CS	p = 0.013, d = 0.475
No. of words	Pre	G<CS	p = 0.011, d = 0.218
	Post	G<CS	p = 0.019, d = 0.332
Task Requirements	Pre	G<CS	**p = 0.082**, d = 0.333
	Post	G<CS	p = 0.004, d = 0.673
Coherence/ Cohesion	Pre	G<CS	p < 0.001, d = 0.857
	Post	G<CS	p = 0.043, d = 0.316

Note. The bold numbers for the p value indicate marginal significance.

The tests for the simple main effects revealed some differences between the topics. Specifically, the mean length of T-unit (MLT), number of words, and coherence/cohesion were higher for the content-specific topic in the pre- as well as post-tests. In addition, task requirements were significantly higher in the post-test and marginally higher in the pre-test. Conversely, lexical diversity (MTLD) was higher for the general topic in the pre- and post-tests, and accuracy was also higher for the general topic in the pre-test. Effect sizes ranged from small to medium according to Plonsky and Oswald (2014).

To answer the second research question "*How do students perform in the content-specific writing topic and general writing topic differently before and after the theme-based instruction?*," the results for the time factor were also examined. Though they were not significant, marginal interaction effects with medium effect sizes were found for accuracy (Time x Topic, F [1, 31] = 2.2262, p = 0.1458, η^2 = 0.067) and coherence/cohesion (Time x Topic, F [1, 31] = 3.8790, p = 0.0579, η^2 = 0.1112). Although not significant, marginally significant main effects for time with large effect sizes were found for comprehensibility (F [1, 31] = 5.7763, p = 0.0224, η^2 = 0.1571) and task requirements (F [1, 31] = 6.9361, p = 0.0131, η^2 = 0.1828). There was no change between pre- and post-writing for fluency, mean length of T-unit and content. When compared to the results of the topic difference of which comprehensibility was the only variable that remained unaffected, it can be said that the topic difference caused more differences in the writing performance than receiving instruction in the current research context.

Subsequent tests for simple main effect comparisons revealed that the task requirements were significantly higher in the post-test for the content-specific topic. Furthermore, accuracy was also found to be marginally higher in the post-test for the content-specific topic. As for the general topic, coherence/cohesion was higher, and comprehensibility was marginally higher in the post-test. The values and effect sizes are summarized in Table 3.8. The effect

sizes were small. These results will be discussed in relation to CBI/CLIL writing research, TBLT research, and second language writing research.

Table 3.8

Significant Results from the Post hoc Analysis in terms of Time Factor

Measures	Topics	Higher in pre- or post-writing	p values and effect sizes
Accuracy	CS	pre<post	**$p = 0.088$**, $d = 0.308$
Comprehensibility	G	pre<post	**$p = 0.082$**, $d = 0.312$
Task Requirements	CS	pre<post	$p = 0.010$, $d = 0.547$
Coherence / Cohesion	G	pre<post	$p = 0.025$, $d = 0.397$

Note. CS indicates content-specific topic and G indicates general topic. The bold numbers for the *p* value indicate marginal significance.

3.3.2 Results of the Learner Perceptions about the CBI Unit

To evaluate the current CBI unit from the learners' perspectives, their written reflection was thematically analyzed. Eleven codes were obtained, and their percentages are shown in Table 3.9. Some of the mentions were in a neutral tone while others were both positively and negatively connoted (e.g., I did not understand it well, but I want to try hard to keep up [まだあんまりわからないのでついていけるように頑張りたい。]). In this case, the sentence was counted as both positive and negative in Table 3.10. Similarly, when two skills (e.g., listening and vocabulary) were mentioned in one positively-charged sentence, positivity was counted for both listening and vocabulary in Table 3.9, but it was counted as one positive sentence in Table 3.10. Therefore, the positive and negative mentions in Table 3.9 do not necessarily add up to the number in Table 3.10.

Table 3.9

Number of Mentions for Each Code and their Positivity and Negativity

Codes	Total no. of mentions	Positively-charged mentions	Negatively-charged mentions
Code 1: Target content of the unit	39	24	4
Code 2: Target language of the unit	28	20	3
Code 3: Reading	12	4	9
Code 4: Listening	25	7	19
Code 5: Speaking	4	1	3
Code 6: Writing	17	8	9
Code 7: Vocabulary	21	6	16
Code 8: Grammar	4	0	4
Code 9: Holistic comments about the class and class delivery	8	0	7
Code 10: Understanding level	30	10	23
Code 11: Others	5	0	0

Table 3.10

Number of Positive, Negative, and Neutral Mentions

Codes	The number of mentions
Positive	64 (41%)
Negative	75 (48%)
Neutral	17 (11%)

The analysis revealed that the students reflected on both the target content and language forms, which are mostly positive reflections (e.g., "I haven't thought about them [power generation problems] deeply, but it was nice that I could learn about it in depth [あまり考えたことがない話だけど、詳しく知ることができてよかったと思う。]," and "I understood how to use the conjunctions we learned today [今日出てきた接続詞の利用方法が理解できた。]"). These aspects were paid attention to by the researcher in the preparation and delivery phases, and the students may also have felt that they learned these aspects. Various other language aspects were mentioned (e.g., codes 3 to 8 in Table 3.9), but all of them were more negatively connoted. Especially, listening and vocabulary induced negative wording, suggesting that

listening to the videos and lectures and learning unfamiliar vocabulary particularly raised the students' cognitive complexity. They also reflected on their own understanding levels shown in code 10, which were more negatively connoted.

The number of positive, negative, and neutral mentions is summarized in Table 3.10. Slightly more negative mentions were found. In Table 3.11, the most frequently used expressions for the positive and negative mentions are summarized. As for the positive expressions, students reflected on their understanding and what they could do in class. Additionally, they showed motivation for studying more about the content and language targets and reviewing what was covered in class (e.g., "I will review how to compare/contrast and how to write sentences [比較の仕方、文の作り方をよくみなおしておこうと思った]"). Furthermore, they wrote about their satisfaction for learning about new content (e.g., "I'm glad that I learned about the latest proportion of different power generation methods used in Japan [日本の発電所の種類の割合の最新のものを知れてよかった。]") and expressed enjoyment and interest in learning about the topic (e.g., "There were many things I learned for the first time, and it was fun [初めて知ることが多くて楽しかった。]," and "My knowledge and interest toward power generation issues have become higher than before [発電に関しての知識や関心が今までよりも高まった。]"). It can be concluded that the focus on content may have positively influenced some of the learners' affective state toward learning.

Table 3.11

Top Five Most Frequent Expressions to Show Positivity and Negativity

Ranking	Examples of positive comments	Examples of negative comments
1	Understood, learned (21)	Difficult (44)
2	Was able to~ (16)	Wasn't able to~ (15)
3	Want to study, review (14)	Did not understand (10)
4	Satisfaction (6)	No interest (2)
5	Enjoyment (4), interest (3)	Tiredness (1), frustration (1)

In contrast, the negative expressions were predominantly about the difficulty of the class, followed by what the students could not do or understand (e.g., "Vocabulary was difficult and hard to memorize, so I won't be able to recognize it when I see it next time [単語が難しくて覚えきらないので次見てもわからないと思う]"). For the current participants, it was challenging to learn about new content in L2, and this may have caused negative feelings, frustration, and tiredness. While some students were motivated by the inclusion of content learning, a small number of comments (two comments) showed that the students had no interest in this topic.

Analysis of the reflection sheets highlighted the learners' focuses in the CBI class and brought to light the possible positive and negative impacts on them. These results will be discussed in relation to the first two research questions.

3.4 Discussion

3.4.1 Topic Difference Variable

In this first study, one of the CBI-related writing task variables, namely the content familiarity variable, was examined by comparing the performances of the content-specific power generation topic and the everyday, general topic. Theme-based instruction unit based on the power generation topic was also created by referring to the several existing CBI lesson planning models (Echevarría et al., 2017; Mohan, 1986), and writing performance before and after the instruction was compared. The results indicated the existence of topic effects on EFL high school learners' writing performance. For instance, the content-specific topic writing task received higher scores both in the pre- and post-tests in terms of syntactic complexity (e.g., MLT), fluency, task requirements, and coherence/cohesion. Contrastingly, the general topic writing's lexical diversity was higher both in the pre- and post-tests, and accuracy was higher in the pre-test. As for the impact of the theme-based instruction, the

different aspects of writing changed for the two topics: task requirements and accuracy were higher in the post-test for the content-specific topic whereas comprehensibility and coherence/cohesion were higher for the general topic. There were more statistical differences and higher effect sizes for the topic difference variable than for the instruction variable.

First, the results for the topic difference will be discussed in reference to the related TBLT and second language writing research. Comparison of the content-specific power generation topic, which was perceived as being more difficult by 60% of the students, and the general topic, which can be written by utilizing the learners' general world knowledge and personal preferences, revealed that the content-specific topic elicited higher performances than the general topic in four measures. This overall finding only partly supports Robinson's (2003) Cognition Hypothesis. It predicts that possession of prior knowledge (in this study, the general topic condition) in the resource-dispersing dimension heightens all CAF measures. In the current study, lexical complexity (lexical diversity, both pre- and post-tests) and accuracy (pre-test only) were higher for the general topic as predicted in the hypothesis, but one of the syntactic complexity measures (i.e., mean length of T-unit) and fluency were higher for the content-specific topic. Thus, it suggests that a cognitively less demanding task in terms of content familiarity does not necessarily elicit better performance from EFL secondary school learners. In addition, it can be said that the task condition of the current study may elicit the performance characteristic of the resource-directing dimension since the task with higher complexity (the power generation topic) elicited higher syntactic complexity (but not higher accuracy), fluency, and rating scores (task requirements, coherence/cohesion). Reviewed from Skehan's (1998) Limited Attentional Capacity model, this study supports it since content-specific topic conditions elicited higher syntactic complexity but lower accuracy compared to the general topic, thus demonstrating a trade-off between complexity and accuracy.

The results suggest that using a writing prompt that requires content knowledge and is perceived to have a certain level of difficulty may elicit higher performance for certain linguistic aspects and rating. Since authorized English textbooks in Japanese high schools are increasingly becoming content conscious by reflecting the revised Courses of Study (e.g., expanding knowledge and thoughts on the SDGs in Ikeda, 2022), this type of writing prompt is compatible with the new textbooks and may be able to elicit different aspects of learner language that would not have been utilized in a general writing topic that necessitates general world knowledge and personal preference. The use of thought-provoking writing prompts with sufficient content and language support is a critical addition in Japanese high school English class to support the learners' L2 development.

Next, the results of each measure will be discussed. For the content-specific topic (i.e., the less familiar topic), syntactic complexity (MLT) was higher in both the test occasions, which is contrary to Abdi Tabari et al. (2021), Kessler et al. (2022), McDonough and Crawford (2020), and Yoon (2017) in that the familiar topics received higher length-based syntactic measures. In contrast, a similar pattern to the current study was seen in Tedick (1990) and Yang and Kim (2020). In Yang and Kim (2020), a less familiar topic (Internet's benefits and problems for people in underdeveloped areas vs. for university students) elicited a higher mean length of T-unit, although the statistical result was not significant with a small effect size. Focusing on the less familiar writing topic both in Yang and Kim (2020) and the current study, it may be speculated that a less familiar topic may have encouraged the students to explain the content by making each sentence longer than that in the general topic. It might be suggested that if the content familiarity condition does not arise from a complete lack of content knowledge (e.g., asking Japanese high school students to compare/contrast tax deduction systems in Japan and the U.S.) but rather from conscious efforts to create ideas and organize them [for Yang and Kim's (2020) example, students could imagine the benefits and

problems of the Internet in underdeveloped areas, but with more efforts and imagination than thinking about their familiar situation], it may not deteriorate the syntactic complexity.

In a similar vein, a higher number of words was observed in the content-specific topic in both the pre- and post-tests, which corresponds to Tedick (1990). Although this task type is thought to be cognitively difficult by the students, it actually pushed them to create longer output on both the test occasions. The use of a topic which requires some knowledge (in this case, the characteristics of each power generation method) and has reasoning demand (e.g., explaining a better power generation option with objective reasons and not personal preferences) may be more suitable than the topic which requires the learners' general world knowledge and preference for eliciting extended written output in L2 from them. In fact, higher-order thinking skill is considered as an important building block of lessons in CBI and CLIL literature. For example, Echevarría et al. (2017) recommend preparing higher-order thinking questions in the lesson planning stage that require analysis, synthesis, and evaluation of information in Bloom's taxonomy (1956) in their SIOP model. As for CLIL, Coyle et al. (2010) also emphasized the essential role of students' cognitive engagement in the content, which is included in one of the 4Cs (i.e., cognition). Therefore, cognitively challenging questions prepared for CBI/CLIL class, which encourage students to analyze, synthesize, and evaluate, may elicit more production for explanation both in the length of a whole essay and a sentence.

In addition to cognitive engagement, students' interest in the current power generation topic might have worked positively in the production amount. The theme was chosen according to the needs analysis of the current participants, and the topic related to their major was chosen. Reflection comments also confirmed that the content was intriguing to some of the students (e.g., "I want to know more about solar power generation [太陽光についてもっと知りたくなった。]," and "It was nice that the content is aligned with our major. There were

some difficult points, but it was more interesting than the content in the normal textbook [高専の電気科らしい内容で、少し難しい部分もあったが教科書よりも興味を持ちやすくてよかった。]"). For teachers to choose cognitively fewer complex prompts (general topics with less higher-order thinking) for EFL high school students is expected when taking these students' proficiency levels into consideration, and this type of prompt is actually dominant in the textbooks. Although the topic itself requires some content knowledge and the functional use of L2 (e.g., compare/contrast), the task may work positively in terms of fluency and overall syntactic complexity if the topic is aligned with the students' interests.

Next, two of the rating scores were better for the content-specific topic, which was in line with Tedick (1990). Task requirements were higher in the post-test with a medium effect size and marginally higher in the pre-test for the content-specific topic with a small effect size. The larger effect for the post-test is understandable since the students practiced comparing different power generation methods mentioning the pros and cons in class and may have been able to apply their in-class practice to their power generation writing task, which is the same topic as the class content. This can be supported by transfer appropriate processing (TAP; Roediger & Guynn, 1996), which predicts that the language items learned in one setting are transferable to a similar language use setting that involves similar cognitive processes. Interestingly, the task requirements score did not change for the general topic even though the required function was the same for the two tasks (i.e., compare/contrast the benefits and downsides of the two options and choose a better one with reasons). This supports the above point (Roediger & Guynn, 1996) in that the skill is most easily applicable in a similar context where it was first practiced.

Moreover, coherence/cohesion was also higher in the pre- and post-tests, and the significance was stronger in the pre-test with a large effect size. The content-specific topic may have required the students to explain their ideas to a greater extent with cohesive

devices, such as conjunctions and adverbs, as the reasoning demand for the content-specific topic might have been higher than the general topic since the students had to justify their opinions using the specific content information rather than their personal preferences in the general topic writing.

As for the more familiar general topic writing, it elicited lexically more diverse writing for both times, which is in line with many prior studies (He & Shi, 2012; Johnson, 2017; Kessler et al., 2022; Yang & Kim, 2020). Students may have been able to utilize their existing vocabulary in the general topic to discuss their preferences with reasons. Conversely, in the content-specific topic, they needed to use unfamiliar content-related vocabulary, which may have limited the use of their existing vocabulary. This tendency continued to be observed in the post-writing even after receiving the CBI instruction with a medium to large effect size ($p < 0.001$, $d = 0.886$). As mentioned in the student reflections, the content-specific words were suggested to have caused challenges in understanding the content, and they may not have been able to use the variety of words as they wished (e.g., "Reading materials were difficult because there were many specialized terminologies [専門的な語句が多く読みづらい]"). Therefore, teachers should be cognizant of the fact that content-specific vocabulary may become a hurdle for learning in CBI/CLIL and that they need to pay conscious attention to the students' understanding and utilization of the vocabulary by increasing the use of core vocabulary in teacher talk, adjusting the level of vocabulary in the input material, and utilizing tasks for vocabulary use practice. When interpreting the results, however, it should be pointed out that there were three different general writing prompts with different writing topics, which may require a different set of vocabulary, while the three power generation prompts required a similar set of vocabulary. Although having three different prompts was to counterbalance the topic effect, this condition may have influenced the result of higher lexical diversity for the general writing topics.

In addition, accuracy was higher only for the pre-test in the general topic writing with a small effect size, which is in line with He and Shi (2012) but contrary to McDonough and Crawford (2020), Kessler et al. (2022) and Yang and Kim (2020) since these studies found no difference or lower accuracy. However, it is in line with Robinson (2003) and Foster and Skehan (1996) as a cognitively less complex task (i.e., a general topic) elicited accurate performance. It is speculated that the participants may have been able to spare their attentional resources for accuracy in the general topic more than the content-specific topic since content formulation may have been aided by the learners' general knowledge and accessible vocabulary. However, accuracy in the content-specific topic showed a slight increase from pre- to post-test (0.63 to 0.67) although it was not statistically significant ($p =$ 0.088, $d = 0.308$). Due to this improvement, the difference in accuracy between the writing topics disappeared in the post-test. Improvement may have occurred because the students gained content and language knowledge through the instruction and may have been able to pay attention to accuracy in the post-test. The results of the current and previous studies are summarized in Table 3.12. It can be seen that lexical complexity tends to be higher for the + topic familiarity prompt, but mixed results have been obtained for the other measures. The current study obtained similar results with Tedick (1990), both of which compared a content-specific topic that is aligned with learners' majors and a general topic.

So far, it has been shown that the cognitive task complexity difference operationalized as topic familiarity (a content-specific and general topic) has a different impact on learner writing. One highlight is that a cognitively complex task does not necessarily elicit lower performance from EFL high school students, and this type of writing prompts should be used to develop their language skills and enhance cognitive engagement in a meaningful task. Next, the influence of theme-based instruction on pre- and post-writing is discussed.

Table 3.12

Results of CAF and rating for the + topic familiarity prompt (a general topic) in comparison to – topic familiarity prompt (a field-specific topic) for the current and previous studies

	Syntactic complexity	Lexical complexity	Accuracy	Fluency	Rating
Current study	Lower or no difference	Higher (both pre- and post-tests)	Higher (pre-test only)	Lower	Lower or no difference
Abdi Tabari et al. (2021)	Higher	–	–	–	–
Yoon (2017)	Higher	Higher	–	–	–
Tedick (1990)	Lower	–	–	Lower	Lower
Yang & Kim (2020)	No difference	Higher	No difference	No difference	–
Kessler et al. (2022)	Higher	Higher diversity, lower sophistication	Lower	No difference	–
McDonough & Crawford (2020)	Higher	–	No difference	–	Higher
He &Shi (2012)	–	Higher use of academic words	Higher	Higher	Higher

3.4.2 Pre- and Post-instruction Variable

Compared to the topic variable, the instruction caused fewer changes in the writing (i.e., one significant and one marginal change for both the content-specific and general writing). Content knowledge gained through theme-based instruction may have helped satisfy the task requirements and marginally increase the accuracy. As discussed in the previous section, more successful achievement of task requirements in the post-test may have been supported by theme-based instruction where the tasks that have the same task requirements (e.g., compare contrast pros and cons of power generation methods and justify the optimal ways) were conducted in class (TAP; Roediger & Guynn, 1996).

A similar discussion can be conducted for accuracy, since gaining knowledge about content and language in the theme-based instruction may have supported the learners in paying attention to accuracy by lightening the cognitive load for idea generation, which is in line with both Robinson's (2005) and Foster and Skehan's (1996) hypotheses. However, in Vandommele et al. (2017), which compared the writing performances before and after the content-focused instruction (e.g., creation of a tourist website), the syntactic complexity, lexical complexity, fluency, and rating developed, but accuracy did not. The results of this study are opposite from that of the current research, which may be owing to the different participant populations. Vandommele's et al. (2017) participants were beginner-level adolescent immigrant students in the Netherlands while the current study's participants were beginner-level Japanese high school students. The former group may have emphasized conveying more information for communication while the latter group may have prioritized accuracy over complexity and fluency, which is typical in grammar-focused EFL contexts. In addition, the current results contradict Ikeda (2013), where the participants of both studies are Japanese high school students, in that complexity and fluency improved and accuracy deteriorated in Ikeda (2013). It may have been influenced by the research period (i.e., one year vs. 12 weeks) and timing of data collection. CAF measures do not develop linearly, but they gradually improve with numerous fluctuations (Baba & Nitta, 2014) as suggested in the complexity theory, which views language development as dynamic and accompanying a phase shift (Larsen-Freeman, 2006). Therefore, the timing of the data collection may have reflected the fluctuations at that point. Additional long-term empirical CBI/CLIL studies in the EFL contexts with frequent data collection occasions are called for.

The theme-based instruction also had positive effects on general writing in terms of coherence/cohesion at a significant level, which is in line with Whittaker et al. (2011), and comprehensibility at a marginal level. Thus, learning a language based on specific content

may have some positive effects on the students' writing performance in other writing topics. In the general writing, the linguistic aspects (i.e., CAF) did not change, but two dimensions of the functional adequacy scale did. Therefore, learning a language with specific content may not improve the linguistic aspects of other topics, but the holistic aspects of writing rated by a rating scale may be positively influenced if the task structure is compatible with the language functions dealt with in CBI/CLIL (e.g., compare and contrast for the current study).

Unlike other CBI/CLIL studies which investigated the influence of instruction and confirmed the development in various aspects of written language (Ikeda, 2013; Kong, 2015; Shibata, 2021; Roquet & Pérez-Vidal, 2017), the current research found only one significant change and one marginal change for each prompt type. It may be due to the brief instructional duration (30 minutes) over a short period of time (12 weeks) compared to other longitudinal CBI/CLIL research, which typically last more than a year. Moreover, the current participants' proficiency was still low, and they may have needed more time and intensity of instruction for the changes to occur. The results of the current and previous studies about the impact of CBI/CLIL on writing are summarized in Table 3.13. Compared with other previous studies that confirmed the positive impact of CBI/CLIL on CAF and rating, the current study's impact seems to be lower presumably due to the short instructional period.

3.4.3 Limitations and Pedagogical Implications

The study findings indicate that distinct levels of task complexity operationalized as topic difference bring about different learner performances, and the writing for the complex topic was shown to be higher in length measures and rating. Thus, the writing task that requires some content knowledge may push the learners to produce more language, write coherently, and satisfy the task requirements. A slightly challenging task may support the diverse aspects of writing development than the general topic writing, which is typical of

Japanese high school English textbooks. Conversely, the influence of instruction was less pronounced compared to the topic difference variable possibly due to the limited amount of instruction time. However, it has been shown that not only specific topic writing but also general writing was positively affected by theme-based instruction. These results may add empirical evidence for the implementation of theme-based instruction in Japanese high schools.

Table 3.13

Results of CAF and rating after the theme-based instruction in comparison to the results before the instruction for the current and previous studies

	Syntactic complexity	Lexical complexity	Accuracy	Fluency	Rating
Current study	No change	No change	Higher (content-specific topic)	No change	Higher (content-specific and general topics)
Vandommele et al. (2017)	Higher	Higher	No change	Higher	Higher
Ikeda (2013)	Higher	Higher	Lower	Higher	Higher
Kong (2015)	–	–	–	–	Higher
Shibata (2021)	–	Higher	–	–	–
Shibata (2019)	–	–	–	Higher	–
Whittaker et al. (2011)	–	–	–	–	Higher (coherence)

Note. – represents that the measure was not dealt with in the study.

Although the results illuminated the effect of topic difference and instruction with an under-researched population (i.e., Japanese high school students), there are certain limitations. The first one is that the construct definition of topic familiarity would need to be specified. It could include two distinct components, namely difficulty and interest, which could differently affect the quality of writing. Moreover, difficulty can be further classified

115

into content difficulty and language difficulty as Tedick (1990) has mentioned. These three components are intertwined and collectively affect the task complexity level. In other words, even if the topic is less challenging in terms of the content and language, the participants' low interest in the topic may negatively affect the writing quality. Thus, the interest component may have been an intervening variable since it was not controlled in this study. The construct definition of topic familiarity needs to be considered further as this is an essential task variable for writing assessment (He & Shi, 2012; Kessler et al., 2022; Saiki et al. 2022) and task-based teaching (Yang & Kim, 2020). As for methodological operationalization, the current study asked students to choose one of the two more difficult prompts or choose both if they thought that it was challenging to compare the cognitive complexity level of the two types of prompts. However, as Kessler et al. (2022) and Qiu and Lo (2017) pointed out, it would have been better to ask the students to rate their familiarity level instead of asking them to choose to take into account the degree of each learner's perceived topic familiarity.

As for the effect of instruction, the instructional period was short (12 weeks) due to the constraints of the class schedule. A longer implementation period may have yielded more robust changes, and such exploration is needed in future research. In addition, the students were concurrently taking other English classes (2 hours to 2.5 hours of comprehensive English, 45 minutes of English conversation, and 45 minutes of grammar per week) at the time of data collection; therefore, the changes in writing may also be influenced by these classes. In addition, since the data were collected from one intact class, the number of participants was limited, and the results may not be generalizable to a wider population.

Despite these limitations, this study provides pedagogical implications for CBI/CLIL implementation in EFL contexts, especially for secondary school students who are in the process of developing their language skills. First, the type of writing prompts used in class for secondary school students does not have to be cognitively less complex everyday topics that

require superficial information. As Robinson (2003) suggested, high cognitive task complexity also requires additional language resources and facilitates access to these resources; therefore, tasks with a slightly higher cognitive complexity also greatly contribute to language learning. As for the current study, a content-wise and linguistically less familiar power generation topic, which might have intrigued the students' interest, elicited higher performances in terms of length measures and rating. These positive results may support the inclusion of cognitive prompts that are slightly challenging into secondary school English classes, which is in accordance with the revision of the Courses of Study. Use of outside information (e.g., content knowledge covered in the CBI/CLIL class) and answering higher-order questions (e.g., justifying the use of one option considering the benefits and downsides in this study) is a promising addition to task repertoires.

Another implication is that the CBI/CLIL instruction had a positive impact on the different aspects of the content-related and unrelated topic writing tasks. Although the effect sizes were small, the findings suggest that focusing on one particular content deeply for an extended period of time in CBI/CLIL may not limit the learning effect on the same topic writing.

Lastly, pedagogical implications can be derived from the analysis of the reflection sheets. Comments highlighted the pros and cons of CBI/CLIL in EFL secondary school students who are in the process of developing their proficiency. Holistic and multimodal learning of content (i.e., not limited to a particular skill, such as a speaking class) may help students focus on various language skills as in the codes 3 to 8 in Table 3.9 and enable them to achieve a sense of understanding and achievement of content and language learning. The inclusion of thought-provoking content may positively affect the learners and stimulate their motivation to learn more and review, raise their interest in the topic, and induce enjoyment. In addition to affective benefits, such a class was shown to cause positive change in writing

as well. In contrast, the comments also revealed that learning new content that requires new vocabulary and higher-order thinking may overwhelm the students as most of the negative comments point to the difficulty level of the class. Therefore, teachers should attempt to strike a balance between the content levels to intrigue the learners' cognitive involvement and the language levels for content learning. As Echevarría et al. (2017) suggested in their model, providing ample scaffolding in various forms (e.g., visual support by using a written form and pictures, modified input in written and oral forms) with careful planning of a lesson is vital for the learners' positive experience in CBI/CLIL. Moreover, the supportive role of L1 in integrated learning in L2 (e.g., translanguaging, Lo & Lin, 2019; Nikula & Moore, 2019) needs to be further cultivated, especially for EFL secondary school learners with relatively low proficiency so as to not demotivate students with cognitively challenging content and language learning. Overall, shifting the focus from language to content for Japanese high school learners' English class is shown to have positive and negative effects arising from higher content and linguistic demands. Since the writing data in this study confirmed the different writing strengths of the learners in content-specific and general topics, it is worthwhile for teachers to strive for maximizing the cognitive and motivational benefits of content integration while controlling for the difficulty level.

Study 1 investigated the topic difference (i.e., one topic aligned to the CBI content and a general one) and the effect of CBI instruction on Japanese high school students' written language. One of the essential pillars of CBI/CLIL is content learning, which is brought about by integrating various information from different sources of input (e.g., lectures, readings, presentation slides, realia, videos, and peer discussion) into the learners' cognitive space. Among these, reading is common and one of the major sources of content input for learners, and integration of reading and writing can naturally occur in CBI/CLIL. Therefore in Study 2, Japanese college students' learning in the CBI task (i.e., integrated writing tasks) and writing

output will be more closely investigated. Additionally, the impact of the learning of content

and language (L2 writing skills in this study) will be explored from the students'

perspectives.

Chapter 4. Study 2: Content-Based Writing Class: Two Types of Integrated Writing Tasks and Students' Class Perceptions

4.1 Introduction

Integrated writing task has been recognized as one of the important and versatile writing task variables in CBI/CLIL (Plakans, 2015). Numerous studies on reading and writing integration conducted in content-based writing classes have been conducted (Kavanagh, 2019; Shih, 1986; Underwood, 2019) since CBI/CLIL can provide a solid foundation for skill integration and argument construction. Although the writing tasks, especially for Japanese learners, tend to be highly structured and similar to grammar exercises (e.g., translation of a short sentence; Kobayakawa, 2011) or require the writers' personal experience as a sole source of information, the necessity of integrating ideas from other sources into their academic and professional writing is increasingly identified as important. Moreover, studies that view integrated writing as a language learning task are still scarce in TBLT and second language writing research (Abrams, 2019) as this task type is mainly investigated in the field of language testing (Yu, 2013; Plakans, 2015). Differences in task complexity in terms of different integrated writing task conditions (e.g., a degree of information integration) also require additional research so that teachers can effectively utilize this task type in their language classes.

Therefore, the second study explores how novice foreign language (FL) writers' writing performance differed according to the level of source text integration in writing tasks as they engaged in systematically designed content-based writing instruction. Furthermore, it examines their task perceptions on two writing task types through interviews. Previous research on this task type mainly targets ESL pre-sessional university students who need to prepare themselves for formal university classes (Neumann et al., 2019), and fairly few studies focus on the language learning effects of a source integration writing task in EFL

settings. Moreover, the language learning impact of such a task has not been investigated with learners with emergent language proficiency who are in the process of developing their FL output skills. The research questions are as follows.

Study 2-1: *Is there any difference between Japanese college students' writing performances on the two types of integrated writing (i.e., source-text integration required or not)?*

Study 2-2: *How did the students perceive the difference in the two writing task types?*

Study 2-3: *How did the students perceive the combined learning of content (SDGs) and English writing?*

The following section details the research context and second CBI course created for the current writing class. Then, the results of the writing and interview analysis will be presented and discussed.

4.2 Method

4.2.1 Participants

Participants were 40 Japanese college students aged 18 to 19 years with low to intermediate proficiency levels ($M_{TOEIC} = 345$, SD = 89, range = 225–570). They were in the same five-year college as the participants in the first study and took the current elective academic writing class. Their majors varied (mechanical engineering, electrical and computer engineering, electronic control engineering, materials science, and architecture). A background questionnaire survey that was created in reference to Hirose and Sasaki (1994) revealed that their English writing experience during the previous semester was limited. Around 27 students (67.5%) answered that they had no opportunity to write in English while the other 13 students had completed at least a page of English writing. About 39 students

(97.5%) found English writing either extremely difficult or difficult. Their main purpose for taking the class was to increase their TOEIC score and prepare for the university transfer exams[3]. From the instructor's estimation by reading their writing samples, their level of writing can be regarded as being at the intermediate-level as per the American Council on the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines. As for their previous knowledge about the content, which was the SDGs and especially the goals related to environmental problems, only five students answered they could explain the concept of the SDGs and some goals. Therefore, it can be inferred that their background content knowledge was relatively limited. In addition, as for their knowledge about writing, 29 students (72.5%) had already learned how to cite outside sources in their previous classes held in Japanese, such as in social science general education classes or specialized classes in their department.

At the beginning of the first class, the researcher informed the research purpose to the students, informed about their authority to refrain from participating in the study, and explained that there was no relationship between their grade and their withdrawal from the study. The explanation was given in oral and written forms, and consent was obtained. Four students claimed withdrawal, and their writing samples were not included in the analysis. Thus, a total of 36 students participated in the classroom study and agreed on the use of writing data.

4.2.2 Content-Based Writing Class Development

Another content-based class for writing was created for this study. Content-based approach to teaching writing is described by Shih (1986) as follows: "Writing is integrated with reading, listening, and discussion about the core content and about collaborative and

---

[3] The research site is a five-year college where students can receive an associate degree. After graduation, some of the students transfer to the third-year in undergraduate courses in four-year national universities.

independent research growing from the core material (p. 618)." As specified, content serves as a base for skill integration, and the current study especially focused on reading–writing integration.

Content and language objectives for the course were set. Content objectives were to be able to explain the basic concepts of the SDGs (e.g., what it stands for, background of its establishment, and content of the 17 goals) and form opinions and suggestions toward environmental problems that are related to Goals 7 (affordable and clean energy), 14 (life below water), and 15 (life on land). To achieve these objectives, the students conducted various activities, including reading the related reading materials, watching video clips, independently searching for information, discussing, and writing. Conversely, the language objectives were to be able to write well-organized paragraphs and essays, use discourse markers, and appropriately use outside sources by paraphrasing and summarizing. Moreover, they practiced incorporating their own opinions in writing. Table 4.1 shows the overall schedule of the course.

Table 4.1

Content and Language Objectives and Tasks Used in the Course

Class	Content objectives	Language objectives	Tasks
Class 1 (10/6)	To understand the aim of the SDGs and their background	To familiarize students with the paragraph format	Listening for information by watching a video clip Reading about the aims of the SDGs and answering comprehension questions Doing an error correction task for learning about mechanics
Class 2 (10/20)	To learn what the 17 goals are and classify them into categories	To review the paragraph format To learn the structure of a paragraph	Classification of goals and comparison with two classification models Practicing identifying and writing a good topic sentence, supporting sentences, and conclusion Practicing writing an opinion paragraph

Class 3 (10/27)	To consider the impact of the SDGs on companies	To learn the structure of an essay To learn the concepts of coherence and cohesion	Brainstorming on the impact of SDGs on companies Reading about the same topic in paragraph and essay formats and comparing them Identifying the use of connectives and cohesive ties Practicing writing an opinion paragraph
Class 4 (11/10)	To learn about Goal 14 To consider the problems of marine animals and their habitats	To be able to summarize and paraphrase	Picture–vocabulary matching Listening for information by watching a video clip Summarization and paraphrasing practice
Class 5 (11/17)	To learn about Goal 7 To understand Japan's current power supply configuration	To review coherence and cohesion	Picture–description matching Listening for information by watching a video clip Reading comprehension questions
Class 6 (11/24)	In-class writing 1 related to Goal 7 (without source integration writing 1)		
Class 7 (12/8)	To learn about Goal 7 To be able to explain Japan's energy policy To understand Australia's energy policy as one example	To review the rules about mechanics	Processing feedback on writing and doing error correction tasks Listening for information by watching a video clip Reading comprehension questions
Class 8 (12/15)	In-class writing 2 related to Goal 7 (with source integration writing 1)		
Class 9 (12/22)	To learn about Goal 15 To reflect on the importance of forests in our lives To be able to read graphs on deforestation	To incorporate linguistic, organizational, and mechanical feedback into their writing	Processing feedback on writing Brainstorming the reasons for forest conservation Reading graphs and explaining the trends Listening for information by watching a video clip
Class 10 (1/5)	In-class writing 3 related to Goal 15 (without source integration writing 2)		
Class 11 (1/12)	To learn about Goal 15 To think about why biodiversity is being lost To learn about the	To incorporate linguistic, organizational, and mechanical feedback into their writing	Processing feedback on writing Brainstorming the reasons for biodiversity loss Listening for information by watching a video clip

	three kinds of ecosystem services		
Class 12 (1/26)	In-class writing 4 related to Goal 15 (with source integration writing 2)		
Class 13 (2/2)	Wrap-up of the SDGs	To incorporate linguistic, organizational, and mechanical feedbacks into their writing Test preparation	Processing feedback on writing
Class 14 (2/5)		Final test	
Class 15 (2/16)		Test return session	

Lesson sequences were constructed by first covering the basic concepts of both content (the SDGs) and English writing, followed by a series of writing tasks answering higher-order questions related to the goals covered. More specifically, from classes 1 to 4, the aims and background of the SDGs, what the 17 goals represent, their impact on not only countries and individuals but also companies, and one of the goals (Goal 14, life below water) was focused on and considered by using various pedagogical tasks that necessitate the use of all four skills and peer interactions. Simultaneously, the crucial concepts of English writing (i.e., mechanics, paragraph and essay structures, coherence and cohesion, paraphrasing and summarizing) were also dealt with by utilizing the same content to base the tasks for the writing practices. The practices included, for example, identification of a paragraph/essay structure, writing of a topic sentence, supporting sentences and conclusion, identification of connectives and cohesive ties in reading materials, and summarization of a paragraph in the reading (see Appendix F for the materials for the first class).

For classes 5 to 12, two consecutive lessons were paired. In the first lesson, one reading material related to either goal 7 or 15 was assigned to the students as homework. They were asked to read it at home and make a brief summary of each paragraph in Japanese to facilitate their understanding of the text. Then, the summary was checked in pairs, and additional

reading materials and listening tasks that aid an understanding of the environmental problems raised in the reading material were given. Example tasks were a picture–definition matching task and a listening task by watching a video clip for learning new content knowledge (see Table 4.1 for the content and language objectives and tasks used). In the first lesson, the focus was on the development of the content knowledge by providing multiple forms of input (e.g., reading, audio input, videos, pictures) and conducting output tasks (e.g., writing short answers, discussing in pairs, discussing as a whole class). In the second class, students were given a writing prompt related to the previous lesson and were asked to write an essay in class. This first draft of in-class writing was analyzed, and a detailed writing procedure is explained in the following section. The researcher gave feedback on each writing, and the common errors were shared with all the students. They wrote a second draft as homework. This reading + content learning and writing cycle was conducted four times, followed by a wrap-up class, final examination, and test return session.

As in study 1, the 30 principles of the SIOP model (Echevarría et al., 2017) were referred to in setting the content and language goals, creating the tasks, sequencing them, and increasing comprehensibility in delivery. The researcher read the principles and their explanation and used them as a checklist to consider the effectiveness of her teaching plan. In addition, each of the first four lessons and a set of the two lessons in classes 5–6, 7–8, 9–10, and 11–12 began with a focus on content learning, shifting it to the language aspects (i.e., English writing), and moving back to the content concepts as suggested by Lyster (2017). To integrate the learning of the SDGs and English writing coherently, connections among the themes, topics, texts, and tasks were consciously constructed in reference to the Six Ts model (Stoller & Grabe, 2017), and the example concept map for lesson planning is shown in Appendix G.

For material development, the researcher read five books on the SDGs (two in English

and three in Japanese) to gain insight into the content before she developed the materials. In addition, books and web resources, such as the homepage of the United Nations (UN) and YouTube videos created by the UN and the University of New South Wales, were used for the researcher's study and as teaching materials. Around 10 existing writing textbooks were reviewed to create the practice questions for studying writing principles. Finally, the handouts for content learning (including, for example, reading materials, listening practice questions, graph reading, and higher-order thinking questions), handouts for writing practice (including the practices for learning about the basic writing concepts listed in Table 4.1), and PowerPoint slides were created. Besides the materials created, five reading passages were chosen from the existing reading textbook (units 2, 6, 7, 9 in Yasunami & Lavin, 2020), which was assigned to the students as a textbook for this class.

4.2.3 Two Types of Integrated Writing Tasks

Writing performance for research analysis was collected in four in-class writing sessions (see Table 4.1) in classes 6, 8, 10, and 12. The aim of the analysis was to examine whether the EFL learners' writing performance differed in terms of linguistic and rating assessment according to the two types of integrated writing: writing related to reading but no need for source text integration (task A) and writing with the requirement of source text integration (task B). As Table 4.2 below summarizes, the reading materials that the students read before writing ranged from 428 to 479 words, and according to the Flesch Kincaid reading ease score, which is a widely used readability score (Kincaid et al., 1975), these readings can be easily understood by L1 English high school to university students. However, the instructor found that the readings were a little challenging for the current participants; therefore, they read the materials as homework and summarized each paragraph in Japanese before class, and their understanding was checked with peers in the first lesson.

Table 4.2

Number of Words and Readability Scores of the Four Reading Materials

Reading materials	No. of words	Flesch Kincaid Reading Ease
Writing related to a source reading but no need for the source text integration (task A-1, topic: power generation, unit 7 in the textbook)	479	39
Writing with the requirement of the source text integration (task B-1, topic: power generation, unit 6 in the textbook)	465	43
Writing related to a source reading but no need for the source text integration (task A-2, topic: deforestation and biodiversity loss, unit 9 in the textbook)	428	54
Writing with the requirement of the source text integration (task B-2, topic: deforestation and biodiversity loss, unit 2 in the textbook)	469	49

Table 4.3 below shows the four writing prompts used. The two prompts (tasks A-1 and A-2) did not require the integration of a source text, and the participants needed to come up with their own ideas and organize them. The reading material was not necessary for the completion of the task but was used as a springboard for writing. In contrast, the other two prompts (tasks B-1 and B-2) required the integration of information in the reading material. More specifically, the underlined parts in the prompts required the writers to read the material and extract the necessary information to accomplish the task requirements. The order of the prompts was as follows: task A-1, B-1, A-2, and B-2. It was impossible to counterbalance the order of the prompts among the participants because all of them were in the same class and followed the same instructions; however, the type of task (task A or B) was alternated.

Table 4.3

Writing Prompts Used for Data Collection

Type of tasks	Prompts
No source text integration (task A-1)	What can each of us do to reduce the amount of electricity use? (written 1st)
No source text integration (task A-2)	What are the things that we can do to stop deforestation? (3rd)
With source text integration (task B-1)	Learning from the countries in the reading, what can Japan do to improve its energy policy? <u>Choose (an) example(s) among four countries described in the textbook to support your idea.</u> (2nd)
With source text integration (task B-2)	<u>Why do we need to conserve biodiversity according to the textbook?</u> What should we do to protect biodiversity? (4th)

The prompts were piloted with two students from the same institution having similar English proficiency to the study participants. They were asked to read the reading materials, explain the content in Japanese paragraph-by-paragraph to the researcher, and complete two types of writing. Their understanding of the reading materials and writing prompts were checked through interviews, thereby prompting the researcher's decision to provide reading support in class before conducting in-class writing. The time for the writing task was decided as 45 minutes by observation. For task B (source-text information needed), both students included the information in the reading as the researcher intended.

4.2.4 Data Collection Procedure

In the second lesson, an in-class writing activity was conducted in the sequence as shown in Table 4.4. Using the first 10 minutes, the researcher reviewed the content covered with the PowerPoint slides used in the previous lesson and introduced the writing prompt (the difference between the two types of prompts will be detailed in the writing prompt section). Next, 15 minutes were given for individual planning using the handout for writing (see Appendix H). Students jotted down ideas in an open space in the handout by referring to the reading material, handout used in the previous lesson, and Internet resources of their choice.

They also considered the organization of the entire essay by writing down ideas for a topic sentence, supporting points, and conclusion. With the plan, students were put into pairs, and they discussed their plans with their partners mainly in Japanese. They introduced the examples they were going to use in their essays and checked their relevance to the topic sentence. They were told to note down any advice they received from their peers in their handouts. After the peer discussion, they were given 45 minutes to write their essays in class. The class observation conducted by the researcher confirmed that the students looked at their summary homework to review the content of the reading. In addition, there were no restrictions on using their dictionary and smartphone (including translation applications such as Google Translate) so that they were not amotivated by the challenge of writing. However, a translation of the complete essay written in Japanese into English was not allowed, and they were encouraged to use apps for a phrase-level search. In real language use situations, it is natural for L2 writers to use these devices to express their thoughts in writing; therefore, their use of these tools was not restricted, thus recognizing that their writing performance was aided by these tools. In-class writing, that is, their first drafts without feedback, was used for the analysis. The researcher provided feedback on language, organization, and content, and they handed in their revised drafts in the next class, which were not analyzed in this research.

Table 4.4

Instructional Sequence for the In-class Writing

Sequence	Activities
1	Review of the content covered in the first lesson using PowerPoint slides and introducing the writing prompt (10 minutes)
2	Writing an outline individually (15 minutes)
3	Sharing the outline with a partner (10 minutes)
4	Individual writing in class (45 minutes)

Those students' writing was excluded from the analysis, leaving 27 students: those who

wished to withdraw their data from this study, who were absent and thus did not write under the same condition even once and who wrote less than 50 words even once.

4.2.5 Writing Assessment and Statistical Analysis

All the hand-written data was typed into MS Word for analysis. To facilitate the comparison with study 1 and other related previous studies, the linguistic and rating measures used in study 2 were mostly the same as those in study 1. For example, as for the complexity and fluency measures, the following measures in Table 4.5 were obtained from automated analysis tools. Since it was assumed that the lexical sophistication level may be influenced by the integration of reading material, lexical sophistication 2 (LS2) was added and computed by the Lexical Complexity Analyzer (Lu, 2012). LS2 can be calculated by dividing the number of sophisticated word types, which are beyond the most frequent 2,000 words in the British National Corpus, by the total number of word types.

Table 4.5

Complexity and Fluency Measures Used for the Analysis

Kind of measures	Specific measures	Automated analysis tools used
Syntactic complexity	Mean length of T-unit (MLT)	Syntactic Complexity Analyzer (Lu, 2010)
Syntactic complexity	The number of dependent clause per T-unit (DC/T)	Syntactic Complexity Analyzer (Lu, 2010)
Lexical diversity	Measure of textual lexical diversity (MTLD)	Text Inspector (Bax, 2012)
Lexical sophistication	Lexical sophistication 2 (LS2)	Lexical Complexity Analyzer (Lu, 2012)
Fluency	The total number of words	Syntactic Complexity Analyzer (Lu, 2010)

As for accuracy, the Weighted Clause Ratio (Foster & Wigglesworth, 2016) was calculated (see section 3.2.5 of study 1 for how to obtain the score). The researcher and another rater who had 14 years of experience teaching secondary school students rated all the essays, and good inter-rater reliability was obtained (Cronbach's alpha = 0.86). Before each proceeded to the individual rating, the researcher conducted a rating session which took about an hour, explaining the different levels of accuracy using the examples in the paper (Foster & Wigglesworth, 2016) and sample essays obtained from a similar writer population. They then collaboratively rated the sample essays for norming, and any questions regarding score assignment were discussed and resolved. The average score of the two was used for the analysis.

To assess the quality of writing, the Functional Adequacy Scale (Kuiken & Vedder, 2017) was used as in study 1. The original score descriptors of the three scales (e.g., comprehensibility, content, coherence/cohesion) were used, but the descriptors of task requirements were adapted to suit the two types of tasks used in this study (see Tables 4.6 and 4.7). As for the without integration task, there were two requirements: (1) suggestion(s) required to be made in the prompt are present and they are explained and (2) an essay with three basic components (i.e., introduction, body paragraphs, conclusion) is written.

Table 4.6

Adapted Score Descriptors for Task Requirements for the Without Source Integration Task

Score	Descriptors
6	Both requirements have been successfully satisfied
5	Both requirements have been satisfied
4	One of the two requirements has been satisfied, and the other requirement has been attempted but not successful
3	Only one of the two requirements has been satisfied
2	Attempts have been made to satisfy the requirements, but none has been satisfied
1	None of the requirements has been attempted and satisfied

On the other hand, with integration task had three requirements: (1) Suggestion(s) required to be made in the prompt are present, and they are explained, (2) An essay with three basic components (e.g., introduction, body paragraphs, conclusion) is written, and (3) Citation of appropriate information from the reading is appropriately conducted (inappropriate citation includes a choice of inappropriate information, direct citation without explanation and copying). The first two are the same as the other task, but the third criterion was added to account for the task difference.

Table 4.7

Adapted Score Descriptors for Task Requirements for the With Source Integration Task

Score	Descriptors
6	All three requirements have been successfully satisfied
5	All three requirements have been satisfied
4	Two of the three requirements have been satisfied
3	Only one of the three requirements has been satisfied
2	Attempts have been made to satisfy the requirements, but none has been satisfied
1	None of the requirements has been attempted and satisfied

Score descriptors are adapted so that they reflect the degree of satisfying these two or three requirements. The degree of satisfaction (e.g., the number of requirements satisfied) in the adapted descriptors reflects the explanation in the original descriptors by Kuiken and Vedder (2017). After the researcher repeatedly rated the sample essays written by a similar population of students and adjusted the wording of the descriptors, two Ph.D. students with several years of teaching experience at secondary and tertiary level schools were asked to rate all the data. The first rating session for the without source integration task with each rater was conducted online for about 90 minutes. The researcher explained the writing prompts, scale, and benchmark essays with their scores to set the common criteria for the evaluation. Then, the sample essays were rated individually to determine the degree of correspondence, and any

disagreement was resolved through discussion. After completion of the assessment of the first set of writing, the second rating session was held to evaluate the writing with source integration. In the one-hour session, raters read the reading materials and prompts, and their attention was directed to the descriptors for task requirements, which were different from the first task. Similar to the first rating session, the sample essays were rated, and the criteria were set. The two raters and the researcher rated all the data, and the intraclass correlation coefficients among the three were calculated, which are indicated in Table 4.8. According to Koo and Li (2016), 0.50 to 0.75 indicates moderate level, and 0.75 to 0.90 indicates good reliability. Most of the reliability coefficients were at a moderate level. This is assumed to be due to the relatively long intervals (one to two months) between the rating session and the completion of the rating owing to the raters' work and research schedule. Although most of the reliability was at a moderate level, the average score of the three raters was used for the analysis.

Table 4.8

Intra-class Correlation Coefficients for the Functional Adequacy Rating

Scale	Without source integration	With source integration
Comprehensibility	0.58	0.64
Content	0.66	0.55
Task requirements	0.83	0.73
Coherence/cohesion	0.60	0.68

As for the statistical analysis of CAF measures and functional adequacy rating, the scores for the two tasks in each task type (without or with source integration) were averaged and the difference between the task types was compared by paired-samples t-test (Zhan et al., 2021) and Wilcoxon signed-rank test for the dataset that was not normally distributed. Effect sizes were interpreted as small $d = .40$, medium $d = .70$, and large $d = 1.00$ (Plonsky & Oswald, 2014) for paired-samples t-test and small $r = .10$, medium $r = .30$, and large $r = .50$

for Wilcoxon signed-rank test. By using Bonferroni correction, the overall alpha level was controlled ($\alpha = 0.05/10 = 0.005$); however, the results approaching the significant level were also considered to infer weak but possible results.

4.2.6 Interviews and Analysis

To investigate the trend of all the students, a questionnaire survey that enquired about the task difficulty and usefulness was conducted in the last class with 38 students (2 students were absent). In the questionnaire, they were asked to choose either the with integration task or the without integration task in terms of a higher difficulty or higher usefulness for language learning. In addition, their perceived usefulness of the course for English learning was assessed using a Likert scale (ranging from 1 = Disagree to 4 = Agree).

An interview is commonly conducted in qualitative second language writing research to gain the learners' emic perspectives toward writing (Polio & Friedman, 2016). To further investigate their perspectives on task difficulty and task usefulness for language learning and to supplement our understanding of the numerical results, 13 students were invited for a semi-structured interview at the end of the semester. They were also asked about their perception on the combination of content learning (about the SDGs) and learning of English writing. They were interviewed individually in Japanese for 16 to 32 minutes (M = 24.68, SD = 4.86) at the researcher's office, and the interviews were recorded and transcribed for analysis. The interview was conducted following the interview guide, which is shown in Appendix I. The guiding questions aimed to illustrate the learners' perceived difficulty and usefulness of the two kinds of tasks and their learning experiences in the CBI writing class. Example guiding questions included the following:

How did you find the difficulty of the two kinds of tasks?

How did you use the reading material? Did its use affect your writing and language learning?

How was your experience of learning about the SDGs and the writing task?

To help the students recall their writing experience, their writing samples were shown during the interview.

To answer research questions 2 and 3, the interviews were transcribed, and thematic analysis (Braun & Clarke, 2006; Terry et al., 2017), which is thought to be "a foundational method for qualitative analysis" (Braun & Clarke, 2006, p. 78), was conducted. This study utilizes thematic analysis as an analytic method since it takes an experiential orientation that seeks to identify what the participants think, feel, and do, and it is considered especially compatible with this orientation (Terry et al., 2017). The analysis followed the six phases suggested, which are shown in Table 4.9. The researcher started by reading the printed transcription, taking notes in the margins, and highlighting important ideas to be used for later analysis. Then, she imported the text data to a qualitative data analysis tool (MAXQDA) to assign as many semantic codes as possible so as to not limit the number of possible latent codes or themes. After assigning the semantic codes across the data set, they were compared against each other to form more overarching, abstract themes. The similarities among the codes and themes and the hierarchical relationships between the themes and the subthemes were considered. Themes were refined by reviewing the consistency among the codes within a theme and the distinctiveness between different themes, and the map of the codes and themes was visualized in a thematic map. Each theme was defined and named, and it was summarized with relevant interview quotes and descriptive codes in coding tables. As the coding process overlaps with that used in grounded theory (Strauss & Corbin, 1998; Corbin & Strauss, 2008; Saiki-Craighill, 2017), the essence of the coding strategies (open coding, constant comparative method, awareness about properties, and dimensions of a category)

were also referred to so as to deepen the understanding of the data. In addition, the use of interview data in second language writing studies, such as Neumann et al. (2019) and S. M. Lee (2020), and resource books, such as Polio and Friedman (2016), were referred to in deciding how to integrate the interview data in this study.

Table 4.9

Phases of Thematic Analysis in Braun and Clarke (2006)

Phase	Description of the process
1. Familiarizing yourself with your data:	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.
4. Reviewing themes:	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme.
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

4.3 Results

4.3.1 Results for the Two Types of Integrated Writing Analysis

To compare if there is any writing quality difference between the two types of integrated writing, the descriptive statistics for CAF measures are shown in Table 4.10. As for the time, increasing trends for the second writing task can be observed for both task types, except for the lexical sophistication of the with source integration writing task. In terms of the task type difference, similar figures can be seen for most of the CAF measures, but fluency is seemingly higher for the with source integration.

Before performing the statistical analysis, normality was checked by examining the kurtosis and skewness values and performing the Shapiro-Wilk test. Most of the data [90%, except for accuracy for the with integration task ($p = 0.003$) and the number of dependent clauses per t-unit for the without integration task ($p < 0.001$)] showed normal distribution (see Appendix J for kurtosis, skewness, and Shapiro-Wilk test result values for all the datasets).

Table 4.10

Descriptive Statistics of CAF Measures for the Two Types of Tasks

CAF measures (n = 27)	Without source integration		With source integration	
	1st writing	2nd writing	1st writing	2nd writing
Syntactic complexity (mean length of T-unit)	12.0 (2.13)	12.21 (3.23)	11.99 (2.0)	12.3 (2.52)
Syntactic complexity (no. of dependent clause per T-unit)	0.38 (0.26)	0.42 (0.45)	0.18 (0.13)	0.45 (0.23)
Lexical diversity (MTLD)	57.92 (15.37)	68.55 (17.55)	58.07 (17.02)	71.05 (24.20)
Lexical sophistication (LS-2)	0.16 (0.05)	0.18 (0.05)	0.19 (0.05)	0.17 (0.04)
Accuracy	0.80 (0.12)	0.84 (0.09)	0.78 (0.12)	0.81 (0.10)
Fluency (no. of words)	113.37 (28.31)	127.41 (31.98)	135.44 (38.47)	135.96 (29.42)

To statistically check whether there is any difference in CAF measures in terms of the task type difference, paired-samples t-tests were run. As for accuracy and the number of dependent clauses per T-unit, nonparametric Wilcoxon signed-rank tests were performed. The results are shown in Table 4.11. There was a statistical difference for fluency ($p = 0.0032$, $d = 0.57$), which shows a small to medium effect size. Participants produced more written production for the with source text integration writing. Although not at the significant level, lexical sophistication is slightly higher for the with source text integration writing ($p =$

0.08, $d = 0.45$) with a small to medium effect size. For the other measures (the mean length of T-unit, the number of dependent clauses per T-unit, lexical diversity, and accuracy), there was no significant difference between the tasks. No measures were higher for the without integration task.

Table 4.11

Results of Paired-samples T-tests and Wilcoxon Signed-rank Tests for CAF Measures

CAF measures (n=27)	Without source integration	With source integration	Paired-samples t-tests/Wilcoxon signed-rank test
Syntactic complexity (mean length of T-unit)	12.11 (2.20)	12.14 (2.01)	$p = 0.93$ Cohen's d = 0.02
Syntactic complexity (no. of dependent clauses per T-unit)	0.40 (0.29)	0.32 (0.16)	$p = 0.16$ $r = 0.27$
Lexical diversity (MTLD)	63.23 (12.47)	64.56 (17.33)	$p = 0.72$ Cohen's d = 0.09
Lexical sophistication (LS-2)	0.17 (0.03)	**0.18** **(0.03)**	**$p = 0.08$** **Cohen's d = 0.45**
Accuracy	0.82 (0.10)	0.80 (0.10)	$p = 0.27$ $r = 0.21$
Fluency (no. of words)	120.39 (25.03)	**135.70** **(29.53)**	**$p=0.0032$** **Cohen's d=0.57**

Next, the descriptive results of the quality of the writing assessed by the functional adequacy scale are shown in Table 4.12. All the ratings are in the range of 3 to 4, which indicates a somewhat unsuccessful to minimally successful performance level. The increasing trend between the first and second writing task cannot be seen for the rating although it is observed for the CAF measures. When the means of the two task types are compared, there was no significant difference for any of the four scales, as shown in Table 4.13. With or without integration of the source text did not bring about a noticeable difference in the ratings by human raters.

In summary, the task type difference is shown to have an effect on fluency, and a weak effect is expected for lexical sophistication. Rating results were not affected by the task difference. In the next section, the students' perceptions of these tasks elicited by interviews in terms of the difficulty and usefulness for language learning will be indicated. Moreover, their experience of learning about content and English writing in this CBI writing class will also be explored.

Table 4.12

Descriptive Statistics of the Functional Adequacy Scale Measures for the Two Types of Tasks

Functional adequacy scale (n=27)	Without source integration		With source integration	
	1st	2nd	1st	2nd
Comprehensibility	3.26 (0.62)	3.58 (0.54)	3.58 (0.52)	3.56 (0.59)
Content	3.58 (0.83)	3.57 (0.54)	3.60 (0.66)	3.48 (0.51)
Task requirements	3.72 (1.01)	3.89 (0.82)	3.79 (0.65)	3.67 (0.63)
Coherence and cohesion	3.49 (0.65)	3.60 (0.61)	3.57 (0.52)	3.36 (0.53)

Table 4.13

Results of the Paired-samples T-tests for the Functional Adequacy Measures

Functional adequacy scale (n = 27)	Without source integration	With source integration	Paired-samples t-tests
Comprehensibility	3.42 (0.48)	3.57 (0.47)	$p = 0.15$ Cohen's d = 0.29
Content	3.57 (0.61)	3.54 (0.45)	$p = 0.80$ Cohen's d = 0.05
Task requirements	3.80 (0.75)	3.73 (0.53)	$p = 0.63$ Cohen's d = 0.09
Coherence and cohesion	3.55 (0.52)	3.46 (0.44)	$p = 0.31$ Cohen's d = 0.20

4.3.2 Interview Results for the Student Perceptions of the Two Types of Integrated Writing Task

To answer the second research question (*How did the students perceive the difference between the two writing task types?*), the numerical results of the questionnaire survey are shown (see Table 4.14) to capture the whole class tendency, followed by the thematically coded interview results. These results are shown in Tables 4.15, 4.16, and 4.17 as suggested by Creswell and Poth (2018) to facilitate the visibility and understanding of the interview data.

Table 4.14

Questionnaire Results on the Difficulty and Usefulness of the Two Tasks

Question	With integration	Without integration	Both
Which writing task was more difficult for you, with integration task (Unit 6, 2) or without integration task (Unit 7, 9)?	22 (58%)	16 (42%)	0
Which writing task was more useful for your English learning, with integration task (Unit 6, 2) or without integration task (Unit 7, 9)?	25 (66%)	8 (21%)	5 (13%)

Table 4.14 shows that a slightly higher number of students (i.e., six students) perceived the with source integration task as more difficult, but both types of tasks were chosen by a decent number of students. This implies that each of them had different sources of challenges, which is confirmed by the interview results below. Conversely, when asked about the usefulness for English learning, 66% of the students chose the with integration task, which may reflect different characteristics of the two types (with or without source integration) of integrated writing used in this study.

Coding of the interview resulted in 492 descriptive codes, which were divided into task difference codes to answer the second research question and codes for the perception of the CBI course to answer the third research question. For the task difference codes, a total of 180 codes were assigned, of which 89 concerned the with integration task, 49 concerned the without integration task, 23 concerned both task types, and 19 were classified as others. The researcher repeatedly analyzed the initial codes to formulate the categories and themes, referring back to the original interview scripts in order for the categories and themes to represent the prominent cases in the interviews.

4.3.2.1 Themes for the With Integration Task

Codes were grouped into themes and categories according to the task that the codes were concerned with (i.e., with integration, without integration, or both). As for the with integration task, three themes (understanding of reading materials, reading material as a resource for writing, and influence on language learning) were obtained, with all the themes having two related categories. All the themes, categories, category definitions, example descriptive codes given by the researcher at the early stage of coding, and example quotes from the interviews are summarized in Tables 4.15, 4.16, and 4.17. Since the interviews and their analysis were conducted in Japanese, the researcher's descriptive code examples and students' interview quotes are shown in Japanese to respect their original nuance in Japanese. In-text quotes are presented in both English and Japanese. Quotes written in Japanese in the tables are translated into English and shown in Appendix K.

4.3.2.1.1 Theme 1. Understanding of Reading

The students frequently mentioned the importance of deep processing and understanding of the reading materials and their relationship to improved writing, which reflects the task characteristics. Theme 1 is further divided into two categories: 1-1. Deeper processing of reading materials and 1-2. The level of understanding affects writing. Four out of 13 interviewees answered that this task type requires them to read the source material more carefully for the later use of that information and to process it more deeply. Student F said, "For the with integration task, I need to understand the content of the textbook thoroughly (教科書のほうは、教科書の内容をちゃんと理解してないといけないし)." Student E also describes the reading process for the with integration task as follows: "I tried to understand the sentences line by line, checked the meaning of words, and wrote them down as I proceeded (こっちはめっちゃ、一行一行、文を理解してっていうか、単語調べて意味をちゃんと書いてって、やりました。)." From their explanation, it can be said that this student paid attention to the content of the reading and tried to understand it, which was to be referred to in the writing stage. Moreover, student B said that they understood and retained the content of the reading material for later use in writing: "The content of the textbook was in my head when I was writing (これを書いてる中で教科書の内容はちゃんと頭にはありましたね。)." Therefore, this task manipulation may have enabled the students to engage with a deep processing of reading, which may have positively affected their language skills development.

The other category highlights the dependence of writing quality on the understanding of reading. Five interviewees said that the difficulty of the reading material negatively affected their writing and, conversely, a better understanding of reading boosted their writing quality. Student C emphasized the role of reading:

If I understand the reading material, it makes writing easier. However, if I can't really understand it, I need to try to understand it fully. Otherwise, my citation makes little sense even if I try to integrate the source reading. (文書読んで理解できたら引用したら多分すごい書きやすいんですけど、読んでみていまいち分かんなかったら１回ちゃんと理解して読めるまでやらないと、多分引用してもよく分かんないことになっているかと思います。)

This student's quote highlights the necessity to understand reading to be able to integrate source information into writing; otherwise, according to student C, the low understanding level may lessen the writer's certainty about the appropriateness of source integration. In fact, student K stated that "unless you can read the material, you can't write (読みものを読んでないと書けないから書けないっていうか)" and, consequently, they could not write and cite the reading materials, thus experiencing enormous challenges in this task. In addition, it can be said that reading comprehension forms the foundation for the second (reading as a resource for writing) and third (impact on language learning) themes to positively affect the writers.

In summary, this task type may play a facilitative role to allow the students to become deeply engaged with the reading material and raise awareness about their understanding level of it. Moreover, the level of reading material understanding is a crucial element for task completion and lays the foundation for reading to work positively as a writing resource and for language learning.

4.3.2.1.2 Theme 2. Reading as a Resource for Writing

The second theme features the resource aspect of source reading. Ten interviewees mentioned that the reading materials served as an ideational as well as organizational resource for their writing. However, as it was raised by student K in the previous section,

whether or not students can use reading as a resource seems to be influenced by their proficiency and reading skills. In addition, it was pointed out that the requirement for reading integration adversely raised difficulty in organization for some writers. As for the first category, 2-1. Resource for writing content, most codes highlighted the positive role of the reading material in providing content for writing. Student G described the reading material as a "trigger (きっかけ)" for writing: "The textbook served as a trigger for creating ideas (アイデアのきっかけ、教科書がそういうきっかけになってると思うんで。)." Similarly, students G and I mentioned that they developed their own ideas by first obtaining information from the textbook. Therefore, the reading material functioned as a springboard for developing the essay. Moreover, having predetermined content to be included in the writing seems to have lessened the difficulty level of writing since "blocks for the writing are ready (こっちのほうがブロックがそろってる分、組み立てやすかったのかなって思います [student H])," and two students said they used textbook information as examples, which made the writing process easier (書きやすかった[student I, J]). Student B also said that it enabled them to increase "the amount of information (情報量)."

In contrast, two students, who also recognized reading as a resource, pointed out the downsides. One refers to the fact that the content to be written by the students is to some extent limited to the source text ideas. Student F said that the integration of the source text "limited how to start the writing, and it became difficult to expand on (the content) later (初めの書き出しが限定される感じがして、それは、後、広げにくいなって思ったりしてました)." Additionally, student D mentioned that using ideas from the source text caused them to think they were mimicking it, which included a slightly negative connotation (まねしてる感っていうか、ちょっと見ましたよ感が内容的にもあって、ちょっとまあまあまあみたいな). To summarize, the inclusion of a source text was mostly reflected as having a positive impact on

the enrichment of the content in writing by providing examples to be used and a trigger to develop their own ideas.

As for the second category, 2-2. Source of difficulty or support for the organization, both the facilitative and difficult aspects of source integration writing for the organization were mentioned. Three students pointed out that the organization of the source reading can be referred to when writing and supported the better organization of their own writing. Student C said that they "found and cut out an impressive part (from the reading) and based the writing around that part, which made the writing easier (自分の印象に残ったものを見つけて取り出して、そこから軸に書いていけるので書きやすくはあるのかなとは思いました。)." In the end, student D reflected that they were able to improve their skill to mix the source text information and their own opinion in the text.

In contrast, as student D perceived positive attainment through source-based writing, three students received it as a difficult aspect of this kind of writing task. Student M compared integrated writing with non-integrated writing and emphasized the difference:

> The process of integrating my own ideas with the source text information was difficult to some extent. If I wrote with only my own ideas, I could write smoothly from the beginning, but for this one, I needed to align with the cited sentence, which increased the time taken for thinking (文章に自分の考えをつなげるっていう工程がなかなか難しいんで。初めの段階から自分の意見だとすらすら書けるのが、こっちだと、引用した文とかに合わしてないといけないんで、その分考える時間が増えたかなって思いました。).

The quote suggests that the inclusion of source-text information raised the cognitive task difficulty in terms of organization, and accordingly, the thinking time and thinking for task completion increased. It was interesting that some students received the same task

manipulation as being supportive for writing while the other students pointed out its cognitively demanding aspect.

For the second category, it was confirmed that the reading material had provided ideational support for writing. As for organization, some students perceived it as support for their own writing while some others recognized it as a source of task difficulty.

4.3.2.1.3 Theme 3. Influence on Language Learning

As for the third theme, the integration of the source text seems to have created opportunities to deeply process the text by summarizing and paraphrasing it and develop various aspects of the students' language skills. The first category is 3-1. Processing language through paraphrasing and summarizing, which indicates the difficulty of paraphrasing and summarizing and learning through them. Two students specifically mentioned the challenges of modifying the source text for their writing: "I need to say things differently for citation, and that was difficult (やっぱり引用とかも言い換えなきゃいけないじゃないですか。それも大変だったなっていうふうに思います。Student A)." However, two other learners perceived the challenge of summarizing and paraphrasing as "gaining a skill to paraphrase (言い換える力がちょっとついた, Student D)" and "learning for myself because I need to express it in different words (別な言葉で言い替えるんで自分の勉強にもなるし, Student I)." Therefore, it can be suggested that the cognitive challenge induced by source integration caused the feeling of difficulty as well as provided a chance to expand their language repertoire through summarizing and paraphrasing.

The other category, 3-2. Positive impact on L2 learning, features students' comments about the various positive impacts on their English skills triggered by source integration. Four students pointed out the influence on fluency, which is in line with the increase in the number of words in the with integration writing task shown in the numerical results section. Student

G said, "Integrating opinions from the textbook and combining them with my own opinion increased the amount of writing. This (using a source text) was easier for me to write (教科書の意見を入れたほうが内容の量自体は自分の意見とプラスして結構書けたんで、そっちのほうがやりやすいかなと思ってます)." Thus, the ideas and examples provided by the reading helped the writers to write more. Two students mentioned the effect of the source text on their grammar use as student M "looked for useable sentences and grammar and changed it a little (for use) (いい感じの文章をというか、文法があるかを見て、で、それを自分なりに変えて、少し変えたぐらいなんで)." Student F reflected on the sentence length of the source text being "complicated (複雑)" and "long (長い)" and thought that "it would be better to write such (long) sentences, and I tried to write as such (そういう文、書かないといけないのかなって思って、長くしたりとかはしてた)." Therefore, the process of source text integration may have raised awareness about grammar and syntax for some students and changed their language use. A similar positive influence on vocabulary is also mentioned by two students: student E pointed out that they looked up the meaning of the unknown words, and student G "discovered (発見がある)" new vocabulary and expressions from the text and used them in the writing. Lastly, student L highlighted the benefits of this task type on reading and writing skills.

> When I cite, I need to read (the source text). Then my reading skill will improve, and if I can cite, I can improve my writing skill to write about it (引用するってなると、やっぱり読むじゃないですか。そしたら読む力も付くし、引用できたらそれについて書く力も付くのかなって思いました。).

This quote is also related to the first theme, that is, understanding the reading materials, but this student recognized the possible positive influence of source integration on their language

skill development. To sum up, the above quotes from the interviewees show the variety of perceived positive effects on language learning, such as on fluency, grammar, vocabulary, reading, and writing. The unique task condition, more specifically, the careful reading of the source material and rewording the part to integrate it into the writing, seems to have been perceived as an opportunity to promote a few students' language use.

In this section, the interviewees' comments on the with integration task were classified into three themes, all of which relate to the unique task demand caused by the integration of the reading material. Completion of this task may be affected by the understanding level of the reading material, but the presence of it seems to have supported idea generation and better organization (causing difficulty in the organization at the same time) and opportunities for language learning. The next section summarizes the comments on the without integration task and compares the results with those of the with integration task.

4.3.2.2 Themes for the Without Integration Task

As for the other task, that is, the without integration task, fewer codes were assigned than the with integration task (e.g., 49 vs. 89), and three themes were obtained through repeatedly assigning and modifying the semantic and latent codes (Terry et al., 2017): variability in the use of reading, impact on the organization, and impact on content. Overall, compared to the prominent role of the reading material emphasized for the with integration task, the level of reading material usage varied for this task as described in the interviews. Moreover, the interviewees' comments about content and organization contained positive and difficult aspects of this task type. Besides, even though there were comments on the language use in their writing induced by the requirement of source integration for the with integration task, no comments about their language use were obtained specifically for this task type. The following section will detail each theme with the interviewees' quotes.

149

Table 4.15

Themes for the With Integration Task

Theme	Categories	Definition	Example descriptive codes	Example quotes
1. Understanding of reading	1-1. Deeper processing of the reading materials	Deeper reading and understanding materials for task completion	引用ありは書くために教科書の内容が頭に入っている Reading の読み込みの深さが大事	これを書いてる中で教科書の内容はちゃんと頭にはありましたね。(student B) 文の意味、理解しないと入れようがないじゃないですか。だから、こっちのほうがちょっと調べるのに時間はかかったけど、いいんじゃないですかね。(student E)
	1-2. The level of understanding affects the writing	Better understanding of reading positively affects writing and vice versa	Reading を読めていないから書けない 引用ありの書きやすさは Reading の理解度による	僕、これ多分、入れてないんです。それ。ただテキストブックから引っ張ってきた記憶がないんです。(student K) 文書読んで理解できたら引用したら多分すごい書きやすいんですけど、読んでみていまいち分かんなかったら 1 回ちゃんと理解して読めるまでやらないと、多分引用してもよく分かんないことになっているかと思います。(student C)
2. Reading as a resource for writing	2-1. Resource for writing content	The use of ideas from reading in writing	書くとっかかりとしての引用の役割 引用ありは具体例として使える→書きやすい	教科書を読んだ上で、そこで得た知識を起点として自分の意見が浮かんでくるっていう。(student G) 具体的にあって、あれはそう言ってるからとかっていうふうに、何か、ちょっと例を挙げれるんで、書きやすいじゃないですか。(student J)
	2-2. Source of difficulty or support for organization	Reading materials can both support organization and make it difficult	Reading の論理展開を参照→つじつまが合うように書ける 引用ありの方が引用箇所と自分の考えをつなげる作業が難しい	こういうふうに文章構成したら全く一緒にはならないですけど、ちょっと話がつじつま合うように書けれるっていうのがあったんで。(student I) 文章に自分の考えをつなげるっていう工程がなかなか難しいんで。初めの段階から自分の意見だとすらすら書けるのが、こっちだと、引用した文とかに合わしてないといけないんで、その分考える時間が増えたかなって思いました。(student M)
3. Influence on language learning	3-1. Processing language through paraphrasing and summarizing	Difficulty of paraphrasing and summarizing and learning through them	引用ありだと paraphrase, summarize など書き換えが難しい 言い換えるのが自分の英語の勉強になる	やっぱり引用とかも言い換えなきゃいけないじゃないですか。それも大変だったなっていうふうに思います。(student A) 別な言葉で言い替えるんで自分の勉強にもなるし。(student I)

| | 3-2. Positive impact on L2 learning | Positive impact on fluency vocabulary, grammar, and reading/writing skills | 引用ありの方が長く書ける

引用ありは Reading を読む・書く力が付く | 作文をこなすんだったら、僕的にはこっちの、いっぱい量書けたほうが多分いいと思う、思ってるんで。(student H)
引用するってなると、やっぱり読むじゃないですか。そしたら読む力も付くし、引用できたらそれについて書く力も付くのかなって思いました。(student L) |

Table 4.16

Themes for the Without Integration Task

Theme	Categories	Definition	Example descriptive codes	Example quotes
1. Variability in the use of reading	1-1. Reading used for writing	Confirming the use of ideas in reading for writing	引用なしでも Reading を参考にした	私は一番最初のときかな、最初から割と教科書のことを自分の意見みたいに使ってしまっていて。(student D)
	1-2. Reading not used for writing	Using students' own ideas rather than ideas from reading	引用なしでは Reading はあまり使っていない。自分の経験と知識で書く	あんまり使ってなくて、自分の経験と今までの経験とか知識を基に書いてたと思います。(student G)
2. Impact on organization	2-1. Smooth organization	Writing flows smoothly by using students' own ideas	引用なしで自分の意見で書いた方が論の流れがスムーズ	自分が元々知ってたこととかで書き始めれるじゃないですか。だから内容を、じゃあこれ書いたら次はこう書いてっていって、つなげやすかったんですよ。(student F)
	2-2. Detrimental impact on organization	Increased organizational difficulty	書くことの整理に時間がかかる	何を選ぼうかっていう感じで、何ですか、自分の記憶とかから使えそうなものを探すっていうのが要るんで、ちょっと書くことの整理に時間かかったなっていう印象がちょっとありました。(student H)
3. Impact on content	3-1. Freedom in choosing the content	Less regulation positively perceived by students	引用なしは自由に書けるのがいい	自分の考えのほうはまだ縛りがあまりなくって、同じ単語を使わないようにもしないといけないけど、あまりそういう縛りがなかったから、自分の意見書くだけのほうがまだ多分書きやすかったです。(student B)
	3-2. Difficulty in creating content	Difficulty coming up with their own ideas for writing	引用なしは書く内容を考えるのが難しい 引用なしの方が考えることが多い（構成・単語・内容）	自分が意見を言うとかだったら、やっぱりその背景知識とかそういうのがないと書けないっていうのがあるし。(student A) こっちは自分で考えて書くから、こっちと比べて考えることが多い(student D)

151

Table 4.17

Themes Concerning Both With and Without Integration Tasks

Theme	Categories	Definition	Example descriptive codes	Example quotes
1. Reading material and the use of vocabulary/ expressions in it	1-1. The use of vocabulary/ expressions in reading for writing	Referring to reading for vocabulary/ expressions for both types of writing	Reading があることで新たな表現や文法を見つける Reading material から単語を拾って使う	表現ですかね、文法とか、こういう書き方もあるんだなとかですかね、多分(student L) 語彙とかもテキストからピックアップして、自分なりに文章がきれいにまとめれました (student M)
	1-2. Vocabulary/ grammar learning unaffected by reading	Vocabulary/grammar learning in writing unaffected by reading regardless of task types	文法力が増えたかは疑問、Reading があってもかわらない 語彙の学習については考えなかった	文法に関しては、そんな、読みものを読んだから、じゃあ、文法が書けるっていうのはなかった(student K) 語彙はやっぱりそんなにというか、やっぱり自分が使える英単語じゃないと使えないっていうのがあるじゃないですか。だから、それほど、何ていうんですかね、考えなかったっていうか、深くは考慮しなかったですね。(student A)
2. Perceived task similarities and differences	2-1. Similarities between the two tasks	Similarities across tasks in terms of the writing process, difficulty, and use of reading	読んで書くという点で引用ありもなしも同じなのでタスク差をあまり感じず 引用あり・なしの難しさは同等	自分の意見で書くだったらやっぱそれも結局は読んでみて理解した後にそこで自分がどう思ったかを書いていくって感じなんで。でも、どうなんだろう、結局同じように自分はやってたような気はしました。(student D) どっちも難しかった(student I)
	2-2. Difference between the two tasks	Differences across tasks in terms of the writing process and skills to be gained	引用あり・なしで書くプロセスが大きく異なる【差の程度】 引用ありなし鍛える部分が違う。どっちもいい。	もう、全然。こっちとこっちで全然書き方が、僕は、分かれました。全然違いました。(student M) どっちにも多分良さがあると思ってて。どっちかいいっていう感じではなくて、それぞれ鍛えられる部分というか、能力が違うかなって思いました(student J)

4.3.2.2.1 Theme 1. Variability in the Use of Reading

The without integration task did not require the students to cite specific information from the reading, and they needed to build on their ideas to construct the writing. Thus, the reading for this task type was provided as an awareness-raising material prior to writing. Given this task condition, five students said that they referred to the reading material even though they could complete the task without it. Student D said, "I tended to use the textbook (information) as my own opinion in my writing from the beginning (私は一番最初のときかな、最初から割と教科書のことを自分の意見みたいに使ってしまっていて。)." Therefore, for some students, reading materials may naturally serve as writing resources without specific instructions to use them.

In contrast, six students said they did not use the reading material for their writing, and their writing was unaffected by it for the without integration task. Student G mentioned that "I did not use (the reading material) that much, but wrote based on my experience and knowledge (あんまり使ってなくて、自分の経験と今までの経験とか知識を基に書いてたと思います。)." Student A also said, "I had some knowledge about the topic, so I didn't use the textbook that much (それについての知識がある程度あったからっていうのが大きいと思うから。それだったら、教科書はあんまり何か活用しなかったっていうのが大きいですね。)." Thus, some students relied more on their own knowledge for this task type. Although the task condition was the same across the participants, the task elicited different approaches to writing in terms of the use of reading. Compared to the with integration task, which is characterized by the notable importance of the reading materials understanding, this task elicited both the use and non-use of the reading materials. The availability of their own ideas for writing as mentioned by students A and G may have influenced their approaches to the task.

4.3.2.2.2 Theme 2. Impact on Organization

The second category is related to the organization of writing, which contains positive and negative reflections on this task type. Five students perceived this task type to have a better impact on their organization compared to the with integration writing since it was easier for them to write using their own ideas rather than utilize ideas from a different source. Student F described the ease as follows:

> I can start writing with my previous knowledge. So, it was easy for me to connect the content, like knowing what to write next (自分が元々知ってたこととかで書き始めれる じゃないですか。だから内容を、じゃあこれ書いたら次はこう書いてっていって、つな げやすかったんですよ。).

In comparison to the difficulty of combining the source text information and the writer's own ideas for the with integration task, student F and the other four students said that they felt the organization for this task type ran more smoothly. Student J also pointed out that coherence was better for this task type. Therefore, it seems to have been easier for some students to organize writing without source integration.

In contrast, two students highlighted the difficult aspects of this task type on the organization. Student H said that choosing information from their memory bank and organizing it for writing was difficult.

> It was necessary to search for information in my memory to use it for my writing, and it took time to organize the content before writing (自分の記憶とかから使えそうなもの を探すっていうのが要るんで、ちょっと書くことの整理に時間かかったなっていう印象 がちょっとありました。).

154

Additionally, Student F, who mentioned the positive impact of this task type on the organization, pointed out a difficulty as well. They said that it was difficult to narrow down the point in writing, and their writing became scattered. Student F reflected: "I didn't go into detail for one point. I wrote a little on one point and then moved on to another (1 個のことをあんまり掘り下げてなくて、それで、何かちょっと書いて、また次の例にいってみたいな感じで書いてて)." They also mentioned that they were able to write in detail on one point for the with integration task; hence, it can be assumed that the source text helped them explain the main point in greater detail. Therefore, according to each student, the without integration task seems to have both positive impacts and difficulties in organization.

4.3.2.2.3 Theme 3. Impact on Content

The last theme is the impact on content, which includes the following two categories: 3-1. Freedom in choosing content and 3-2. Difficulty in creating content. As in the previous section, the interviewees pointed out both the positive and difficult aspects of this task type in terms of content. Four students talked favorably about the freedom in choosing the content of the writing. Student D first experienced some difficulty as they needed to come up with the content themself but explained that ideas developed freely in their mind once they started writing.

> Once I started thinking, it was unexpected, but the ideas expanded in my head, and I was able to write freely. When I started writing, the ideas flowed freely (1 回考えてみたら、案外頭の中でアイデアが膨らんで自由に書けるから、書き始めたらつらつらつらっていろんなことが出てきたりして).

In contrast, student B focused on the non-requirement of citation, which they thought made the task doable.

> For the without integration task, there were fewer requirements (For the with integration task) I was not allowed to use the same words (from the reading material), but there were no such requirements in this task. Hence, this task type was easier for me to write than the other one (自分の考えのほうはまだ縛りがあまりなくって、同じ単語を使わないようにもしないといけないけど、あまりそういう縛りがなかったから、自分の意見書くだけのほうがまだ多分書きやすかったです。).

The key words are "freedom (自由)" and "no requirements (縛りがない)" when the interviewees explained the positive side of this task type. Since this task type did not induce the cognitive burdens related to restrictions about source integration, some writers may have considered it as more doable.

The opposite side of the same phenomena was described as the difficult side of this task type, which is the challenge in creating their own content. Owing to the freedom it offered in terms of the content to be written compared to the other task type, six students felt that it was difficult for them to gain ideas for writing. Student A pointed out the necessity of background knowledge for writing their own opinion: "When I need to write my opinion, I can't write without background knowledge (自分が意見を言うとかだったら、やっぱりその背景知識とかそういうのがないと書けないっていうのがあるし。)." Student I said, "I needed to think to come up with the solution, and it took a lot of time (自分で考えないとちょっと解決策出てこないんで、その探す時間が余計に増えちゃって)," thus pointing out the time required for idea generation. Student G also pointed out that it was difficult for them to write more since they could not necessarily express all that they wanted to in English: "Only with my own opinions,

it is sometimes difficult to express them (in English). So, I can't write much (ちょっと自分の意見だけだと、なかなか表現しづらいときとかもあるんで、なかなか量が書けないというか)."

As shown in these quotes, the without integration task elicited both positively and negatively charged comments on the content aspect of writing. The positive side was characterized by the freedom in choosing the content while the negative side was indicated by the content creation. As the following student J's comparison of the two task types suggests, whether the writers have something to write about or not may influence their perceived difficulty of task.

> When I can come up with many ideas, this one (the without integration task) may be easier to write. (…) When I can't think of any ideas, writing with examples in the textbook may be easier (自分でたくさんアイデアが出るときは、こっちが書きやすいかもしれないです。[…] 思いつかないときは、教科書の例を入れながら書くほうが書きやすいかなと思いました。).

4.3.2.3 Themes Concerning Both Types of Tasks

Some of the comments concerned both task types, and such comments have been summarized in this section. There are two themes, namely reading material and the use of vocabulary/expressions in it and perceived task similarities and differences.

4.3.2.3.1 Theme 1. Reading Material and the Use of Vocabulary/Expressions in it

The first theme addresses the effect and non-effect of the reading materials on writing. As for the first category, 1-1. The use of vocabulary/expressions in reading for writing, three students said that they used the vocabulary and expressions in the reading material for both tasks. Student M said, "I picked up vocabulary from the textbook and was able to write my

essay (語彙とかもテキストからピックアップして、自分なりに文章がきれいにまとめれました。),” referring to both the with and without integration tasks. Therefore, for some students, the use of vocabulary occurred regardless of the task type.

With regard to the other category, 1-2. Vocabulary/grammar learning unaffected by reading, three students mentioned that they did not feel that they learned certain vocabulary or grammar even though they read the reading material before writing for both task types. Student K described, “About grammar, reading a material did not simply lead to the use of that grammar in writing (文法に関しては、そんな、読みものを読んだから、じゃあ、文法が書けるっていうのはなかった).” Student M also doubted the increase in grammar. In addition, student A said they did not think about vocabulary learning because they could only use their productive vocabulary.

> Vocabulary was not influenced that much (because of the use of the reading material).
> I can only use vocabulary that I know how to use. So, I didn't think or deeply
> consider (the influence on vocabulary) (語彙はやっぱりそんなにというか、やっぱり自
> 分が使える英単語じゃないと使えないっていうのがあるじゃないですか。だから、考え
> なかったっていうか、深くは考慮しなかったですね。).

Therefore, regardless of the task type, some students did not feel that the series of writing activities had a positive influence on the vocabulary or grammar learning. Although there were positive comments regarding these aspects of learning for the with integration task type discussed in the previous section, it seems that some variations exist.

4.3.2.3.2 Theme 2. Perceived Task Similarities and Differences

The second category is the perceived task similarities and differences, which is divided into two sub-categories: 2-1. similarities of the two tasks and 2-2. differences of the two tasks. The first concerns the similarities across the tasks in terms of the writing process, difficulty, and use of reading. Seven students mentioned similar points, and Student C explained that the process of writing in both tasks was similar: "after all, (both tasks required) reading and understanding it, and then I wrote about what I thought. So, I think I performed the task in the same way (結局は読んでみて理解した後にそこで自分がどう思ったかを書いていくって感じなんで。でも、どうなんだろう、結局同じように自分はやってたような気はしました。)." Student D also said that they gained knowledge through reading and used it as a foundation of writing for both tasks. In terms of difficulty, two students said (students E, I) that both had the same level of difficulty, and two others (students A and J) mentioned that their level of reading material understanding was the same across the tasks.

In comparison, various differences were raised by most interviewees and were coded separately in each theme for the with and without integration task sections. Therefore, the differences will not be listed in this section. One illustrative explanation was provided by student M saying that, "I wrote differently for the two tasks. It was completely different (もう、全然。こっちとこっちで全然書き方が、僕は、分かれました。全然違いました。)." Student M used the grammar in the source text in the writing for the with integration task, but no source text information was reflected in the writing for the without integration task. As student J summarized, each task trains different aspects of writing (e.g., using the without integration task when practicing writing about own ideas in English and using the with integration task when practicing writing a term paper); therefore, student J said, "Maybe both have good points. It's not that either one is good, but the aspects or skills to be trained are

different (どっちにも多分良さがあると思ってて。どっちかいいっていう感じではなくて、それ ぞれ鍛えられる部分というか、能力が違うかなって思いました。)."

4.3.2.4 Synthesis of the Task Type Difference Results

So far, the students' interview results about the perceived differences between the with integration task and the without integration task were categorized into representative themes and explained using descriptive quotes. As suggested by Terry et al. (2017), thematic maps were created to compare the characteristics of the comments on the two kinds of writing tasks. Upon creating the thematic maps that graphically represent the main findings of the thematic analysis, all the themes, categories, interview quotes, related raw data, and written descriptions were reread.

Figure 4.1 shows two thematic maps for the with and without integration tasks. The map on the left side depicts the relationship of the themes for the with integration task. Reflecting on the crucial role of reading materials and understanding of it for task

Figure 4.1

Thematic Maps of the Two Writing Tasks
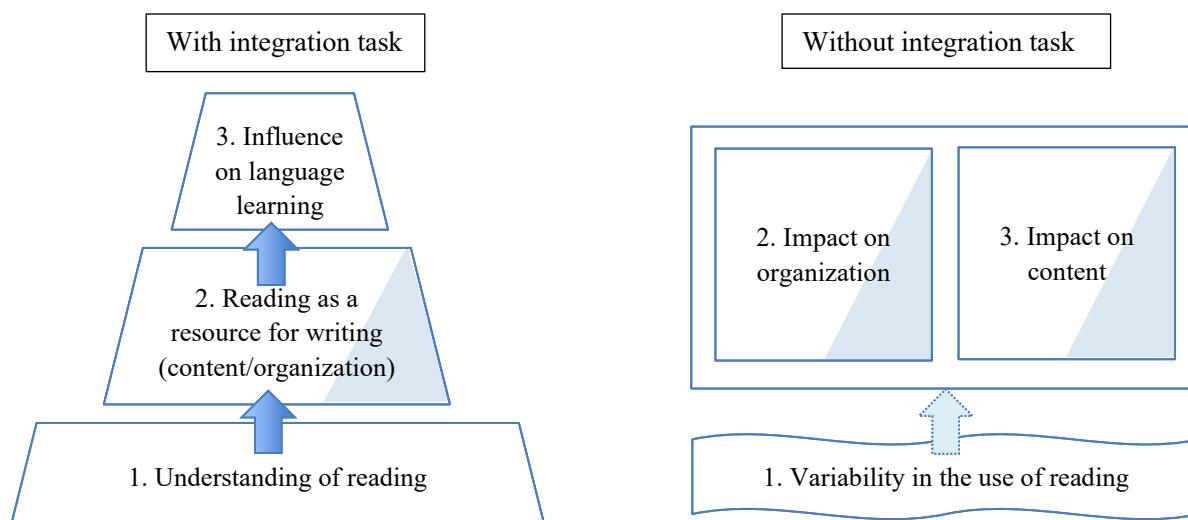
completion, the first theme, 1. Understanding of reading, is placed at the bottom as a base and affects both themes 2 and 3. Above theme 1 is theme 2, that is, reading as a resource for writing. If the reading material is understood, it can serve as a resource for content creation and refining the organization of writing. In contrast, as indicated by the shadowed triangle, the integration of information from reading can cause difficulty in the organization since the students have to combine their own opinions and source information. Through the process of source integration for the content and organization of writing, some students said that they processed the reading material more deeply; used words, expressions, and grammar in the source reading; and improved their reading and writing skills. Therefore, theme 3 comes above themes 1 and 2. However, there were also a few students who stated that their writing was unaffected by the presence of reading materials in terms of grammar and vocabulary; therefore, the shape of theme 3 is indicated as being smaller than that of theme 2. From the thematic map, it can be noted that the requirement of the source integration can make the task revolve around the reading and that may affect the content and organization of writing and their language learning experience through the task.

Conversely, when the relevant reading material is present but the task does not necessitate information from it, the students' comments were focused on both the positive and difficult sides of this writing task in terms of content and organization. As in the other task, the shadowed triangle represents that there were both positive and negative comments. The students' independent thinking for completing the writing tasks may have been highlighted for this task type. Although the relevant reading material was read before writing, its use varied depending on each writer: some used the ideas and vocabulary in it while the others did not. This variability of reading use is indicated as a wavy square below themes 2 and 3, and its possible but variable influence on learner writing is indicated as a dotted arrow which begins from theme 1 and targets themes 2 and 3.

In summary, this task manipulation created different writing and learning opportunities for the students. By knowing these differences perceived by the learners, teachers may be able to make a reasonable decision when choosing a task for their CBI/CLIL class in which the students' language output tends to be based on the content covered in class. Although the study tried to identify the dominant task perception patterns through the thematic analysis, there seems to be an individual difference as to how they perceive the tasks according to their proficiency levels, writing strategies, and motivation for English learning as the students' comments in Tables 4.15, 4.16, and 4.17 suggest. Further analysis may need to take these individual differences into account.

4.3.3 Interview Results for the Student Perceptions of the CBI Writing Course

This section summarizes another part of the interview results, which are the students' comments on the current CBI course. The third research question for study 2 was as follows: *How did the students perceive the combined learning of content (SDGs) and English writing?* There were 164 initial codes concerning this research question, and following a similar method to answer the previous research question, these codes were classified to form themes and categories according to Braun and Clarke (2006) and Terry et al. (2017). With the careful reading of the raw interview data and comparisons among the themes and categories, two themes and seven categories were obtained, which are shown in the coding table (see Table 4.18). As with the task difference interview results, the researcher's descriptive code examples and students' interview quotes are shown in Japanese to share their original nuance. In-text quotes are shown in both English and Japanese. English translation of Table 4.18 is shown in Appendix L.

4.3.3.1 Theme 1. Difficulty due to Content, Language, and Both

The first theme is 1. Difficulty due to content, language, and both, which contains 67 descriptive codes. All interviewees stated that the course was challenging. The source of difficulty lies in the novelty of the content, linguistic challenge, and combination of these two elements. Three categories below theme 1. reflect these three points. The first category is 1-1. Difficulty arising from the novelty of the content, which contains comments from six students that point to the newness of the content, academic nature of this topic, and difficulty in learning and writing about it even in Japanese. Student I said, "I have not thought about deforestation, so that was difficult. Also, I haven't thought that much about biodiversity ([deforestation について] そこまで気にしたことなかったんで、それが難しかったし。あと、なんか絶滅の話も、[...] あれもそこまで考えたことなかったんで。)." It can be said that the novelty of the content adds to the students' perceived difficulty of the course, thereby suggesting that understanding of the content should be supported by, for example, adjusting the language level, using visual aids, and using L1 supplementally. Furthermore, student G pointed out that the academic nature of this content adds to the difficulty ("The SDGs itself have a social aspect and are difficult [SDGs 自体、結構社会的で難しい内容。]"), and student F said that they checked the concept of the content in Japanese first to further understand it ("When I spoke with my partner, we first checked how to say the concept in Japanese and then thought about how to say it in English [日本語で何て言うんだっけみたいになって、そもそも。それで、さらに英語にしたら何て言うんだろうみたいな話はしてました。]"). Since the content was novel, students' L1 seems to have worked as content scaffolding. Nevertheless, two students (Students K and L) said that it felt challenging for them to learn about the content even in Japanese ("I found it difficult to write about the SDGs in Japanese [日本語で書くのは難しいと思ったんで、SDGs について。]," student L). From these comments, it can be inferred that the newness of the content can be a source of difficulty in CBI.

The second category focuses on the language aspect of the course: 1-2. Difficulty arising from writing and reading skills. Specifically, it highlights the lack of experience in writing and difficulty with reading due to unfamiliar words, which were mentioned by 11 students. The current participants claimed that they had had only a few opportunities to write in English before taking the course. Student J said, "I think the difficulty was, rather than the theme, simply writing in English (難しかったことは、テーマというよりも、シンプルに、英語を書くのが難しかったと思います。)." Similarly, student B realized that the difficulty arose from the fact that they had forgotten the grammar: "English writing was difficult because I didn't remember the grammar. It took a long time to complete one task (英語の作文はちょっと文法とかも全然覚えてなかったから結構難しかったというか、時間はすごいかかりました。1つ書くのに。)." Moreover, their lack of writing experience caused feelings of their struggle. Student C said, "I tried hard to write, but I couldn't do it well (頑張って作って、でもいい感じにできなくてみたいな感じになってた。)." In addition to the difficulty that arose from writing, unfamiliar vocabulary in reading also seems to have raised the difficulty level, which was mentioned by seven students. Student C explained the difficulty in relation to the specificity of vocabulary for content learning: "There were many specific terms, and I felt difficulty at first (専門というか難しい英単語が結構あって、最初は結構難しいなと思ってました。)." In addition to the unfamiliar words, Student J said that grasping the meaning of a whole sentence was also challenging: "I looked up the words and combined them to understand them in Japanese, but there were many sentences I couldn't understand even in Japanese (言葉、調べてつなげて、日本語にしたときに、日本語でも、これ何、どういう意味なんだみたいなのが多くて。)." Therefore, it is assumed that both the vocabulary and sentence structures of the reading were difficult for the students. It is suggested that the control of the language level for the materials is crucial, and it needs to be thoroughly verified in planning the CBI course, keeping in mind that the focus on content learning may heighten the language level.

Additionally, since the course aimed to practice writing based on the reading about the SDGs, these two language skills were predominantly mentioned.

The third category, 1-3. Difficulty arising from the combination of content and language learning, concerns the increased difficulty of learning about new content and writing about it in an unfamiliar language, which was mentioned by nine students. Student J said, "It was difficult to learn new content and write about it at the same time (新しい知識も学びながらそれを英語で書くっていうのは、難易度は高い。)," highlighting cognitively demanding nature of CBI. It was mentioned that the difficulty of the content and of English mutually influenced each other. Student F felt that lack of content knowledge created issues in reading in L2:

> I didn't know about the SDGs at all at the beginning, so I didn't understand the content by reading about it in English. It was very difficult (SDGs の内容を元々全然知らなかったので、初めは内容を英語で読んでも分からなかったので、すごい難しかった。).

From this example, the significant connection between content and language in CBI learning can be observed. Another example of this link is pointed out by student K, saying that even a common word has a specific meaning in the content field, and they were unable to understand the concept by looking up the word in a dictionary, thus adding to the perceived difficulty of this type of instruction. Specifically, the example that student K was talking about was active and passive solar, which refer to the different ways of harnessing solar power (i.e., with or without other devices to distribute solar energy). Student K said,

Active, static, and dynamic energy, I can't understand these without looking them up. I can't understand proper nouns, and I don't know the meaning behind these proper nouns (active とか、静的な、動的な、何かエネルギーみたいなやつも、調べないと出てこないから、あんなの。固有名詞が分からない上に固有名詞の裏の意味も分からないから。).

Therefore, this example suggests that the field-specific use of vocabulary increased the perceived difficulty of the content and English understanding. Students in a language classroom where content is not necessarily focused on may not face this kind of content-related difficulty, but it was noted in the current CBI course. All in all, the students experienced challenges in the current course in terms of content, English, and the synergy of these two aspects.

4.3.3.2 Theme 2. Positive Effects of Having Content Focus for Language and Affect

The second theme is 2. Positive effects of having content focus for language and affect, which contains 99 descriptive codes. Despite difficulties due to content integration and language required for learning about it, the students reflected on the various positive aspects of the course. These are mainly related to the effect on language learning and students' affect, and two categories for each (four categories in total) were generated.

The first category is 2-1. Positive effects on writing and reading by having the SDGs as content and a comparison with familiar topics. Although it was pointed out in the previous section that the learners perceived difficulties with the language used in CBI, the integration of the content into language learning can also be perceived positively and affect their language skills, such as writing and reading. Student J suggested that gaining knowledge about the writing topic supported their writing in L2.

166

I think if there is that kind of theme, then it makes writing easier, even in English. Learning about a new thing and keeping it in mind while writing makes writing easier (そういうテーマ、あったほうが書きやすいと思います、英語とかでも。そういう新しいことを学んで、そういうのがあるんだって思って書いたほうが書きやすいかなとは思いました。).

According to student J, even in an English class where the primary aim is to develop the students' language ability, focusing on learning about new content can benefit their use of English. Moreover, student I commented that having a sustained theme (i.e., the SDGs) set the direction for writing and supported their writing process, and student B favorably perceived the SDGs as a theme because it enabled them to search for various information to be included in the writing. In addition, students D, F, and G pointed out that the continuation of the content learning over a semester enabled them to reuse the content and vocabulary from the previous writing to the next one, thus making it easier to handle the task. Student D explained the connection among the topics and how that made writing slightly easier.

I started with no knowledge of the SDGs. When I studied deforestation, the content reappeared in the ocean topic, for example. They share common topics, so it was a little easier to write (全然知らない状態から、例えば森のことだけを勉強してもちょっと海のところとつながってる、同じようなとこがあったりして、［…］ちょっと書きやすかったです。).

As the comment suggests, learning about related topics over a certain period of time (e.g., the whole semester) in the CBI course may facilitate the repeated use of vocabulary and content concepts, and thus may make their writing experience doable.

Table 4.18

Themes about the Students' Perceived Learning Experiences in the Current CBI Course

Theme	Categories	Definition	Example descriptive codes	Example quotes
1. Difficulty due to content, language, or both	1-1. Difficulty arising from the novelty of the content	Difficulty due to the newness of the content, even in Japanese	SDGs 自体、結構社会的で難しい内容 日本語でさえ SDGs について書くのは難しい	SDGs 自体、結構社会的で難しい内容なので (Student G) 日本語で書くのは難しいと思ったんで、SDGs について(Student L)
	1-2. Difficulty arising from writing and reading skills	Lack of experience in writing and difficulty of reading due to unfamiliar words	難しさに影響するのは内容よりも英語で書くことに慣れていないこと 書くための文法を忘れている Reading の単語が難しい	難しかったことは、テーマというよりも、シンプルに、英語が、書く、英語を書くのが難しかったと思います。(Student J) 英語の作文はちょっと文法とかも全然覚えてなかったから結構難しかったというか、時間はすごいかかりました。1 つ書くのに。(Student B) 専門というか難しい英単語が結構あって、最初は結構難しいなと思ってました。(Student C)
	1-3. Difficulty arising from the combination of content and language learning	Increased difficulty of learning about new content, and writing about it in an unfamiliar language	CBI は難易度高い SDGs を知らないから Reading が難しかった 固有名詞の表す概念が難しい	新しい知識も学びながらそれを英語で書くっていうのは、難易度は高い(Student J) SDGs の内容を元々全然知らなかったので、初めは内容を英語で読んでも分からなかったので、すごい難しかった(Student F) 単語は調べたら出てくるけど、固有名詞が難し過ぎてよく分かんなかったし、言い回しが難しかったから。(Student K)
2. Positive effects of having content focus for language and affect	2-1. Positive effects on writing and reading by having the SDGs as content and a comparison with familiar topics	Connection between content learning and writing/reading development and comparison of learning with the use of familiar topics	新しい内容の知識を得て書くほうが書きやすい 内容の Theme 同士のつながりがある→書きやすさにつながる	そういうテーマ、あったほうが書きやすいと思います、英語とかでも。そういう新しいことを学んで、そういうのがあるんだって思って書いたほうが書きやすいかなとは思いました。(Student J) 全然知らない状態から、例えば森のことだけを勉強してもちょっと海のところとつながってる、同じようなとこがあったりして、[…]ちょっと書きやすかったです。(Student D)

			正確に読み取るよりも雰囲気で読む（あいまいさに耐える）	文法は、正直言うとなんでこの文がこうなるかあんまりよく分からんみたいなとこもあったんですけど、そこはなんとなくで感じる感じで、単語を見て大体こういう意味かなっていうふうにやってました。(Student D)
			身近な話題だと単調な文になって広げられない	好きな食べ物とか、そっち系だと単調な文になっちゃいそうだなと思って。あんまり深く掘り下げれないじゃないですか。(...) 文を広げるのが難しいんじゃないかなと思って。(Student E)
	2-2. Wide range of vocabulary input	Various vocabulary input and learning of it through repeated exposure	SDGs の方が単語のインプットが多い。	今までは習ってきた単語とか少なかったんですけど、(...) いろんな文書に触れることで結構いろんな、こんな単語あったんだっていうのが結構思いました。(Student C)
			繰返し単語が出てきて覚える	文法とか語彙も新しい言葉がすごいいっぱいで、何回も使っていくことで自然と身に付いている語彙、言葉もあ（った）(Student D)
	2-3. Value of learning about the SDGs	Expressing values of learning about the SDGs and appreciating the connection between the students' previous learning and the SDGs	新たな内容を知れることがよかった	SDGs もそれまで知らなかったので、自分にとっては結構プラスな面が多かったかなと思ってます。(Student H)
			過去 SDGs について少し習った経験とつながった。	4年生になって最初の専門科目で SDGs ちょっとさらっと習った程度で、そこまで深く知らなかったんで。それで英語の授業で SDGs を絡めて学習できてんのが、すごく勉強になったなっていう。(Student I)
	2-4. A sense of achievement	A sense of achievement learning about the SDGs and writing about them	難しかったがかなり力はついた	SDGs 自体、結構社会的で難しい内容なので、それに加えて普段、英語で自分の意見を書くっていうことをなかなかしないので、それが合わさってかなりハードな内容だと思ってましたけど、その分かなり力は付いたなと思ってます。(Student G)
			頑張ったという思い	自分の背景知識をフルに生かして、頑張っていけたから、結果としては良かったんじゃないかなっていうふうに思います。(Student A)

Next, two students mentioned that their reading skills also increased in relation to content learning. Student G especially noted that their reading skills improved, and it became easier to read college transfer exams. Student D reported that the type of reading (i.e., explanatory) was different from the materials they had read before (i.e., narrative). Due to this difference, student D explained that their way of reading in this course changed, stating that they endured the ambiguity of meaning and tried to guess the meaning of the words instead of focusing on an accurate understanding of the reading.

> For grammar, I didn't completely understand why the sentence was as it was. But when I was reading, I tried to feel the nuance, and I looked at words and guessed the meaning (文法は、正直言うとなんでこの文がこうなるかあんまりよく分からんみたいなとこもあったんですけど、そこはなんとなくで感じる感じで、単語を見て大体こういう意味かなっていうふうにやってました。).

It may be suggested that reading materials that were full of information for content learning and were relatively more difficult than the reading students had read may have required different approaches to reading.

Differences between the current CBI topic (the SDGs) and familiar topics were also compared by six of the students in the interview. Student E remarked that the current topic had depth and breadth to write about compared to cognitively less complex and familiar topics.

> I think topics like my favorite food will end up as plain sentences. I can't dig into these topics deeply. (…) It may be difficult to develop writing with these topics (好き

な食べ物とか、そっち系だと単調な文になっちゃいそうだなと思って。あんまり深く掘り下げれないじゃないですか。[…] 文を広げるのが難しいんじゃないかなと思って。).

Therefore, to practice writing a fully developed essay, a topic that offers depth is perceived to be suitable. Student E also mentioned that the SDGs as the writing topic were far better than cognitively less complex familiar topics for their learning. In contrast, four students favorably perceived familiar topics in terms of the degree of difficulty. Student K said, "I feel like I can write if the topic is something like what I did during the weekend, like junior high school students write about, and develop on this topic (週末何したかとか、中学生がやる英作文みたいなトピックで、文を広げてみたいなライティングだったら、まだ書けるかな。)." Therefore, some students felt that the current writing on the SDGs benefited their learning in terms of content and language knowledge expansion while others noticed that familiar topics could lower the difficulty level and encourage them to write. Although there were some comments that favor cognitively less demanding familiar topics to mitigate the perceived difficulty in writing, the beneficial effects of learning about new content for a period of one semester on writing and reading were also mentioned by the students.

The second category is 2-2. Wide range of vocabulary input, which features various vocabulary input and learning of it through repeated exposure in CBI. Seven students mentioned the breadth of vocabulary and effect of the repeated use of these words on their learning. Student C said, "I had learned only a few words before. But by reading various texts, I found that there were various new words (今までは習ってきた単語とか少なかったんですけど、[…] いろんな文書に触れることで結構いろんな、こんな単語あったんだっていうのが結構思いました。)." Since EFL textbooks that the students had used before the current CBI course limited the range of vocabulary for understanding, they may have felt that the materials (e.g., reading, handouts, slides, videos) for content learning contained a wider range

of vocabulary. They also mentioned the type of vocabulary they gained, which was more content-specific (e.g., words used for describing environmental problems). Student J said that their receptive vocabulary specific to the content topic may have increased: "There were newly-learned content-specific words, although not many, and the number of such words that I can recognize may have increased (自然系の言葉とかは、たくさんじゃないですけど、新しく知った単語とかも多かったので、そこは、ちょっとは、この言葉、聞いたことあるなぐらいが増えたのかなと思います。)." As related content was dealt with for a semester, repeated exposure to certain vocabulary enabled students to memorize some of these words. Student D said, "There was a lot of grammar and new words, and some of the words were naturally learned by using them repeatedly (文法とか語彙も新しい言葉がすごいいっぱいで、何回も使っていくことで自然と身に付いている語彙、言葉もあ[った])." Thus, the interview comments suggest that the focus on content enabled a wider range of vocabulary, and the repeated use of these words helped students increase their receptive vocabulary.

The third category is 2-3. Value of learning about the SDGs, which is defined as expressing values of learning about the SDGs and appreciating the connection between the students' previous learning and the SDGs. Almost all students (except student K) mentioned that they genuinely valued learning about the topic, expressing their positive feelings with such words as "interesting (面白かった, students C, D, J)," "encouraging (やる気がわく, student J)," and "important (有意義, student G)." Evidently, the engagement level and usefulness of the content itself had influenced their class perceptions. For instance, student H positively reflected on their experience of gaining new knowledge: "I didn't know much about the SDGs, so I think there were many positive aspects (SDGs もそれまで知らなかったので、自分にとっては結構プラスな面が多かったかなと思ってます)." More specifically, student G observed that the knowledge of the SDGs will be "useful when I start working (社会に出て就職するときとか、そういうときに役に立つ)," and student M also said that the topic "is

attracting attention in various places (いろんなところで注目されている)," and it is thus

beneficial to know about it "from a global perspective (グローバル的な観点から見ても)."

While acknowledging the challenge of writing about newly-learned content, student J

mentioned that having a decent topic to write about encouraged them to learn.

> I feel that it (the course) was better than only learning about English. It was difficult
> to gain new knowledge and write about it in English, but personally, I'm encouraged
> if there is a theme to be learned, so it was good (ただ英語だけを単独で学ぶよりは、や
> りやすかったかなと思います。新しい知識も学びながらそれを英語で書くっていうの
> は、難易度は高いですけど、個人的にはそういうふうに何かテーマがあったほうがやる
> 気とかも湧くんで、良かったと思います。).

Therefore, learning meaningful content may positively affect students' minds about learning

in a language classroom.

    Moreover, the students felt the connection between the topics of the SDGs and their

previous learning/everyday life experiences, which led them to feel positive about the course.

Eight students reflected on this connection favorably. Student I highlighted their experience

of learning about the SDGs further in the current English course.

> I had learned only a little about the SDGs in the first technical course of my
> department in the fourth year, so I didn't know much about it. Then, I felt I learned a
> lot in the English class which incorporated learning about the SDGs (4年生になって最
> 初の専門科目でSDGsちょっとさらっと習った程度で、そこまで深く知らなかったん
> で。それで英語の授業でSDGsを絡めて学習できてんのが、すごく勉強になったなって
> いう。).

As student I said, when the topic dealt with in the English class and the students' background learning experiences match, that may benefit their affective state for learning. In addition, it is assumed from student I's comment that learning about content, not about the language itself, in a language class may have been perceived as a positive surprise. Other students (students B, C, E, and G) stated that they found a connection between the current SDG topics and their everyday life experiences, and that the relationship seems to have become their motivational support. Student E said they realized that their way of viewing companies had changed after the class and reflected on the class saying, "It was hard but good (大変だったけど、良かったです。)."

> By learning (about the SDGs) in class, when I see their marks on the companies' homepages, I understand how they may be dealing with them (social problems). My way of looking at companies has changed (授業ですることで、他の企業とかの見たりするじゃないですか。ホームページとか。あれとかを見るときに、SDGs とかのマークがあったら、あ、ここはこういうことを多分やってるんだろうなみたいな感じで、見方が変わってきました。).

As suggested by the quotes, the content itself plays an important role and carries value in language-driven CBI for effectively supporting and promoting students' learning. Furthermore, it was pointed out that if the connection between the topic in the CBI course and the students' background and everyday experiences was felt, it may have a positive impact.

The last category is 2-4. A sense of achievement, and it was mentioned by five students. These learners explained their learning experiences as having a sense of achievement in learning and writing about the SDGs. As depicted in the comments in the first

theme, paying dual attention to content and language raised the perceived difficulty level of the course for all students; however, some of them reported their experiences of making efforts and feeling of a sense of accomplishment. Student G summarized their experience as having both challenging and beneficial aspects.

The SDGs itself is a social and difficult topic. In addition, I usually don't write my opinions in English; hence, I felt that this combination made the course very hard. But, I think I gained a lot of skills (SDGs 自体、結構社会的で難しい内容なので、それに加えて普段、英語で自分の意見を書くっていうことをなかなかしないので、それが合わさってかなりハードな内容だと思ってましたけど、その分かなり力は付いたなと思ってます。).

According to the comment, student G certainly experienced difficulties but recognized the positive change in their language skills. Moreover, student A said, "I made use of my background knowledge and tried hard (to write), so it was good in the end (背景知識をフルに生かして、頑張っていけたから、結果としては良かった)," thus depicting their challenge in a cognitively taxing task. Student L explained that it pushed them to use difficult expressions in writing: "If only language was covered in class, maybe I would not have used difficult expressions and would have ended the class (片方[の内容]しかない場合は、多分あんま難しい表現とか自分使わずにその授業終えるんだろうなって感じがしますね)." Student E also valued the experience of pushed output for memory retention.

It was hard but good. Simply learning about something doesn't make the experience memorable, but in writing, I had to think about examples. Not just the input but the combination of input and output makes our memory strong. Thus, I gained knowledge

175

(大変だったけど、良かったです。学ぶだけだと、あ、そうなんだで終わるけど、書くことで具体例とか考えたりして、インプットしただけじゃなくて、インプットしてアウトプットしたら記憶が伸びるじゃないですか。そのことについての。だから、知識になったなみたいな感じ。).

Therefore, the difficulty of pushed output may have been interpreted as a springboard for gaining knowledge. In summary, although it was not the same for all the students, some of them considered the challenging nature of the course as a source of feeling a sense of achievement.

4.3.3.3 Summary of the Interview Results for the CBI Course Perceptions

The student interview comments suggest that they perceived both difficulties and positive influences by having a dual focus on content and language learning in the current CBI class. More specifically, the challenges were owing to the novelty of the content, language used for content learning, and the synergy of these two aspects. Despite these difficulties, they also perceived positive effects on their language use and development, especially in writing, reading, and vocabulary. Moreover, they perceived value in learning about the content (i.e., the SDGs) itself and that positively affected their motivational base for learning in the CBI class. In addition, some of the students mentioned that they experienced a sense of achievement by completing the class despite the difficulties they faced.

A visual summary of the interview coding for the CBI course perception is shown in Figure 4.2 to illustrate the relationship between each theme and category. The numbers (e.g., 1, 1-1) indicate the themes and categories identified in this analysis. Results highlighted dichotomous feelings, which are difficulties and positive effects, placed in squares on the left

and right sides. These two feelings seem to have arisen from the focus on both content and language learning, which is shown in the colored circles in the squares, and they variously

Figure 4.2

Visual Summary of the Interview Coding for the CBI Course Perceptions
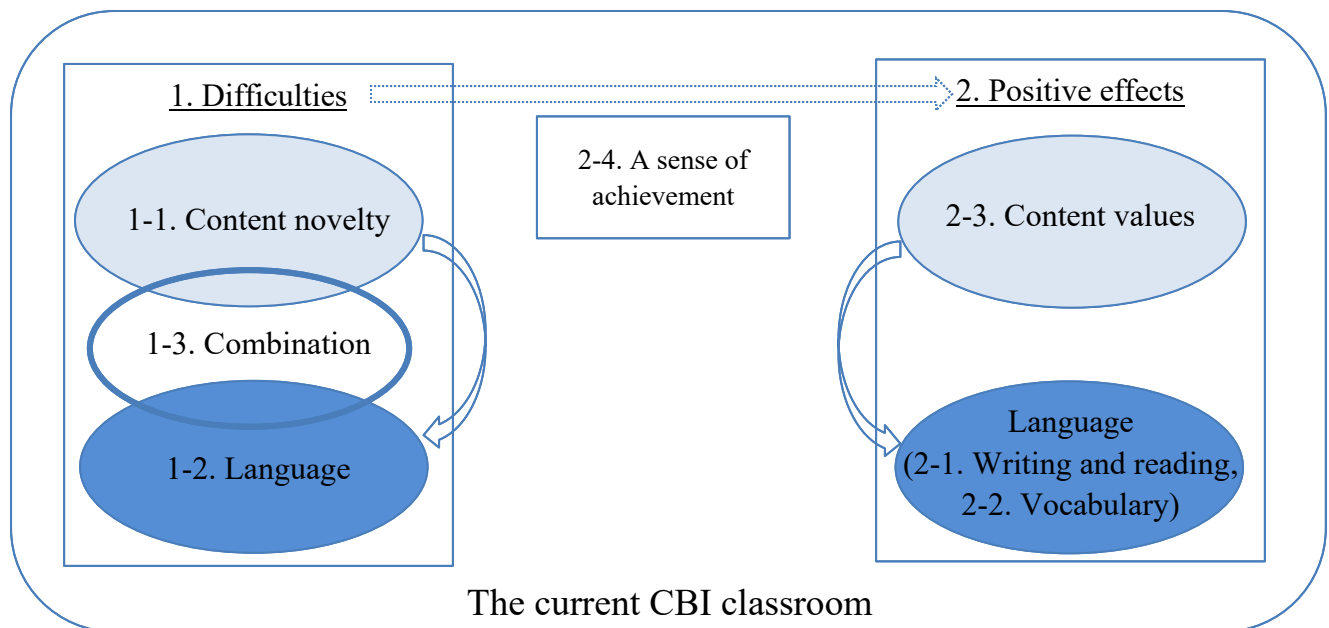
affected the students' perceptions of the difficulties and positive effects. For the difficulties, the novelty of the content led to the feeling of facing a challenge, and learning about the new concepts that they had not learned even in Japanese seems to have been especially difficult for them. Next, the interview results revealed that the language to input new content information and output what they learned and thought caused feelings related to these challenges. This feeling of language difficulty appeared to have arisen from cognitively more complex content to be learned compared to the topics that they had dealt with in English classes. In addition, the higher language level in the input materials for content learning and their lack of experience in producing written output added to their difficulties. As shown in the center circle between content novelty and language, the combination of content and

language learning also greatly influenced the feeling of difficulty. Students' comments asserted that the dual focus on different sources of difficulties seems to have raised their cognitive complexity.

In spite of these challenges, the learners also pointed out the positive aspects of this type of instruction. They appreciated the presence of content learning itself since they recognized the value in learning about the SDGs for both intrinsic (e.g., genuine interest in and importance of the topics) and instrumental (e.g., preparation for college transfer exams and jobs after graduation) reasons. Moreover, the focus on content-supported vocabulary learning through repeated exposure and use of some of the prominent words in content learning provided ideational and linguistic support in writing and facilitated attentive reading. A dotted arrow from difficulties to positive effects shows the possibility that the learners felt a sense of accomplishment in class by facing challenges and trying to deal with them. The two themes and seven categories summarize the students' perceptions that stood out in the current CBI course, indicated by the square covering all these themes and categories in the figure.

As shown in the colored circles in the figure, the focus on content and language learning are present in both squares for the difficulties and positive effects of CBI. As for content, dealing with intellectually denser content in the language class raised difficulty in content learning because of new and concentrated content information; at the same time, the added cognitive complexity in content learning supported the students' perceived value for content learning and increased positive perceptions toward the class. Similarly, both difficulty in English and positive effects for English learning emerged from the focus on content, indicated in the arrows from content to language circles in the figure. Therefore, in planning their classes, teachers should be aware that the inclusion of sustained and thought-provoking content has these influences. Control of the cognitive complexity level in the

classroom needs to be properly adjusted in language-focused CBI classes, especially in EFL classrooms in Japan where greater focus is typically placed on linguistic aspects.

4.4 Discussion

Study 2 explored one writing task difference variable (i.e., with or without the requirement of source integration) in a content-based writing instruction classroom conducted for Japanese college students. More specifically, the effect of the task variables on the students' writing quality was investigated through linguistic analysis and rating, and their perceptions of the task differences were explored through interviews with 13 participants. This section will discuss the task difference variables referring to both quantitative and qualitative results. In addition, learners' perceptions toward the current CBI course, in which they learned about both English writing and the SDGs, will be discussed in relation to the previous studies.

4.4.1 Quantitative and Qualitative Analysis of the Source-integration Task Variables

This section aims to discuss the results of writing performances and student perceptions about the task variables in terms of the numerical evaluations of the students' writing and their task perception interviews. To summarize the first quantitative analysis, it was revealed that the with source integration writing obtained significantly higher fluency and marginally higher lexical sophistication. There was no difference in terms of the syntactic complexity, accuracy, and rating by human raters. These results suggest that the current task manipulations elicited different writing focus although these tasks are sometimes treated as one task type (source-based writing) (Knoch & Sitajalabhorn, 2013).

The current study's finding regarding fluency is in line with Abrams (2019), which also compared writing performances on the with and without source integration writing tasks of

lower-level learners of German. In contrast, the current study did not support Ong and Zhang's (2010) results that found no difference in fluency between one group, which was only given the writing topic, and the other group, which was given the topic, ideas, and macro structure, written by Singaporean university students. Similarly, the finding was contrary to Cumming et al. (2005), which found higher fluency for the without integration task; however, this result may have been influenced by the task instruction that set the higher minimum word limit (e.g., 300 vs. 100–200 words) for the without integration task. Although both tasks for the current study were accompanied by the relevant reading materials, the necessity of integrating information into writing caused students to produce longer output. Integration of the source-text information may be suitable for encouraging low–intermediate writers create more written output. In this respect, integration of the source text information may have lessened the resource-dispersing aspect of cognitive task complexity (creating + prior knowledge condition by the source text's ideas and language inclusion). Student interviews supported this view. Many students confirmed that for the with integration task, the reading provided concrete examples to develop their writing, and thus enabled them to write longer. In addition, student J mentioned that being able to write more in this type of task made them feel positive about the task, saying that it was "easier to write (書きやすい)" and "It looks good if the writing is longer (見た目、いいじゃないですか、長いほうが。)." Therefore, this task manipulation may affect students' minds about the task in a positive way.

Second, with regard to the syntactic complexity measures, the mean length of T-unit and dependent clauses per T-unit did not differentiate between integrated and independent writing. The current result is in line with Cumming et al. (2005) in which the independent and reading-writing tasks showed no difference in the mean length of T-unit. In addition, the finding contradicts with Abrams (2019) in which the mean length of T-unit was marginally longer for the integration task ($p = 0.078$, $d = -0.273$ in Abrams, 2019), but it is in line with

Abrams (2019) in that the clause per T-unit did not differentiate between the integrated and independent writing ($p = 0.241$, $d = −0.35$ in Abrams, 2019). From the results of the above previous and current studies, which are inconsistent and show only marginal results, it is speculated that with or without source integration variables may not affect syntactic complexity in writing. According to Robinson (2005), higher reasoning demand may elicit higher complexity for an explanation, and since both tasks belong to the same genre (opinion writing), the reasoning demand for explaining opinions may not have differed across the tasks. Moreover, students may not have borrowed syntactically complex structures when citing the source text information as doing it may have been demanding for them. Student K mentioned that they could use vocabulary from the reading but were unable to instantly use grammar in the reading.

> I learned a word and immediately used it. The image (of the word) was instantly associated with writing. I was able to write each word. However, there was not much that I do with grammar (単語を知れてすぐに書けるようにみたいなのが、イメージがすぐにそのライティングにつながるような、単語一個一個を書けるようになったとかはありましたけど、文法がどうとかっていうのはあんまりなかったです).

Teachers can be informed that both types of writing may elicit the same level of syntactic complexity from learners at the current English level (e.g., a low- to intermediate-level). To encourage students to complexify their writing, it may be suitable to raise awareness about the possible structures to be used in the instruction and not solely use the current task manipulation. Additionally, this study allowed students to use assisting devices for writing, such as a dictionary and translation tool. This task condition may have aided their syntax skills, and thus the syntactic complexity was not different between the two tasks.

As for lexical complexity, there was no difference in diversity between the task types, which was contrary to Cumming et al. (2005), Frear and Bitchener (2015) and Shin and Kim (2014). For the current study's writers, the source text integration did not necessarily widen the range of words used. Since the student interviews revealed that some of them used words in the source texts for both types of tasks, the lexical diversity did not differ. In contrast, lexical sophistication was marginally higher ($p = 0.08$, $d = 0.45$) for the integrated writing as in Kyle and Crossley (2016) and Shin and Kim (2014). Writers may have been able to include sophisticated words used in the reading material for the with integration task. Integrated writing may offer learners opportunities to explore and use words outside their current lexicon and increase the chance to use sophisticated words. However, the processing of these advanced words may not have been deep enough to be stored in the students' memory as they mentioned their improvement in receptive vocabulary in the interviews. In summary, the source integration increased lexical sophistication presumably due to using high-level words in the source text, but the variety of words was not affected. In other words, the source integration may affect the use of vocabulary but may not lead to the acquisition of vocabulary due to a lack of deep and repeated processing of that vocabulary.

The current study did not find any differences in accuracy and rating scores (comprehensibility, content, task requirements, coherence, and cohesion). The result for accuracy corresponds to Cumming et al. (2005). In fact, there were fairly few interview comments regarding accuracy for both types of tasks, implying that they may not have paid a lot of attention to accuracy but rather may have focused on the content to be written and the organization of the overall writing, which constituted a major part of the interview. Therefore, in addition to task manipulation, explicit feedback on accuracy is necessary. Moreover, as the interview comments suggested, since the writing was conducted in class with a time limitation, they may have focused on finishing the task with all the necessary

parts. In addition, to simulate the actual writing condition for EFL learners, this research did not prevent students from using writing assistant tools, such as dictionaries and smartphones owing to which they may not have spared attention to the differences in the linguistic aspects between the two task types. Hence, the integration of source text may not have influenced the current-level learners' accuracy in writing.

These CAF results of source text integration tasks can be interpreted from TBLT task complexity discussions (Robinson, 2005). The with integration task can be considered as difficult along the resource-directing aspect in terms of the reasoning and number of elements (Abrams, 2019). The hypothesis states that higher complexity along the resource-directing dimensions negatively affects fluency while positively affecting accuracy and complexity. The current study result for the with integration task partly supports this view by showing higher lexical sophistication but no higher syntactic complexity, accuracy, and fluency. Conversely, the hypothesis states that lower complexity along the resource-dispersing dimensions positively affects all CAF measures. The with integration task can have lower complexity along resource-dispersing aspect in terms of prior knowledge since the source text gives writers this knowledge. The current findings partly support the hypothesis in terms of higher complexity for fluency and lexical aspects but not for syntactic complexity and accuracy. Consequently, the source integration task may not yield the performance results assumed in this framework since it includes both resource-directing and dispersing aspects in one task. In other words, the source integration writing task has several different variables in one task (e.g., + reasoning demand, – few elements, + prior knowledge) that may work in combination to affect the task complexity and performance of each learner differently.

As for the rating results, there was no difference in any of the functional adequacy scales. More specifically, there was no difference in content ($p = 0.80, d = 0.05$) and task requirements ($p = 0.63, d = 0.09$), and there was a small effect for comprehensibility ($p =$

0.15, d = 0.29, higher for the with integration task) and coherence and cohesion (p = 0.31, d = 0.20, higher for the without integration task). This study's results did not support Cumming et al. (2005), who found a better argument structure and quality of claims for the without integration task with large effect sizes. This study also did not support Cho (2019), which obtained better rating scores for the with integration task with moderate effect sizes and Shin and Kim (2014), which obtained better coherence for the with integration task. However, it should be reminded that the rating methods are different between the current study and these previous studies, and thus it may not allow simple comparison. In sum, previous studies found that the without integration task elicited a positive performance over the with integration writing, but the current study only weakly and partly supports these findings. The comparisons of the results of the current study and previous studies are shown in Table 4.19.

When considering these results with the interview comments, it was revealed that both types of tasks can elicit low and high perceived task difficulty from students in terms of idea generation and organization. In addition, the number of students who chose either of the tasks as difficult did not differ substantially (22 for with and 16 for without integration task). In other words, it cannot be established decisively that either one was easier or more challenging than the other for the learners to write. Therefore, there was no particular difference in writing quality assessed by the current rating scale for both tasks having different sources of task difficulty, and the with or without source integration conditions did not differentiate the learners' writing performance rated by human raters. Referring to each subscale for the rating, comprehensibility was slightly higher for the with integration task; it can be assumed that this was due to the fact that the students could use the source text as their linguistic and ideational support for writing as mentioned in the interviews. However, since the statistical result was not significant and the effect size was only small, this cannot be said strongly. As for content and task requirements, there was no statistical difference. Content represents the

Table 4.19

Results of CAF and rating of the with integration task in comparison to the without integration task for the current and previous studies

	Syntactic complexity	Lexical complexity	Accuracy	Fluency	Rating
Current study	No difference	No difference (diversity) Higher (sophistication)	No difference	Higher	No difference
Abrams (2019)	Higher	Higher (diversity)	Higher	Higher	–
Ong & Zhang (2010)	–	No difference (diversity)	–	No difference	–
Cumming et al. (2005)	No difference	Higher (diversity)	No difference	Lower	Lower
Kyle & Crossley (2016)	–	Higher (sophistication)	–	–	–
Frear & Bitchener (2015)	Lower	Higher (diversity)	–	–	–
Shin & Kim (2014)	Higher	Higher (diversity)	–	–	Higher (coherence)
Cho (2019)	–	–	–	–	Higher

Note. – represents that the measure was not dealt with in the study.

number of ideas in the writing, and it is speculated that both types of tasks enabled students to include a certain number of ideas. According to the interview results, students could utilize examples in the source text for the with integration task while they could freely use their own ideas for the without integration task without limitation for the content to be included as in the with integration task. Lastly, coherence/cohesion was slightly higher for the without integration task although there was no statistical difference with a small effect size. Since students mentioned in the interview that it was difficult to combine ideas in the source text with their own opinions and organize the writing for the with integration task, they may have been able to deal with coherence/cohesion better in the without integration task.

Besides the numerical results about the task difference discussed above, students' task perceptions were obtained through interviews by asking them the second research question: "*How did the students perceive the difference between the two writing task types?*" According to the thematic analysis, they seem to have had different task focuses and sources of difficulties for the with and without integration tasks. For the with integration task, it was found that understanding a reading material to be included in writing laid the foundation for the task completion. When it was understood, it worked as a catalyst for content development and organizational support for writing. However, some learners experienced difficulty in combining the source text information and their own opinions and organizing the writing. A few students mentioned that this task type had a positive influence on their depth of language processing and possibly on language development since it requires careful reading of a source text, necessitates summarizing and paraphrasing of a source text, works as a vocabulary springboard, and encourages longer output. In fact, 66% of the students responded that this task type contributed to their language learning more. The interview results coincided with the numerical results: this task type yielded higher fluency and lexical sophistication. In contrast, students' major focuses for the without integration task were on the organization and content to be written for the without integration task, both of which have contributing and debilitating aspects for completing writing. Some of them said that the freer setting (no source integration required) made the writing organization easier, but others said not being provided with any guidance, which was somewhat provided in the with integration task by source integration, made the organization difficult. Similarly, there were some who preferred to choose the content to be included freely from their own ideas while others leaned toward having content support from the source text. Contrary to the with integration task, the use of reading varied among students; some said they referred to it even though they did not need to complete the task and others mentioned that they did not refer to it.

When comparing the task perception results with previous studies, the current results further support the view that understanding reading is crucial for the completion of the with integration task (Leki & Carson, 1997; Neumann et al., 2019). In particular, Neumann et al.'s (2019) findings which aimed to unveil ESL students' perceptions of source-based writing pointed out that the usefulness of a source text tended to be dependent on how easily students can understand it. Similar opinions were elicited from the current EFL participants. When using this task type in the CBI/CLIL classroom, teachers are advised to aid and check the students' understanding of the source text by, for example, providing reading questions, asking them to summarize the content in L1, providing visual and aural support, and supporting building form–meaning connection of vocabulary. These may facilitate content understanding and also writing performance. Moreover, the perceived roles of the source text identified in the previous studies, such as serving as a resource for ideational, organizational, and linguistic aspects of writing (Leki & Carson, 1997; Neumann et al., 2019; Plakans & Gebril, 2012), were also mentioned by the current research participants. Although the integrated writing research participants are typically ESL students (as in Leki & Carson, 1997; Neumann et al., 2019), this study added further support that EFL college students with low to intermediate proficiency level also use the source text similarly.

In addition, the current interview highlighted the language learning potential of the with integration task triggered by source integration (e.g., paraphrasing), which has just begun to be investigated in the TBLT field (Abrams, 2019). Although the current study's participants expressed the difficulty in organization for this task type because of the source integration requirement, which was also pointed out by Neumann et al. (2019), this requirement may have increased the learners' cognitive engagement with reading and possibly contributed to the expansion of language repertoire. In addition, positive comments about vocabulary acquisition for the with integration task in the interviews are in line with Jung (2020). With

187

these positive benefits of this task type derived from a deeper engagement with the reading materials, the with integration task can benefit learners who are building on their various language skills (e.g., vocabulary, grammar, reading, and writing). The use of it as a pedagogical task for language learning can be a welcome addition to a language classroom.

Conversely, for the without integration writing, the task condition that does not require a citation of a source text left the judgement of whether or not to use source text information in writing to each writer: some students referred to it as an ideational and linguistic resource, but others did not and depended solely on their own resources. Therefore, although a source text was provided in both types of writing, how the writers dealt with it differed greatly. As Knoch and Sitajalabhorn (2013) pointed out, these two types of integrated writing tasks are sometimes considered as one type of task; however, this research found that students' source text use, task perception, and writing performance differ between these two tasks. With these differences in mind, teachers can customize the task variables according to what they want their students to focus on in the writing (e.g., with integration task for deeper processing of reading materials for source integration and without integration task for utilizing students' own ideas and linguistic skills).

In addition to the comments about the variable use of reading materials, the students' comments were focused on the content to be used and essay organization. Since they could depend less on a source text for these two aspects of writing, they considered these components on their own to complete the writing, and they appeared frequently in their comments. Interestingly, some students perceived freedom in the choice of content and organization as a positive condition for their task completion while others preferred to have some constraints in content and organization imposed by source text integration. Thus, there seem to be individual differences on which aspects to focus on and decide on the difficulty level of the task.

Up to this point, how the degree of source integration (e.g., a source text provided but no need for integration or integration required) influences EFL college students' writing in terms of CAF measures and rating and their task perceptions have been discussed. These tasks were conducted in the CBI/CLIL writing course created for this study since skill integration (reading and writing) naturally occurs and is necessary for academic content learning (Plakans, 2015). Even in the language-driven CBI/CLIL in EFL classrooms, the use of integrated writing tasks adds to the variety of pedagogical tasks. Moreover, it is important for CBI/CLIL teachers to be aware of the affordances that pedagogical tasks have and to be able to enable both content and language learning to happen effectively. If they are aware of the characteristics of the tasks used for content learning and their influence on language development, they can effectively plan the course. For instance, it was mentioned by the students that they read a source text more carefully for the with integration task and used the text as a vocabulary and grammar reference. We can use this task type when we want to give students opportunities to process language input deeply and to use such L2 input for themselves in their written output. This can be done as a review activity not only for content but also for vocabulary and language structure. In contrast, the without integration task can also be conducted in the CBI/CLIL classroom to enable the students to think further about the content learned. In this task, they are expected to use their language and ideational resource more independently. Like these, the characteristics of the tasks and how they affect language performance and task perception are important information for teachers to imagine the outcomes of the learning tasks of their choice. In this sense, what has been found in the TBLT research that investigates different task variables should be referenced in planning a language-driven CBI/CLIL course for EFL learners.

4.4.2 Student Perceptions toward the CBI Class and Pedagogical Implications

This study also investigated students' perceptions of the CBI course to further understand the EFL college students' views on the course who had not received this kind of instruction before. Thematic analysis of the interviews suggests that the difficulty and positive effects of having content focus are the two dominant themes in their comments. More specifically, the perceived difficulty arises from the novel content learning, language used for content learning, and a combination of these two aspects. In contrast, students experienced the positive effects of having content focus for the genuine interest and value toward the content learning, and they also mentioned its effects on their language learning and development. For some students, the challenge caused by the content and language integration gave them a hurdle to overcome, and when they tackled it, they felt a sense of accomplishment in the course.

The current study coincides with past studies on students' CBI/CLIL perceptions in Japan in terms of positive student reflections on their language skills development (Kavanagh, 2019; Ichimura et al. 2021), content knowledge development (Kavanagh, 2019; Underwood, 2019), and affective states, such as feelings of interest and usefulness of the tasks used in class (Yanagawa, 2017). It can be said that students' positive CBI/CLIL class perceptions may be rooted in a feeling of development in language, gaining new knowledge of content, and their positive affect related to the content and language learning objectives in the classroom. Although it was reported less in previous studies, this type of instruction also entails feelings of difficulties: students may be overwhelmed by the challenge of focusing on both content and language learning. As in the current study, Underwood's (2019) participants also mentioned difficulties in understanding the language for content learning and writing about it and the lack of content knowledge. Despite these challenges, his students' writing achievements were high, and their feelings of interest, enjoyability, and usefulness

characterized the course.

From the previous studies and current research, it can be inferred that the inclusion of content learning in a language class can be a source of both positive effects and difficulties, showing the central role of content. For the positive aspects, content learning in language classrooms can expand the scope of learning, thereby offering students opportunities to deal with L2 in a more meaningful and realistic situation. For example, the current study connected the learning of the SDGs and English writing, and the inclusion of content learning created a need for L2 input to learn about the SDGs and for L2 output to demonstrate understanding and express opinions about the content learned. In addition, the positive affect derived from content learning (i.e., novelty, interest, and value) was also found to be a driving force for language learning as in Ichimura et al. (2021), Kavanagh (2019), and Yanagawa (2017). The students in this study appreciated the content learning as it had a relevance to their other classes and daily lives, which also emerged as a key theme in the CBI course conducted in EFL classes by Cumming and Lyster (2016). This meaningful connection seems to have been a motivational base for pursuing their study of English writing as mentioned by some students in the interview. In addition, the inclusion of content learning has a positive influence beyond language skills development: it may contribute to the students' intellectual development as well. Cammarata et al. (2016) strongly called for the need to develop "thinking-rich" FL programs rather than "thinking-light" ones so that FL class can not only work as a site for language skills development but also support learners' overall intellectual development. Students' comments about the appreciation of content learning for facilitating thinking support the presence of this beneficial aspect of CBI/CLIL in the current study. Overall, this study adds further evidence that meaningful content learning in language classrooms plays a crucial role in further enriching students' learning experience in FL classrooms.

In spite of the positive aspects, teachers should be mindful about the difficulties that students may experience in the CBI/CLIL classrooms. For the content difficulty, teachers could firstly choose relevant content for the students in terms of, for example, their major, interest, and age so that they could continue their learning supported by the interest in the content itself. Connecting the content to be learned and their past learning and life experiences also evoked the learners' positive affect in the current study. Moreover, teachers need to have sufficient content knowledge so that they can teach comprehensively even in L2. To gain knowledge, cooperation with subject teachers may benefit CBI/CLIL teachers, and consequently, the students. When dealing with a new concept wherein simple translation in L1 does not necessarily aid the learners' understanding, sufficient amount of instructional time, various teaching tools (e.g., videos, visuals, realia), and instructional techniques (e.g., repetition, paraphrasing, emphasizing) are required.

For language support, vocabulary in the input materials was emphasized as a source of difficulty in the student interviews. Thus, teachers can help students build form–meaning connections by creating a word list, providing visual images of words, using the words repeatedly in the input, and providing opportunities to use them in the students' output. To make the content linguistically comprehensible for students, it is necessary to have the skills to paraphrase the content using easier words and simpler sentence structures. In addition, providing grammar explanations for a particular point that is causing trouble in students' understanding can also be powerful language support. Use of L1 can also support their L2 performance and learning in the context of content learning, which is represented by the term translanguaging (Cumming, 2014; Lo & Lin, 2018). Previous research suggests that L1 is used in CBI/CLIL classrooms for explaining subject-specific terms, clarifying difficult content concepts, task management, rapport-building, and improving the students' metalinguistic awareness (Lo & Lin, 2018). Especially in EFL settings, such use of L1 can

serve as students' cognitive and academic resources and bring depth and learning efficiencies into the classroom. In addition, as Cummins (2014) suggested, the following skills can be transferable from L1 to L2: conceptual elements of content learning (e.g., understanding the concept of solar power generation system), metacognitive and metalinguistic strategies (e.g., strategies for the visual information organization, strategies for vocabulary learning), and pragmatic aspects of language use (e.g., use of gestures). With these in mind, teachers can draw on learners' such skills in L1 and enable them to use the skills to facilitate their content and language learning in L2. In summary, adjusting the difficulty levels for learning content and language is a major challenge for teachers, but there are ways to accomplish it (e.g., L2 input modification, translanguaging). Despite the difficulties raised, some of the students in this study appreciated the challenge and said they felt a sense of achievement, which was also observed in a study by Cumming and Lyster (2016) conducted on an L2 French CBI course. To benefit from the challenge created by the content and language integration, it is of utmost importance for teachers to pay attention to the students' perceived difficulty level and adjust it to affectively support their learning in CBI/CLIL.

4.4.3 Limitations

In summary, this study featured a CBI writing class and compared the performance and perceptions of two types of writing tasks used in the study. With or without integration of the source text differentiated the writing performance in terms of fluency as well as lexical and syntactic complexity, and different focuses of these tasks were revealed through interviews. Students' perceptions of the current course highlighted the feelings of difficulties and positivity raised by the combined learning of content and language. This study added another empirical investigation of the task variable that reflects the characteristics of CBI/CLIL instruction (i.e., source integration) and students' viewpoints of the tasks and course.

Despite these findings, there are certain limitations. First, since this study only showcases one CBI course with a limited number of available data (n = 27), the generalizability of the results cannot be claimed strongly, and the findings should rather be interpreted as one case study. However, the accumulation of empirical research with participants' data gathered in the classroom will collectively inform the future implementation of CBI/CLIL instruction. Second, the students were allowed to use writing aids, such as dictionaries and translation tools, and the use of these devices certainly affected their writing performance. However, when EFL students write, it is natural to use these aids, and this writing condition was thought to reflect the actual situation. Even so, to more precisely understand the effect of the task variable manipulation on learners' written production, future studies should also investigate writing without the use of these tools. Third, the researcher conducted the current course and also the interviews; therefore, the interviewees may have avoided making certain comments for fear of offending the researcher. However, the interviews were conducted after the last day of the course, and the students were informed that their responses would not affect the course grade. Additionally, since the researcher had the best knowledge of how the class was conducted and shared the classroom experience with them, she could understand the interviewees' comments. Thus, this research condition may have been able to shed light on the aspects which would otherwise have not been addressed.

In the next section, the results of studies 1 and 2 will be summarized, and the overall research question of this thesis, that is, "*How does theme-based instruction affect EFL students' writing performance and development and their class perceptions?*" will be discussed.

Chapter 5. General Discussion and Pedagogical Implications

5.1 Introduction

With a view to creating meaningful language use opportunities in foreign language classrooms (Cammarata, 2016), the implementation of CBI/CLIL, especially language-focused theme-based instruction, has been conducted in language classrooms in Japan. Although its application has been increasing, its effect on learners' language performance and development as well as their course perception has been less investigated and thus needs further exploration (Ikeda, 2013). To understand how learners' L2 use, especially writing in this dissertation, is affected by the inclusion of content learning, the findings in related research fields, such as writing task variable research in TBLT and second language writing, were consulted, which may provide empirical support for language performance and development in CBI/CLIL. Based on the above discussion, this dissertation tried to investigate the effect of two writing task variables related to content inclusion in language learning (i.e., topic familiarity and integrated writing) on Japanese high school and college students' writing to consider the possible impact of CBI/CLIL learning on learner language performance and development. The two task variables were dealt with in two theme-based units in study 1 and 2, respectively. Simultaneously, students' perceptions of these classes were examined through the analysis of reflection sheets and interviews. The overall research question for this dissertation was as follows: *How does theme-based instruction affect EFL students' writing performance and development and their class perceptions?* The results of the two studies will be discussed in terms of task variable differences and students' course perceptions.

5.2 Discussion of the Two Writing Task Variables Related to CBI and Pedagogical Implications

The summary of task variable results for studies 1 and 2 are shown in Table 5.1. The $\pm$ topic familiarity variable and two types of integrated tasks both elicited different writing performances. More specifically, the content-specific topic writing's (i.e., lower topic familiarity) syntactic complexity (i.e., MLT), fluency, task requirements, and coherence/cohesion were higher both in the pre- and post-tests. Lexical diversity was higher both in the pre- and post-tests, and accuracy was higher in the pre-test for the general topic writing. As for the with or without source integration writing, fluency was significantly higher and lexical sophistication was marginally higher for the with integration task. Although there were a few performance variances, student interviews revealed that their focuses on the tasks were different. As seen in Table 5.1, the topic familiarity task variable seems to have caused more variation in writing than the two types of integrated writing. As for the development of writing after the theme-based unit, the results in study 1 (see Table 5.2) showed that writing for both topics changed after the instruction, with the content-specific topic's change being accuracy and task requirements and the general topic's being comprehensibility and coherence/cohesion.

Based on the results, the first part of the overall research question (*How does theme-based instruction affect EFL students' writing performance and development?*) will be discussed. This study tried to investigate the possible impact of language-focused CBI by focusing on the task features that reflect the learning that occurs in CBI/CLIL. For example, it is assumed that the writing activities used in CBI/CLIL require content knowledge dealt with in the class (Kong, 2015). In this point, the prompt is different from a prompt that only necessitates the writer's general world knowledge and preferences (e.g., a prompt asking about the use of e-textbooks or paper textbooks was used in the current study) typically used

for EFL textbooks. Thus, the level of topic familiarity is thought to be different between the content-specific topics and the general topics. Similarly, integrated writing also reflects the learning process in CBI/CLIL (e.g., gaining content knowledge through reading and then writing about it; Plakans, 2015). Examining how different manipulations of this task (i.e., with or without integration of a source text) affects students' writing and learning experiences will inform teachers' choice of tasks. Previous CLIL studies in Europe have attempted to determine its effects on learner language by considering instructional hours of CLIL as

Table 5.1

Summary of the Topic Difference Results

Writing performance measures	Topic familiarity		Writing with or without integration of a source text
	Pre-test	Post-test	
MLT (mean length of T-unit)	G<CS	G<CS	-
DC/T (dependent clauses per T-unit)	-	-	-
MTLD (lexical diversity)	G>CS	G>CS	-
LS2 (lexical sophistication)			**without<with**
Accuracy	G>CS	-	-
Fluency	G<CS	G<CS	without<with
Comprehensibility	-	-	-
Content	-	-	-
Task Requirements	**G<CS**	G<CS	-
Coherence/Cohesion	G<CS	G<CS	-

Note. CS indicates the content-specific topic and G indicates the general topic. Bold letters indicate marginal *p* values ($p = 0.05–0.1$).

an independent variable (e.g., Jiménez Catalán & Agustín Llach, 2017; Gené-Gilet et al., 2015a,b); however, learning outcomes can possibly be identified in the task performance used in the CBI/CLIL class, and such focus on the task in CBI/CLIL settings is called for (García Mayo, 2015; Lyster, 2017).

Table 5.2

Measures Developed after the CBI Unit for Each Topic in Study 1

Topic	Measures
CS	**Accuracy**, task requirements
G	**Comprehensibility**, coherence/cohesion

Note. Bold letters indicate marginal *p* values ($p = 0.05–0.1$).

In study 1, it was revealed that the content-specific topic (i.e., discuss the benefits and downsides of power generation methods) gained higher performance in terms of syntactic complexity (i.e., MLT), fluency, task requirements, and coherence/cohesion in both pre- and post-test occasions compared to the general topic. The content-specific topic was rated as more difficult, but that did not necessarily elicit lower performances from the students; they tried to produce more written output (fluency) with longer T-unit (MLT) and meet the task requirements with higher coherence/cohesion. This may have been because they tried to explain more for the content-specific topic [+ reasoning demand in the resource-directing dimension in Robinson's (2005) Cognition Hypothesis], and higher perceived difficulty may not necessarily lower their performance. In other words, a topic perceived to be slightly challenging in terms of the necessary vocabulary, content knowledge, and reasoning has the possibility to challenge students to perform at their highest level. This finding also suggests that giving opportunities to students to produce pushed output in the classroom is crucial.

Writing activities in Japanese high school textbooks are reported to be limited to controlled activities (e.g., fill-in-the-blanks, sentence combining; Kobayakawa, 2011) rather than free composition, and the topics tend to be everyday life topics (Nakagawa, 2017) that deal with the writers' everyday world knowledge. Thus, creating L2 output opportunities in CBI/CLIL where students learn new knowledge (e.g., global issues) and engage in cognitive processes (e.g., problem-solving) in writing may enrich their language use experience and possibly elicit higher performances for some of the measures mentioned above. As can be seen in the textbook research (Kobayakawa, 2011; Nakagawa, 2017), the cognitive complexity level of the writing tasks in the textbooks are controlled in terms of the language (e.g., controlled practice, such as fill-in-the-blanks) and content (e.g., everyday topics) so that students can handle them with their limited English; however, with the new Courses of Study, to be able to think, judge, and express in English is emphasized, and writing tasks that enable such English usage is necessary. This study contributed to showing one example of such a writing task and the analysis of students' performance.

General topic writing also elicited higher accuracy in the pre-test and lexical diversity in the pre- and post-tests. This task can be written with the learners' own preference rather than having specific knowledge about the topic and considered as + prior knowledge in the resource-dispersing aspects (Robinson, 2005). The results support the hypothesis in that both complexity and accuracy were higher with the + prior knowledge task, suggesting that learners may have been able to utilize their existing vocabulary and pay attention to accuracy owing to lower cognitive task complexity. This task type is also essential for practicing the usage of vocabulary and linguistic knowledge that they already possess. Thus, this type of writing can be done at the beginning of the learning of a new topic by asking students less cognitively complex warm-up questions related to the topic.

As for the effect of theme-based instruction, some changes were observed for both

content-specific (accuracy and task requirements) and general topics (comprehensibility, coherence/cohesion). The 11 theme-based lessons on the power generation topic exerted differential effects on the performances of the two writing topics. It can be suggested that if the writing topic is aligned with the content focused on in CBI/CLIL, students may pay more attention to accuracy (although this was marginally statistical) and attempt to satisfy the task requirements (i.e., compare and contrast the benefits and downsides of power generation methods) after the instruction. This may be because they had learned the language items and ideas for writing during the CBI/CLIL instruction and conducted similar activities that have the same task requirements. However, it should be noted that other measures, namely syntactic and lexical complexity, fluency, and the other three functional adequacy scales were unaffected by the instruction. Therefore, even if students write about the topic they learn in CBI/CLIL, it does not necessarily mean that all aspects of their writing performances are positively influenced. To improve certain aspects of writing, it may be essential to draw their attention to those aspects and practice them during the instruction. Especially for the linguistic aspects (syntactic complexity and accuracy), it has been shown that even in the meaning-focused language learning contexts, such as immersion with ample L2 input, these facets of language skills tend to not reach the native speaker level (Cummins, 2014). In this respect, as Lyster (2017) suggested, proactive focus-on-form in the lesson planning stage will benefit the students' L2 learning. Thus, conducting grammar explanation and rule application tasks within the theme-based instruction, which is represented as awareness activity and guided practice in Lyster's (2017) instructional model, plays a crucial role in developing students' linguistic skills. In study 1, the expressions for compare/contrast were highlighted and practiced using example sentences related to the power generation topic so that students could use them in their writing. The importance of this kind of explicit language instruction should not be dismissed in the CBI/CLIL classroom.

Another point that needs attention is that the general topic writing performance also improved on comprehensibility (marginal) and coherence/cohesion but not on any linguistic measures (i.e., CAF measures). This means that content-based instruction might also have a positive effect on the general topic writing in terms of rating even if it is unrelated to the content in CBI/CLIL. This may be because students practiced compare/contrast functions in the current CBI course and possibly transferred what they had practiced in the general topic writing, which also required them to compare/contrast. In contrast, there was no influence on the linguistic aspects (CAF), which may necessitate more focused practice for the development. This may further indicate that the learning of linguistic aspects may be contextualized in the topics and necessitates extensive attention and practice for learners to utilize them in their L2 production. However, it should be noted that there could be a potential effect from other English classes that students were taking at the time. In summary, it can be said for the teachers that content-based language instruction can positively influence the writing that is on the same topic learned in the instruction and also the general topic writing, but the aspects influenced are different.

The second study investigated writing performances of the two kinds of integrated writing, which is a versatile task type in the CBI/CLIL (Plakans, 2015) and needs implementation in Japanese language classrooms. The Courses of Study for high school English clearly states the following: "Language activities that link multiple domains, such as expressing opinions about what they have read, are not being conducted appropriately (p. 6, translated by the researcher)." The writing task with source text information required was longer (higher fluency, in line with Abrams, 2019) and marginally lexically sophisticated (Kyle & Crossley, 2016; Shin & Kim, 2014). Students' task perception highlighted the crucial role of understanding the reading material, and the support that reading materials can offer for writing and language learning was also mentioned. Learners said that they perceived

the supportive and difficult sides of this task type in terms of organization and content generation. From this study, teachers can predict which writing aspects may be affected by the source text integration task manipulations and how they are perceived by students. When teachers want to ensure that students process the reading material deeply but wish to impose less cognitive burden for idea generation, the with integration writing can be an ideal choice, and when they want their students to write more independently about the topic possibly related to the reading material read in CBI/CLIL, the without source integration task can be selected. As shown in this example, for teachers to know what kind of performance and student perception a task can elicit helps their lesson planning, and ultimately supports the students' language learning.

The two studies discussed in this dissertation tried to explore the learning outcomes of theme-based instruction in terms of writing and student task perceptions by focusing on task as a unit of analysis. Results were interpreted by referring to previous research in TBLT and second language writing to further understand student learning outcomes in the CBI/CLIL. The investigation of tasks is also indispensable in CBI/CLIL research in addition to the comparison research between CLIL and non-CLIL classes. Likewise, the studies investigating the interface between CBI/CLIL and TBLT (García Mayo, 2015: Lyster, 2017) and CBI/CLIL and second language writing (Manchón & Williams, 2016) are called for to further share the research insights. Current studies in this dissertation revealed that the focus on content learning in addition to language learning impacted learner writing performances differently for the content-specific and general topics, and varied levels of source integration writing elicited different writing performances and task perceptions. With such information related to tasks used in CBI/CLIL, teachers may be able to presume students' learning experiences in the classroom and effectively organize language learning in relation to content learning.

5.3 Discussion of the Course Perceptions and Pedagogical Implications

To answer the latter part of the overall research question of this dissertation, *how does theme-based instruction affect EFL students' class perceptions?*, it was explored through reflection sheets in study 1 and interviews in study 2. The analysis of the reflection sheets revealed that students felt both positive and negative feelings toward the classes. The positive comments were mainly about the target content and language items described with words, such as "understood, was able to ~, want to review ~," and the negative comments were related to their understanding level and language skills (especially vocabulary and listening) described with words, such as "difficult, was not able to ~, did not understand." It can be said that the students perceived the learning of content and target language forms (e.g., compare/contrast language) favorably, but at the same time, they negatively reflected on the higher input language levels for content learning and their lack of L2 skills. Commonalities can be observed in the results of study 2 in which the students' course perceptions were more deeply explored in the interviews. There were two themes showing the difficulty and positive sides of the current course. Difficulty was due to the learning of new content, L2 skills necessary for learning, and the combination of content and language learning. In contrast, positive effects were also specified, such as the supportive role that content can play for L2 use, value of learning content, and inducing a sense of achievement. Both studies point out the difficulty of L2 input and output caused by content learning and positive feelings toward learning about the content itself. From the results, it can be summarized that the focus on content learning in Japanese language classrooms can bring about students' positive affect (e.g., content value, language development) and negative affect (e.g., difficulty, inability).

Based on these findings, implications for introducing language-focused CBI/CLIL in Japanese high school and college classrooms can be suggested. As for difficulty, teachers should be reminded that the inclusion of content learning (e.g., environmental problems,

resource and energy problems, population and food shortage problems are the content suggested in the Courses of Study for high school English; MEXT, 2018, p. 129) will heighten the difficulty level of both L2 used for learning about the content and cognitive skills for content learning. Students in the two studies in this dissertation mentioned that L2 required for content learning is more demanding than the L2 they encounter in other English classrooms. This may be because content materials tend to include wider vocabulary and academic language that indicates various discourse functions (e.g., cause and effect, compare and contrast). Therefore, lesson planning is a crucial key for successful learning in CBI/CLIL, and providing scaffolding for both language and content learning is necessary. The SIOP model (Echevarría et al., 2017) indicates the key milestones of lesson planning (component, lesson preparation, building background, and comprehensible input), which provides teachers concrete ideas for content and language scaffolding.

Although the SIOP model and other course planning models had been referred to in the creation process of the two studies in this dissertation, it was still challenging to adjust the language level in the input materials to match the student levels and keep the content intellectually interesting for the current participants. Therefore, extreme care needs to be taken for multiple aspects of CBI/CLIL lesson planning (e.g., content and language teaching objectives, materials development, comprehensible teacher talk) and implementation (e.g., in-class scaffolding, teacher talk modification, pacing). In other words, CBI/CLIL is cognitively challenging for teachers as well. Thus, continuous self-development in content and language knowledge, command of L2, and lesson delivery skills are indispensable. Importantly, paying attention to learners' affect is also a key factor for successful course implementation since the feeling of difficulty may lower the motivation for learning and strengthen the feeling of inability in English as shown in the reflection sheets. To avoid such pitfalls, deliberate adjustment of language and content in the preparation and in-class delivery phase is essential.

The results also underlined the facilitative role of content learning for language learning and learner affect, which is in line with Kavanagh (2019), Ichimura et al. (2021), Underwood (2019), and Yanagawa (2017). The results of study 1 showed that students were able to reflect on the target language items (e.g., the language for compare and contrast) positively while the results of study 2 demonstrated the facilitative role of skill integration (e.g., reading and writing) for deeper processing of reading and learning of language items in the reading material. In other words, if planned proactively, EFL students with low to intermediate proficiency levels may be able to handle their focus not only on the content but also on language, which may support CBI/CLIL introduction in Japanese classrooms. Content can be dealt with along with language learning. Nevertheless, it is true that explicit grammar instruction and repeated practice (DeKeyser, 2007) is crucial, especially for EFL learners at the proficiency level of this study's participants to proceduralize their newly learned language knowledge. Even so, shifting the focus to meaning (e.g., learning content) is also vital to be a language user. Therefore, it is essential for teachers to consider the balance between language learning and content learning (meaningful use of language) within each CBI/CLIL course and also at the curriculum level. The students also mentioned that the integration of different language skills contributed to language learning. CBI/CLIL can provide a base (e.g., content) to use different language skills in combination, which was positively perceived by the students. In Japanese classrooms, the lack of such activities that combine different language skills is pointed out in the new Courses of Study, and these activities are encouraged to be integrated into the class (MEXT, 2018). Introducing CBI/CLIL can be one trigger to seamlessly incorporate such skill integration tasks in the classroom.

Interview results demonstrated that not only language aspects but also student affect was positively influenced by content integration. The results of reflection sheets and interviews showed that the content raised the students' interest level, and they perceived

value in learning about the content for their pure interest and usefulness for college transfer examinations and job hunting. For teachers, it is emphasized that the choice of content in language classrooms is as important as planning for the activities for language skills development. To enrich the content-learning side of instruction, the range of things that teachers should pay attention to may certainly increase. Among many such things, it is necessary for them to be able to teach the content in L2, to encourage learners' deeper processing of the content and language, and to devise tasks that facilitate freer use of learners' language and content knowledge. Thus, CBI/CLIL requires a teacher to not only be a language teacher, but also to be a content teacher, a skillful task planner, and most importantly, a language user. Even so, these challenges have a possibility to result in learners' positive perceptions about the content and language learning, as mentioned in the current interviews and previous studies.

In this section, students' perceptions about the two CBI courses were discussed, which highlighted the positive and difficult sides of this type of instruction. Some believe that CBI/CLIL is extremely difficult, and the learning of basic language items should precede this kind of instruction. However, the author believes that it is worth conducting CBI/CLIL instruction for low- to intermediate-level Japanese learners for the positive comments mentioned by the students in the current dissertation and the base that content can lay for language learning and language use. Underwood (2019) summarizes a similar view:

> In some conversations, it is not uncommon to hear the opinion that a CLIL approach for entry-level classes will be ineffective; either because students lack the language proficiency required to understand the subject content, they should be focused more on mastering basic grammatical structures, or because they lack the motivation for study. The high levels of achievement and the students' voices reported in this study

206

challenge all of these assumptions (pp. 48–49).

It is also stated in the Courses of Study for high school English classes (MEXT, 2018) that

> The subjects of language activities should be appropriate to the students' interests and relate to the content studied in other subjects such as Japanese, geography, history, science, in order to develop the ability to solve problems using English (言語活動で扱う題材は，生徒の興味・関心に合ったものとし，国語科や地理歴史科，理科など，他の教科等で学習した内容と関連付けるなどして，英語を用いて課題解決を図る力を育成する工夫をすること。p. 129).

This means that content-based instruction is aimed at in high school English classes, and the problem-solving skill developed through the integrated learning of content and language is also a crucial skill to be continually developed in college English classes. Therefore, further empirical research about the implementation of CBI/CLIL for Japanese high school and college learners is called for.

5.4 Conclusion

This dissertation investigated the efficacy of theme-based instruction for EFL high school and college learners' writing and provided empirical data on writing performance and development. Furthermore, it provided two examples of theme-based instruction units and explored students' course perceptions through the analysis of reflection sheets and interviews. By examining the two writing task variables (i.e., topic familiarity and integrated writing) related to the characteristic of CBI/CLIL, different writing performances were obtained in terms of linguistic aspects and rating. The analysis of student perceptions indicated both positive and difficult aspects of the current CBI/CLIL courses.

The two studies can be informative for teachers who implement theme-based instruction in high schools and colleges in Japan, especially under the current situation where the new Courses of Study has been implemented in high schools in 2022. Findings indicated that a content-specific topic can elicit longer and syntactically more complex performances with a higher achievement level of task requirements and coherence/cohesion compared to general topic writing, showing that the writing topic which requires specific content knowledge may elicit higher performances for some aspects. Additionally, it was shown that writing prompts with integration of source text requirement elicited longer and lexically more sophisticated writing, and students focused more on the processing of the reading materials. By knowing these task variable characteristics, teachers can effectively choose writing tasks that are compatible with the aims of their instruction. Moreover, qualitative analysis of the student perceptions of courses informs teachers of the benefits of this kind of instruction and students' difficulties that need their deliberate attention in planning and delivering CBI/CLIL. As for the research contribution, this study investigated task variables related to CBI/CLIL by referring to TBLT and second language writing research findings. The focus on task

characteristics added different ways of understanding CBI/CLIL learning, and insights from diverse research fields are necessary to understand classroom instruction.

Further empirical research on how to effectively incorporate content learning in language classrooms through tasks is needed. This study examined writing, but other language skills should also be analyzed. In addition, students' affect needs to be carefully attended to for successful implementation, and a qualitative analysis of their lesson-to-lesson change of affective states will provide detailed information. Lastly, teachers who implement CBI/CLIL need additional research attention; studies on their pedagogical focus in the course planning, materials development, and delivering stages may generate instructive information for other instructors. Furthermore, the development of a materials bank for teachers may support their lesson planning (Cumming & Lyster, 2016). Overall, CBI/CLIL in Japan has the potential to enrich learning in language classrooms and needs further empirical investigation in classroom settings.

References

Abdi Tabari, M., Bui, G., & Wang, Y. (2021). The effects of topic familiarity on emotionality and linguistic complexity in EAP writing. *Language Teaching Research*. *0*(0). https://doi.org/10.1177/13621688211033565

Abrams, Z. I. (2019). The effects of integrated writing on linguistic complexity in L2 writing and task-complexity. *System*, *81*, 110–121. https://doi.org/10.1016/j.system.2019.01.009

Abrams, Z. I., & Byrd, D. R. (2017). The effects of meaning-focused pre-tasks on beginning-level L2 writing in German: An exploratory study. *Language Teaching Research*, *21*(4), 434–453. https://doi.org/10.1177/1362168815627383

Anderson, L. W., David, R. K., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Aoyagi, N., Tanaka, M., & Ikeda, M. (2016). Teaching English to engineering students: A CLIL approach. *J. of JSEE, 64*(6), 56–62.

Artieda, G., Roquet, H., & Nicolás-Conesa, F. (2020). The impact of age and exposure on EFL achievement in two learning contexts: formal instruction and formal instruction + content and language integrated learning (CLIL). *International Journal of Bilingual Education and Bilingualism*, *23*(4), 449–472. https://doi.org/10.1080/13670050.2017.1373059

Baba, K., & Nitta, R. (2014). Phase transitions in development of writing fluency from a complex dynamic systems perspective. *Language Learning*, *64*(1), 1–35. https://doi.org/10.1111/lang.12033

Banerjee, H. L. (2019). Investigating the construct of topical knowledge in second language assessment: A scenario-based assessment approach. *Language Assessment Quarterly*, *16*(2), 133–160. https://doi.org/10.1080/15434303.2019.1628237

Bax, S. (2012). *Text Inspector*. [Online text analysis tool]. Text Inspector. https://textinspector.com/.

Bentley, K. (2010). *The TKT course: CLIL module*. Cambridge University Press.

Bloom, B. (1956). *Taxonomy of Educational Objectives. Book I: Cognitive Domain*. David Mckay.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brinton, D. M., & Snow, M. A. (2017). The evolving architecture of content-based instruction. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom: New perspectives on integrating language and content* (2nd ed., pp. 2–20). The University of Michigan Press.

Brinton, D. M., Snow, M. A., & Wesche, M. (1989). *Content-based second language instruction*. The University of Michigan Press.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42–65. https://doi.org/10.1016/j.jslw.2014.09.005

Bulté, B., & Housen, A. (2019). Beginning L2 complexity development in CLIL and non-CLIL secondary education. *Instructed Second Language Acquisition*, *3*(2), 153–180. https://doi.org/10.1558/isla.38247

Butler, Y. G. (2005). Content-based instruction in EFL contexts: Considerations for effective implementation. *JALT Journal*, *27*(2), 227–242.

Butler, Y. G. (2011). The implementation of communicative and task-based language teaching in the Asia-Pacific region. *Annual Review of Applied Linguistics*, *31*, 36–57. https://doi.org/10.1017/S0267190511000122

Cammarata, L. (2016). *Content-based foreign language teaching*. Routledge.

Cammarata, L., Tedick, D. J., & Osborn, T. A. (2016). Content-based instruction and curricular reforms: Issues and goals. In L. Cammarata (Ed.), *Content-based foreign language teaching* (pp. 1–21). Routledge.

Cenoz, J. (2015). Content-based instruction and content and language integrated learning: the same or different? *Language, Culture and Curriculum*, *28*(1), 8–24. https://doi.org/10.1080/07908318.2014.1000922

Chan, S., Inoue, C., & Taylor, L. (2015). Developing rubrics to assess the reading-into-writing skills: A case study. *Assessing Writing*, *26*, 20–37. https://doi.org/10.1016/j.asw.2015.07.004

Chapple, L., & Curtis, A. (2000). Content-based instruction in Hong Kong: student responses to film. *System*, *28*, 419–433.

Cho, M. (2018). Task complexity and modality: Exploring learners' experience from the perspective of flow. *Modern Language Journal*, *102*(1), 162–180. https://doi.org/10.1111/modl.12460

Cho, Y. A. (2019). The effects of writing prompt types on L2 learners' writing strategy use and performance. *Studies in English Language & Literature*, *45*(3), 295–314. https://doi.org/10.21559/aellk.2019.45.3.013

Cho, H., & Kim, Y. J. (2021). Comparing the characteristics of EFL students' multimodal composing and traditional monomodal writing: The case of a reading-to-write task. *Language Teaching Research*. *0*(0). https://doi.org/10.1177/13621688211046740

Clark, G. (2013). Snapshot of a lower secondary CLIL program in Japan. *The Asian EFL Journal*, *15*(4), 383–394.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications.

Corder, S. P. (1967).The significance of learners' errors. *International Review of Applied Linguistics, 5*, 161–170.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. Sage Publications.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, *10*(1), 5–43. https://doi.org/10.1016/j.asw.2005.02.001

Cumming, J., & Lyster, R. (2016). Integrating CBI into high school foreign language classrooms. In L. Cammarata (Ed.), *Content-based foreign language teaching* (pp. 77–97). Routledge.

Cummins, J. (1980). The entry and exit fallacy in bilingual education. *NABE Journal*, *4*(3), 25–59.

Cummins, J. (2014). Rethinking pedagogical assumptions in Canadian French immersion programs. *Journal of Immersion and Content-Based Language Education*, *2*(1), 3–22. https://doi.org/10.1075/jicb.2.1.01cum

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, *31*, 182–204. https://doi.org/10.1017/S0267190511000092

Dalton-Puffer, C. (2013). A construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education. *European Journal of*

Applied Linguistics, *1*(2), 216–253.

Dalton-Puffer, C., & Smit, U. (2013). Content and language integrated learning: A research agenda. *Language Teaching*, *46*(4), 545–559. https://doi.org/10.1017/S0261444813000256

Dalton-puffer, C., Bauer-marschallinger, S., Brückl-mackey, K., Hofmann, V., & Hopf, J. (2018). Cognitive discourse functions in Austrian CLIL lessons : towards an empirical validation of the CDF Construct. *European Journal of Applied Linguistics, 6*(1), 1–25.

Day, E. M., & Shapson, S. M. (1991). Integrating formal and functional approaches to language teaching in French immersion: An experimental study. *Language Learning*, *41*(1), 25–58. https://doi.org/10.1111/j.1467-1770.1991.tb00675.x

DeKeyser, R. M. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.

Derewianka, B. (1996). *Exploring the writing of genres.* United Kingdom Reading Association.

Doiz, A., Lasagabaster, D., & Sierra, J. M. (2014). CLIL and motivation: The effect of individual and contextual variables. *Language Learning Journal*, *42*(2), 209–224. https://doi.org/10.1080/09571736.2014.889508

Duff, P. A. (2014). Communicative language teaching. In M. Celce-Murcia, D. M. Brinton & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (4th ed., pp. 15–30). National Geographic Learning.

Early, M. (1990). Enabling first and second language learners in the classroom. *National Council of Teachers of English*, *67*(6), 567–575.

Echevarría, J., Vogt, M., & Short, D. J. (2017). *Making content comprehensible for English learners: The SIOP model*. Pearson.

Echevarría, J., Richards-Tutor, C., Canges, R., & Francis, D. (2011). Using the SIOP model to promote the acquisition of language and science concepts with English learners. *Bilingual Research Journal*, *34*(3), 334–351. https://doi.org/10.1080/15235882.2011.623600

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Theory and practice of task-based language teaching*. Cambridge University Press.

Evans, N. W., Hartshorn, K. J., Cox, T. L., & Martin de Jel, T. (2014). Measuring written linguistic accuracy with weighted clause ratios: A question of validity. *Journal of Second Language Writing*, *24*(1), 33–50. https://doi.org/10.1016/j.jslw.2014.02.005

Fitzsimmons-Doolan, S., Grabe, W., & Stoller, F. L. (2017). Research support for content-based instruction. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom: New perspectives on integrating language and content* (2nd ed., pp. 21–35). The University of Michigan Press.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*(3), 299–323. doi:10.1017/S0272263100015047

Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*, *3*(3), 215–247. https://doi.org/10.1177/136216889900300303

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics*, *36*, 98–116. https://doi.org/10.1017/S0267190515000082

Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing*, *30*, 45–57. https://doi.org/10.1016/j.jslw.2015.08.009

García Mayo, M. P. (2015). The interface between task-based language teaching and content-based instruction. *System*, *54*, 1–3. https://doi.org/10.1016/j.system.2015.09.003

Garner, M., & Borg, E. (2005). An ecological perspective on content-based instruction. *Journal of English for Academic Purposes*, *4*(2), 119–134. https://doi.org/10.1016/j.jeap.2004.08.002

Gass, S. M., & Mackey, A. (2015). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 180–206). Routledge.

Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2015a). Development of EFL writing over three years in secondary education: CLIL and non-CLIL settings. *Language Learning Journal*, *43*(3), 286–303. https://doi.org/10.1080/09571736.2015.1053278

Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2015b). Writing development under CLIL provision. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 139–161). Springer. https://doi.org/10.1007/978-3-319-11496-5

Gilabert, R., Manchón, R., & Vasylets, O. (2016). Mode in theoretical and empirical TBLT research: Advancing research agendas. *Annual Review of Applied Linguistics*, *36*, 117–135. https://doi.org/10.1017/s0267190515000112

Golparvar, S. E., & Rashidi, F. (2021). The effect of task complexity on integrated writing performance: The case of multiple- text source-based writing. *System, 99*, 102524. https://doi.org/10.1016/j.system.2021.102524

Goya, H. (2018). Content-based instruction in an EFL writing class : Is English medium instruction (EMI) effective for productive development of L2 lexes among Japanese EFL learners ? *Scripsimus*, *27*, 27–47.

Grabe, W., & Stoller, F. L. (1997). Content-based instruction: Research foundations. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom: perspectives on integrating language and content* (pp. 5–21). Longman.

Grabe, W., & Zhang, C. (2013). Reading and writing together: A critical component of English for academic purposes teaching and learning. *TESOL Journal*, *4*(1), 9–24. https://doi.org/10.1002/tesj.65

Grabe, W., & Zhang, C. (2016). Reading-writing relationships in first and second language academic literacy development. *Language Teaching*, *49*(3), 339–355. https://doi.org/10.1017/S0261444816000082

Graham, K. M., Choi, Y., Davoodi, A., Razmeh, S., & Dixon, L. Q. (2018). Language and content outcomes of CLIL and EMI: A systematic review. *Latin American Journal of Content & Language Integrated Learning*, *11*(1), 19–38. https://doi.org/10.5294/laclil.2018.11.1.2

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts.* Academic Press.

Grim, F. (2008). Integrating focus on form in L2 content-enriched instruction lessons. *Foreign Language Annals*, *41*(2), 321–346. https://doi.org/10.1111/j.1944-9720.2008.tb03295.x

Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. Routledge.

Hanzawa, K. (2019). The development of voice onset time (VOT) in a content-based instruction university program by Japanese learners of English: A longitudinal study. *Canadian Modern Language Review, 77*(4), 501–522.

https://doi.org/10.1017/S0007114518003884

Hauschild, S., Poltavtchenko, E., & Stoller, F. L. (2012). Going green: Merging environmental education and language instruction. *English Teaching Forum*, *50*(2), 2–13.

He, L., & Shi, L. (2012). Topical knowledge and ESL writing. *Language Testing*, *29*(3), 443–464. https://doi.org/10.1177/0265532212436659

Heyden, T. (2001). Using sustained content-based learning to promote advanced ESL writing. *TESOL Journal*, *10*(4), 16–20.

Hirose, K., & Sasaki, M. (1994). Explanatory variables for Japanese students' expository writing in English: An exploratory study. *Journal of Second Language Writing*, *3*(3), 203–229. https://doi.org/10.1016/1060-3743(94)90017-5

Hosogoshi, K., Kanamaru, T., & Takahashi, S. (2016). Scaffolding skill-Integrated tasks for academic English: With special reference to students' proficiency. *CPEHE Annual Report 2016*, *22*, 21–29.

Hunt, K. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the Sociely for Rearch in Child Development, 35*(1), 1–67.

Hymes, D. H. (1971). *On communicative competence*. University of Pennsylvania Press.

Ichimura, K., Uemura, T., Aoyagi, N., Tanaka, M., & Ikeda, M. (2021). Development of CLIL-based English teaching materials for engineering students on the basis of research data obtained from engineers. *J. of JSEE, 69*(4), 26–30.

Ikeda, M. (2013). Does CLIL work for Japanese secondary school students? *International CLIL Research Journal*, *2*(1), 31–43.

Ikeda, M. (2022). *FLEX: English communication 1*. Zoshindo.

Izumi, S. (2019). Teaching English in English in the new era: A CLIL approach. *LET Kyushu-Okinawa BULLETIN*, *19*, 1–17. https://doi.org/10.24716/letko.19.0_1

Izumi, S. (2012). Eigo de saguru eigogaku no sekai: hieigosennkou no gakusei ni CLIL de genngogaku wo oshieru torikumi [The world of English studies explored in English: An attempt to teach linguistics by CLIL to non-English majors] In S. Izumi, M. Ikeda & Y. Watanabe. (Eds.), *CLIL (Content and language integrated learning) New challenges in foreign language education at Sophia University Volume 2: Practices and applications.* (pp. 59–90). Sophia University Press.

Izumi, S., Ikeda, M., & Watanabe, Y. (2012). *CLIL (naiyou gengo tougougata gakushuu) Jouchu daigaku gaikokugo kyouikuno aratanaru chousen dai nikan jissen to ouyou* [CLIL (Content and language integrated learning) New challenges in foreign language education at Sophia University Volume 2: Practices and applications]. Sophia University Press.

Jexenflicker, S., & Dalton-Puffer, C. (2010). The CLIL differential: Comparing the writing of CLIL and non-CLIL students in higher colleges of technology. In C. Dalton-Puffer, T. Nikula & U. Smit (Eds.). *Language use and language learning in CLIL classrooms.* (pp. 169–189). ProQuest.

Jiménez Catalán, R. M., & Agustín Llach, M. P. (2017). CLIL or time? Lexical profiles of CLIL and non-CLIL EFL learners. *System*, *66*, 87–99. https://doi.org/10.1016/j.system.2017.03.016

Johnson, M. D. (2017). Cognitive task complexity and L2 written syntactic complexity, accuracy, lexical complexity, and fluency: A research synthesis and meta-analysis. *Journal of Second Language Writing*, *37*, 13–38. https://doi.org/10.1016/j.jslw.2017.06.001

Johnson, M. D., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing*, *21*(3), 264–282. https://doi.org/10.1016/j.jslw.2012.05.011

Jung, J. (2020). Effects of content support on integrated reading-writing task performance and incidental vocabulary learning. *System*, *93*, 1–13. https://doi.org/10.1016/j.system.2020.102288

Kashiwagi, K., & Kobayashi, Y. (2019). Science in CLIL in a Japanese upper secondary school: Focusing on increasing procedural knowledge with a usage-based model perspective. *The Journal of the Japan CLIL Pedagogy Association (J-CLIL)*, *1*, 1–26.

Kavanagh, B. (2019). The teaching of intercultural communication with academic writing through a CLIL based approach—A case study of a Tohoku University course. *The Journal of the Japan CLIL Pedagogy Association (J-CLIL)*, *1*, 100–119.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–71). Routledge.

Kessler, M., Ma, W., & Solheim, I. (2022). The effects of topic familiarity on text quality, complexity, accuracy, and fluency: A conceptual replication. *TESOL Quarterly*, *56*(4), 1163–1190. https://doi.org/10.1002/tesq.3096

Kim, M. (2021). Exploring longitudinal changes in lexical and syntactic features in beginning-level EFL learner writing. *System*, *103*, 102680. https://doi.org/10.1016/j.system.2021.102680

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. Institute for Simulation and Training.

Kiyota, A. (2022). Problematizing fluent speakers' unintentional exclusion of emergent bilinguals. *International Journal of Literacy, Culture, and Language Education*, *2*(May), 6–19. https://doi.org/10.14434/ijlcle.v2imay.34385

Knoch, U., & Sitajalabhorn, W. (2013). A closer look at integrated writing tasks: Towards a more focussed definition for assessment purposes. *Assessing Writing*, *18*(4), 300–308. https://doi.org/10.1016/j.asw.2013.09.003

Kobayakawa, M. (2011). Analyzing writing tasks in Japanese high school English textbooks: English I, II, and writing. *JALT Journal*, *33*(1), 27–48. https://doi.org/10.37546/jaltjj33.1-2

Kong, S. (2014). Collaboration between content and language specialists in late immersion. *Canadian Modern Language Review*, *70*(1), 103–122. https://doi.org/10.3138/cmlr.1607

Kong, S. (2015). Designing content-language integrated learning materials for late immersion students. *TESOL Journal*, *6*(2), 302–331. https://doi.org/10.1002/tesj.151

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kowata, T. (2018). Differences between Japanese university students' self-evaluation and teacher evaluation on a reading into writing task. *Annual Research Report on General Education, Akita University*, 7–11.

Kowata, T. (2019). Studies on Japanese university students' self-evaluating description about their reading into writing product, *Annual Research Report on General Education, Akita University*, 7–12.

Krashen, S. D. (1981). *Second language acquisition and second language learning.* Pergamon Press.

Krashen, S. D. (1982). *Principles and practice in second language acquisition.* Pergamon Press.

Krashen, S. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45–77). Academic Press.

Kroll, B., & Reid, J. (1994). Guidelines for designing writing prompts: Clarifications,

caveats, and cautions. *Journal of Second Language Writing, 3*(3), 231–255.

Kuiken, F., & Vedder, I. (2012). Speaking and writing tasks and their effects on second language performance. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition*, (pp. 364–380). Routledge. https://doi.org/10.4324/9780203808184

Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing: Towards a new rating scale. *Language Testing, 34*(3), 321–336. https://doi.org/10.1177/0265532216663991

Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, *34*, 12–24. https://doi.org/10.1016/j.jslw.2016.10.003

Lahuerta, A. (2020). Analysis of accuracy in the writing of EFL students enrolled on CLIL and non-CLIL programmes: The impact of grade and gender. *Language Learning Journal*, *48*(2), 121–132. https://doi.org/10.1080/09571736.2017.1303745

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of english. *Applied Linguistics*, *27*(4), 590–619. https://doi.org/10.1093/applin/aml029

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge.

Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, *1*, 30–41.

Lasagabaster, D. (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in Language Learning and Teaching*, *5*(1), 3–18. https://doi.org/10.1080/17501229.2010.519030

Lasagabaster, D. (2017). Integrating content and foreign language learning. *Journal of Immersion and Content-Based Language Education*, *5*(1), 4–29. https://doi.org/10.1075/jicb.5.1.01las

Lee, J. (2020). Assessing the effects of CLIL on Korean high school students' writing. *Linguistic Research*, *37*, 89–112. https://doi.org/10.17250/khisli.37..202009.004

Lee, S. M. (2020). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, *33*(3), 157–175. https://doi.org/10.1080/09588221.2018.1553186

Leki, I., & Carson, J. (1997). "Completely different worlds": EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly*, *31*(1), 39–69. https://doi.org/10.2307/3587974

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge.

Lialikhova, D. (2021). The impact of a short-term CLIL intervention project on Norwegian different ability ninth graders' oral development. *International Journal of Bilingual Education and Bilingualism, 24*(5), 671–692. https://doi.org/10.1080/13670050.2018.1509055

Lindholm-Leary, K., & Genesee, F. (2014). Student outcomes in one-way, two-way, and indigenous language immersion education. *Journal of Immersion and Content-Based Language Education*, *2*(2), 165–180. https://doi.org/10.1075/jicb.2.2.01lin

Llinares, A., & Dalton-Puffer, C. (2015). The role of different tasks in CLIL students' use of evaluative language. *System*, *54*, 69–79. https://doi.org/10.1016/j.system.2015.05.001

Lo, Y. Y., & Lin, A. M. Y. (2019). Curriculum genres and task structure as frameworks to analyse teachers' use of L1 in CBI classrooms. *International Journal of Bilingual Education and Bilingualism, 22*(1), 78–90. https://doi.org/10.1080/13670050.2018.1509940

Lo, Y. Y., & Macaro, E. (2015). Getting used to content and language integrated learning:

what can classroom interaction reveal? *Language Learning Journal*, *43*(3), 239–255. https://doi.org/10.1080/09571736.2015.1053281

Lo, Y. Y., & Murphy, V. A. (2010). Vocabulary knowledge and growth in immersion and regular language-learning programmes in Hong Kong. *Language and Education*, *24*(3), 215–238. https://doi.org/10.1080/09500780903576125

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *Modern Language Journal*, *96*(2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*, 399–432.

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. John Benjamins Publishing.

Lyster, R. (2011). Content-based second language teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 611–630). Routledge.

Lyster, R. (2017). Language-focused instruction in content-based classrooms. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom: New perspectives on integrating language and content* (2nd ed., pp. 109–123). The University of Michigan Press.

Manchón, R. M. (2011). *Learning-to-write and writing-to-learn in an additional language.* John Benjamins Publishing Company.

Manchón, R. M. (2020). *Writing and language learning: Advancing research agendas.* John Benjamins Publishing Company.

Manchón, R. M., & Williams, J. (2016). L2 writing and SLA studies. In R. M. Manchón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing*, (pp. 519–541). De Gruyter Mouton. https://doi.org/10.1515/9781614511335-029

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing*, *29*, 3–15. https://doi.org/10.1016/j.jslw.2015.06.004

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

McDonough, K., & Crawford, W. J. (2020). Identifying effective writing tasks for use in EFL write-to-learn language contexts. *Language Learning Journal, 48*(4), 469–480. https://doi.org/10.1080/09571736.2018.1465990

McDonough, K., Crawford, W. J., & De Vleeschauwer, J. (2014). Summary writing in a Thai EFL university context. *Journal of Second Language Writing*, *24*(1), 20–32. https://doi.org/10.1016/j.jslw.2014.03.001

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, *47*(4), 292–330. https://doi.org/10.1080/01638530902959943

MEXT. (2018). Koutougakkou gakkushu shidou youryo (heisei 30 nen kokuji) kaisetsu gaukokugohen [Commentary on new curriculum guidelines for senior high schools]. Retrieved from https://www.mext.go.jp/content/1407073_09_1_2.pdf

Michel, M., Révész, A., Lu, X., Kourtali, N. E., & Borges, L. (2020). Investigating L2 writing processes across independent and integrated tasks: A mixed-methods study.

Second Language Research, 36(3). 277–304. https://doi.org/10.1177/0267658320915501

Michel, M. (2017). Complexity, accuracy, and fluency in L2 production. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 50–68). Routledge. https://doi.org/10.4324/9781315676968

Mizumoto, A., & Takeuchi, O. (2010). Koukaryou to kentei-ryoku bunseki nyuumon: Toukei-teki kentei wo tadashiku tsukau tameni [Introduction to effect size and power analysis: To use statistical tests appropriately]. *Bulletin of Methodology SIG, Kansai Chapter of LET*, 47–73.

Mohan, B. A. (1986). *Language and content*. Addison Wesley Publishing Company.

Moore, P., & Lorenzo, F. (2015). Task-based learning and content and language integrated learning materials design: process and product. *Language Learning Journal*, *43*(3), 334–357. https://doi.org/10.1080/09571736.2015.1053282

Murphey, T. (1997). Content-based instruction in an EFL setting: Issues and strategies. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom: perspectives on integrating language and content* (pp. 117–131). Longman.

Nakagawa, H. (2017). Koukoukyoukasho to daigakukyouzainai no paragraph writing task no bunseki [Analysis of paragraph writing tasks in high school and university textbooks]. *Hiroshimadaigaku kyouikugaku kennkyuuka kiyou dainibu, 66*, 103–110.

Neumann, H., Leu, S., & McDonough, K. (2019). L2 writers' use of outside sources and the related challenges. *Journal of English for Academic Purposes*, *38*, 106–120. https://doi.org/10.1016/j.jeap.2019.02.002

Nikula, T., & Moore, P. (2019). Exploring translanguaging in CLIL. *International Journal of Bilingual Education and Bilingualism*, *22*(2), 237–249. https://doi.org/10.1080/13670050.2016.1254151

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Nuzzo, E., & Bove, G. (2020). Assessing functional adequacy across tasks: A comparison of learners' and native speakers' written texts. *EuroAmerican Journal of Applied Linguistics and Languages*, *7*(2), 9–27. https://doi.org/10.21283/2376905x.12.175

Ong, J. (2014). How do planning time and task conditions affect metacognitive processes of L2 writers? *Journal of Second Language Writing*, *23*(1), 17–30. https://doi.org/10.1016/j.jslw.2013.10.002

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, *19*(4), 218–233. https://doi.org/10.1016/j.jslw.2010.10.003

Ong, J., & Zhang, L. J. (2013). Effects of the manipulation of cognitive processes on EFL writers' text quality. *TESOL Quarterly*, *47*(2), 375–398. https://doi.org/10.1002/tesq.55

Ortega, L. (2015). Researching CLIL and TBLT interfaces. *System*, *54*, 103–109. https://doi.org/10.1016/j.system.2015.09.002

Osman, S. R. F., Ahmad, E. A., & Jusoff, K. (2009). Nurturing writing proficiency through theme-based instruction. *International Education Studies*, *2*(3), 140–146. https://doi.org/10.5539/ies.v2n3p140

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*(4), 590–601. https://doi.org/10.1093/applin/amp045

Pally, M. (1997). Critical thinking in ESL: An argument for sustained content. *Journal of Second Language Writing*, *6*(3), 293–311. https://doi.org/10.1016/S1060-3743(97)90016-3

Pally, M. (2001). Skills development in 'sustained' content-based curricula: Case studies in

analytical/critical thinking and academic writing. *Language and Education*, *15*(4), 279–305.

Payant, C., McDonough, K., Uludag, P., & Lindberg, R. (2019). Predicting integrated writing task performance: Source comprehension, prewriting planning, and individual differences. *Journal of English for Academic Purposes*, *40*, 87–97. https://doi.org/10.1016/j.jeap.2019.06.001

Pena, C. A., & Pladevall-ballester, E. (2020). Effects of focus on form on primary CLIL students' foreign language performance in task-based oral interaction. *Journal of Immersion and Content-Based Language Education*, *8*(1), 53–79. https://doi.org/10.1075/jicb.17028.alf

Peng, J., Wang, C., & Lu, X. (2020). Effect of the linguistic complexity of the input text on alignment, writing fluency, and writing accuracy in the continuation task. *Language Teaching Research*, *24*(3), 364–381. https://doi.org/10.1177/1362168818783341

Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism*, *15*(3), 315–341. https://doi.org/10.1080/13670050.2011.630064

Pérez-Vidal, C., & Roquet, H. (2015). The linguistic impact of a CLIL science programme: An analysis measuring relative gains. *System*, *54*, 80–90. https://doi.org/10.1016/j.system.2015.05.004

Pessoa, S., Hendry, H., Donato, R., Tucker, G. R., & Lee, H. (2007). Content-based instruction in the foreign language classroom: A discourse perspective. *Foreign Language Annals*, *40*(1), 102–121. https://doi.org/10.1111/j.1944-9720.2007.tb02856.x

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, *13*(2), 111–129. https://doi.org/10.1016/j.asw.2008.07.001

Plakans, L. (2010). Independent vs. integrated writing tasks: A comparison of task representation. *TESOL Quarterly*, *44*(1), 185–194. https://doi.org/10.5054/tq.2010.215251

Plakans, L. (2012). Assessment of integrated skills. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*, (pp.1–8). John Wiley & Sons.

Plakans, L. (2015). Integrated second language writing assessment: Why? What? How? *Language and Linguistics Compass*, *9*(4), 159–167. https://doi.org/10.1111/lnc3.12124

Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, *17*(1), 18–34. https://doi.org/10.1016/j.asw.2011.09.002

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Polat, B., & Kim, Y. (2014). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics*, *35*(2), 184–207. https://doi.org/10.1093/applin/amt013

Polio, C., & Friedman, D. (2016). *Understanding, evaluating, and conducting second language writing research*. Routledge.

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, *26*, 10–27. https://doi.org/10.1016/j.jslw.2014.09.003

Proske, A., & Kapp, F. (2013). Fostering topic knowledge: Essential for academic writing. *Reading and Writing*, *26*(8), 1337–1352. https://doi.org/10.1007/s11145-012-9421-4

Qiu, X., & Lo, Y. Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research*, *21*(6), 681–698. https://doi.org/10.1177/1362168816684368

Révész, A., Kourtali, N. E., & Mazgutova, D. (2017). Effects of task complexity on L2

writing behaviors and linguistic complexity. *Language Learning*, *67*(1), 208–241.

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. H. Long. (Eds.), *Handbook of second language acquisition* (pp.631–678). Blackwell Publishing.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL - International Review of Applied Linguistics in Language Teachings*, *43*(1), 1–32. https://doi.org/10.1515/iral.2005.43.1.1

Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning*, *61*(s1), 1–36. https://doi.org/10.1111/j.1467-9922.2011.00641.x

Roediger, I., & Guynn, M. (1996). Retrieval processes. In E. Bork & R. Bork (Eds.), *Memory* (pp. 197–236). Academic Press. https://doi.org/10.1016/B978-012102570-0/50009-4

Roquet, H., & Pérez-Vidal, C. (2017). Do productive skills improve in content and language integrated learning contexts? The case of writing. *Applied Linguistics*, *38*(4), 489–511. https://doi.org/10.1093/applin/amv050

Ruiz-Funes, M. (2015). Exploring the potential of second/foreign language writing for language learning: The effects of task factors and learner variables. *Journal of Second Language Writing*, *28*, 1–19. https://doi.org/10.1016/j.jslw.2015.02.001

Ruiz de Zarobe, Y. (2010). Written production and CLIL. In C. Dalton-Puffer, T. Nikula & U. Smit (Eds.). *Language use and language learning in CLIL classrooms.* (pp. 191–210). ProQuest. https://doi.org/10.1075/aals.7.10rui

Saiki, M., Okamoto, M., & Shibata, N. (2022). A semester-long study of second language writing development and the effects of topic familiarity and difficulty. *Nagoya JALT Journal*, *3*(1), 41–57. https://doi.org/10.37546/jaltchap.nagoya3.1-2

Saiki-Craighill, S. (Ed.). (2017). *Shitsuteki kenkyuuhou zemina-ru: Grounded thoery approach wo manabu* [Qualitative research method seminar: Learning about grounded theory approach]. Igakushoin.

Salimi, A., & Fatollahnejad, S. (2012). The effects of strategic planning and topic familiarity on Iranian intermediate EFL learners' written performance in TBLT. *Theory and Practice in Language Studies*, *2*(11), 2308–2315. https://doi.org/10.4304/tpls.2.11.2308-2315

Santos, P. (2013). Learning through in-house videos: How one Japanese college integrates subject content in its EAP program. *The Asian EFL Journal, 15*(4), 324–329.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Palgrave Macmillan.

Sasajima, S. (2011). CLIL Atarashii hassou no jugyou: Rika ya rekishi wo gaikokugo de oshieru? [CLIL New Ideas for Teaching: Teaching science and history in a foreign language?]. Sanshu-sha.

Sasayama, S., Garcia Gomez, P., & Norris, J. M. (2021). Designing efficient L2 writing assessment tasks for low-proficiency learners of English. *ETS Research Report Series*, *2021*(1), 1–31. https://doi.org/10.1002/ets2.12341

Shibata, N. (2019). The effectiveness of content-based instruction on beginners' writing skills. *JALT Postconference Publication*, *2018*(1), 352–359. https://doi.org/10.37546/jaltpcp2018-47

Shibata, N. (2021). Selected poster presentations from the American Association of Applied Linguistics conference 2021: Japanese university students' lexical diversity development in essay writing through theme-based instruction. *Language Teaching, 54*(4), 583–586. doi:10.1017/S0261444821000306

Shih, M. (1986). Content-based approaches to teaching academic writing. *TESOL Quarterly*, *20*(4), 617–648. https://doi.org/10.2307/3586522

Shih, M. (1999). More than practicing language: Communicative reading and writing for Asian settings. *TESOL Journal*, *8*(4), 20–25.

Shin, H. W., & Kim, H. J. (2014). Linguistic differences in the writing performance of

adolescent EFL learners: The influence of independent and integrated tasks. *Journal of Asia TEFL*, *11*(1), 125–148.

Shintani, N. (2019). Potentials of writing-to-learn-language activities from second language acquisition research. *Journal of Second Language Writing*, *46*(2019), 100676. https://doi.org/10.1016/j.jslw.2019.100676

Short, D. J., Echevarría, J., & Richards-Tutor, C. (2011). Research on academic literacy development in sheltered instruction classrooms. *Language Teaching Research*, *15*(3), 363–380. https://doi.org/10.1177/1362168811401155

Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*, *18*, 268–286.

Skehan, P., & Foster, P. (1996). The influence of planning and task-type on second language performance. *Studies in Second Language Acquisition*, *18*, 299–324.

Slater, T., & Beckett, G. (2019). Integrating language, content, technology, and skills development through project-based language learning: Blending frameworks for successful unit planning. *Mextesol Journal*, *43*(1), 1–14.

Snow, M. A. (2014).Content-based and immersion models of second/foreign language teaching. In M. Celce-Murcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (4th ed., pp. 438-454). National Geographic Learning.

Stoller, F. L. (2008). Content-based instruction. In N. H. Hornberger (Ed.), *Encyclopedia of language and education* (2nd ed., pp. 59–70). Springer.

Stoller, F. L., & Grabe, W. (2017). Building coherence into the content-based curriculum: Six Ts revisited. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom: New perspectives on integrating language and content* (2nd ed., pp.53–66). The University of Michigan Press.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques.* Sage Publications.

Suzuki, N. (2013). *An investigation of Japanese students' foreign language speaking anxiety in an undergraduate English-medium instruction (EMI) program*. [Unpublished master's thesis]. Waseda University.

Suzuki, N. (2019). Needs analysis for developing a theme-based language unit for students at the national institute of technology. *Bulletin of the Graduate School of Education of Waseda University Separate Volume, 26*(2), 163–178.

Suzuki, N. (2022). Effects of content-based instruction (CBI) on EFL secondary school learners' writing. *Journal of Immersion and Content-Based Language Education, 10*(1), 5–32. https://doi.org/10.1075/jicb.20006.suz

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. Madden (Eds.), *Input and second language acquisition* (pp. 235–256). Newbury House.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review*, *50*(1), 158–164.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Lawrence Erlbaum Associates.

Takagaki, T., & Tanabe, N. (2007). High school freshmen's responses to home economics conducted in a non-native variety of English: A three-year survey on content-based instruction in Japan. *The Asian EFL Journal, 9*(2), 7–16.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*(2), 439–473.

Tedick, D. J. (1990). ESL writing assessment: Subject-matter knowledge and its impact on performance. *English for Specific Purposes*, *9*(2), 123–143. https://doi.org/10.1016/0889-4906(90)90003-U

Tedick, D. J., & Wesely, P. M. (2015). A review of research on content-based foreign/second language education in US K-12 contexts. *Language, Culture and Curriculum*, *28*(1), 25–40. https://doi.org/10.1080/07908318.2014.1000923

Tejada-Sánchez, I., & Pérez-Vidal, C. (2018). Writing performance and time of exposure in EFL immersion learners: Analysing complexity, accuracy, and fluency. In C. Pérez-Vidal, S. López-Serrano, J. Ament & D. J. Thomas-Wilhelm (Eds.), *Learning context effects: Study abroad, formal instruction and international immersion classrooms,* (pp. 101–130). Language Science Press. https://doi.org/10.5281/zenodo.1300630

Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In C. Willig & W. S. Rogers (Eds.), *The SAGE handbook of qualitative research in psychology* (2nd ed., pp. 17–37). Sage Publishing.

Underwood, P. R. (2019). Writing on international relations: an action research study on designing and implementing a university CLIL course. *Toyo Eiwa Journal of the Humanities and Social Sciences, 37*, 15–54.

Valeo, A. (2013). The integration of language and content: Form-focused instruction in a content-based language program. *Canadian Journal of Applied Linguistics*, *16*(1), 25–50.

Vandommele, G., Van den Branden, K., Van Gorp, K., & De Maeyer, S. (2017). In-school and out-of-school multimodal writing as an L2 writing resource for beginner learners of Dutch. *Journal of Second Language Writing*, *36*, 23–36. https://doi.org/10.1016/j.jslw.2017.05.010

Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics*, *38*(1), 90–111. https://doi.org/10.1093/applin/amv002

Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, *21*(3), 239–263. https://doi.org/10.1016/j.jslw.2012.03.007

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Walqui, A. (2006). Scaffolding instruction for english language learners: A conceptual framework. *International Journal of Bilingual Education and Bilingualism*, *9*(2), 159–180. https://doi.org/10.1080/13670050608668639

Watanabe, Y., Ikeda, M., & Izumi, S. (2011). CLIL (Naiyou genngo togogata gakushu) Jouchudaigaku gaikokugokyouiku no aratanaru chousen daiikkan gennri to houhou [CLIL (content and language integrated learning) New challenges in foreign language education at Sophia University Volume 1: Principles and methodologies]. Sophia University Press.

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Weigle, S. C., & Jensen, L. (1997). Issues in assessment for content-based instruction. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom: perspectives on integrating language and content* (pp. 201–212). Longman.

Whittaker, R., Llinares, A., & McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research*, *15*(3), 343–362. https://doi.org/10.1177/1362168811401154

Willis, J. (1996). A flexible framework for task-based learning: An overview of a task-based framework for language teaching. In D. Willis & J. Willis (Eds.), *Challenge and change in language teaching*, (pp. 52–62). Macmillan Education.

Willis, J., & Willis, D. (2007). *Doing task-based teaching*. Oxford University Press.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. University of Hawaii Press.

Xu, Y. (2021). Investigating the effects of digital multimodal composing on Chinese EFL learners' writing performance: A quasi-experimental study. *Computer Assisted Language Learning, 0*(0). https://doi.org/10.1080/09588221.2021.1945635

Yamano, Y. (2013). Utilizing the CLIL approach in a Japanese primary school: A comparative study of CLIL and EFL lessons. *The Asian EFL Journal*, *15*(4), 70–92.

Yanagawa, K. (2017). CLIL gata jugyou no jissen ―gakushuusha ha dou uketome tasuku ha kinou suruka― [Does 'soft'CLIL work? ― How do the students perceive and do the tasks work?] *Annual Review of English Language Education in Japan, 28*, 319–334. https://doi.org/10.20581/arele.28.0_319

Yang, WeiWei. (2014). *Mapping the relationships among the cognitive complexity of independent writing tasks, L2 writing quality, and complexity, accuracy and fluency of L2 writing*. [Unpublished doctoral dissertation]. Georgia State University.

Yang, Weiwei., & Kim, Y. J. (2020). The effect of topic familiarity on the complexity, accuracy, and fluency of second language writing. *Applied Linguistics Review*, *11*(1), 79–108. https://doi.org/10.1515/applirev-2017-0017

Yang, Weiwei., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, *28*, 53–67. https://doi.org/10.1016/j.jslw.2015.02.002

Yang, Wenhsien. (2015). Content and language integrated learning next in Asia: Evidence of learners' achievement in CLIL education from a Taiwan tertiary degree programme. *International Journal of Bilingual Education and Bilingualism*, *18*(4), 361–382. https://doi.org/10.1080/13670050.2014.904840

Yasunami, S., & Lavin, R. S. (2020). *Science quest*. Seibido.

Yoon, H. J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System*, *66*, 130–141. https://doi.org/10.1016/j.system.2017.03.007

Yoon, H. J. (2021). Challenging the connection between task perceptions and language use in L2 writing: Genre, cognitive task complexity, and linguistic complexity. *Journal of Second Language Writing, 54*, 1–14. https://doi.org/10.1016/j.jslw.2021.100857

Yu, G. (2013). From integrative to integrated language assessment: Are we there yet? *Language Assessment Quarterly*, *10*(1), 110–114. https://doi.org/10.1080/15434303.2013.766744

Zhan, J., Sun, Q., & Zhang, L. J. (2021). Effects of manipulating writing task complexity on learners' performance in completing vocabulary and syntactic tasks. *Language Teaching Research, 0*(0). https://doi.org/10.1177/13621688211024360

Appendices

Appendix A: Questionnaire about the Difficulty of the Two Prompts Used in Study 1

アンケート

☆２０１８年度前期に皆さんには Group A, Group B から２つずつ英作文を書いても
らいました。<u>**どちらのグループの方が書くのが難しかったですか？**</u>数字に丸をし
てください。
　　１．Group A
　　２．Group B
　　３．どちらも同じくらいの難易度

☆そのように答えた<u>**理由**</u>を教えてください。

_____

　　　　　　　　　　　　　　　　　　　　　　　　ありがとうございました！

クラス：　　　　　　　　名前：_____

Group A

１．You are responsible for writing a report on two ways of power generation for your
engineering class. Choose two from the list below. Discuss their advantages and disadvantages,
and explain which one will be more important for Japan in the near future and why. You must
write 50 words or more.
● thermal power generation（火力発電）
● nuclear power generation（原子力発電）
● hydroelectric power generation（水力発電）
● solar power generation（太陽光発電）
● wind power generation（風力発電）
● other（その他）：
日本語訳
あなたはエンジニアの授業で２つの発電方法についてレポートを書くことになりま
した。
上記の発電方法の中から２つ選んでください。それぞれの長所と短所について論じ、
どちらが近い将来日本にとってより重要となるか、あなたの意見とその理由を述べ
て下さい。50語以上書くようにして下さい。

２．A newspaper company is hosting a high school student essay contest. The topic is "promising methods of power generation in the next 30 years." All 3rd year Kosen students will participate in this contest. Choose two methods from the list below. Discuss their advantages and disadvantages and explain which you think is more promising and why. You must write 50 words or more.

● thermal power generation（火力発電）
● nuclear power generation（原子力発電）
● hydroelectric power generation（水力発電）
● solar power generation（太陽光発電）
● wind power generation（風力発電）
● other（その他）：

日本語訳
ある新聞社が高校生エッセイコンテストを開催しています。テーマは「これからの30年間で将来性のある発電方法」です。高専の 3 年生全員がこのコンテストに応募します。上記の発電方法の中から２つ選んでください。それぞれの長所と短所について論じ、どちらがより将来性があるか、あなたの意見とその理由を述べて下さい。50 語以上書くようにして下さい。

３．Suppose you work for an electric power company. You are going to submit a report to the city council to suggest which type of power station should be promoted: (a) a solar power plant, or (b) a thermal power plant. Discuss their advantages and disadvantages, and explain which you think should be promoted and why. You must write 50 words or more.

日本語訳
あなたは電力会社に勤めています。あなたは 「(a)太陽光発電所と(b)火力発電所のどちらを推進すべきか」についての報告書を市議会に提出することになっています。それぞれの長所と短所について論じながら、どちらが推進されるべきか、あなたの意見とその理由を述べて下さい。50 語以上書くようにして下さい。

Group B は裏

１．There are many ways to communicate with people. If you communicate with your friends in the U.S. who only speak English, which means would you use, (a) video call or (b) email? Discuss their advantages and disadvantages and explain which means would be better for you and why. You must write 50 words or more.

日本語訳
今日では、たくさんのコミュニケーションの取り方があります。もし英語しか話せないアメリカの友達とやりとりをする時に、(a) ビデオ通話と (b) E メールのどちらの方法を使いますか。それぞれの長所と短所について論じ、どちらの方法があなたにとってより良いか、あなたの意見とその理由について述べて下さい。50 語以上書くようにして下さい。

２．Kosen has started to offer two options for textbooks, (a) e-textbooks in a tablet or (b) paper textbooks. You can choose either of the two. Discuss the advantages and disadvantages of them, and explain the reason for your choice. You must write 50 words or more.

日本語訳
高専では (a)タブレットを用いた電子教科書と、(b)紙の教科書の二種類を選べるようになりました。どちらか好きな方を選ぶことができますが、あなたはどちらを選びますか。それぞれの長所と短所について論じながら、その種類を選んだ理由について述べて下さい。50 語以上書くようにして下さい。

３．These days, some shops and hotels are replacing human workers (e.g., shop clerks and receptionists) with robots. Discuss the advantages and disadvantages of (a) human workers and (b) robots, and explain which you would like to receive service from and why. You must write 50 words or more.

日本語訳
近年、店舗やホテルでは、店員や受付係として、人間の代わりにロボットを用いることが増えてきました。それぞれ (a)人間の店員と (b)ロボットの長所と短所を議論し、どちらからサービスを受けたいか、あなたの意見とその理由について述べて下さい。50 語以上書くようにして下さい。

Writing Practice (Content 1)

以下について、書いてみよう！

You are responsible for writing a report on two ways of power generation for your engineering class. <u>Choose two from the list below</u>. <u>Discuss their advantages and disadvantages,</u> and <u>explain which one will be more important for Japan in the near future and why</u>. You must write 50 words or more.

- thermal power generation（火力発電）
- nuclear power generation（原子力発電）
- hydroelectric power generation（水力発電）
- solar power generation（太陽光発電）
- wind power generation（風力発電）
- other（その他）：

日本語訳

あなたはエンジニアの授業で２つの発電方法についてレポートを書くことになりました。<u>上記の発電方法の中から２つ選んでください。</u><u>それぞれの長所と短所について論じ、どちらが近い将来日本にとってより重要となるか、あなたの意見とその理由を述べて下さい。</u>50語以上書くようにして下さい。

書くときの留意点

　・適切な情報を十分に書けていますか？
　・書くべき内容が書かれていますか？
　・読み手にとってわかりやすく書かれていますか？
　・情報やアイデア同士の繋がりが表現されていますか？
　・様々な表現を使って書いていますか？

```
メモ



```

Appendix C: Content/Language Goals and Tasks for Each Lesson in Study 1

Instr-uction	Date	Content focus	Language focus	Tasks
	May7	Pre-test writing (general)		
	9	Pre-test writing (content)		
1	14	Japan's energy consumption: Introduction Characteristics of each power generation method	Vocabulary definition	Reading a brief description of each method Picture-definition matching Classification
2	21	Review of vocabulary Current breakdown of Japan's energy source	Relative clause for definition Explaining the breakdown (e.g., account for)	Assigning a label for the energy breakdown
3	23	Explaining the change of the energy breakdown from 1973 to present	Explaining the change by comparing three graphs	Writing the explanation Reading the sample explanation
4	24	Explaining the change of the energy breakdown from 1973 to present	Focusing on the language of comparison using the sample explanation (e.g., compared to/ in contrast/ although/ due to/ because of/ in addition)	Identifying the language for comparison
5	28	Learning about Japan's energy policy (S+3E), focusing on fossil fuels and nuclear power	Reading activity Learning comparative forms	Thinking about the important points when deciding which energy source to use to what extent Reading a table and filling it in
6	31	Learning about Japan's energy policy (S+3E) Compare fossil fuels and nuclear power	Learning comparative forms (e.g., connectives) Writing activity	Writing a short passage using a sentence frame to practice comparative forms and conjunctions
7	June4	Compare benefits and downsides of thermal power generation and	Speaking activity Learning comparative forms (e.g., both A and B,	Organizing information using the table used in the previous class

		nuclear power generation	however, first, second…)	
8	June7	Learning about renewable energy	Listening activity Learning expressions for expressing benefits and downsides	Watching a video and organizing information
9	22	Learning about renewable energy	Listening activity Learning comparative adverbs, conjunctions and antonyms	Watching the video again and filling in the blanks to check the expressions for benefits and downsides
10	July2	Thinking about how to achieve the 2030 goal Reviewing the forms of renewable energy	Learning expressions for expressing similarities and for description	Reading to understand pros and cons of each method Information gap task
11	10	Comparing strengths and weaknesses of solar, wind and hydroelectric power generation Making a suggestion to attain 3E+S	Reviewing all the language covered that shows compare/contrast	Wrap-up writing activity
	19	Post-test (general topic)		
	23	Post-test (CBI topic), exit questionnaire		

Appendix D: Example of Coding for the Reflective Comments in Study 1

No.	Reflecction comments	Coder 1		Coder 2		Agreement		Final codes	
		11 codes	Nega. Posi. Codes	11 codes	Nega. Posi. Codes	11 codes	Nega. Posi. Codes	11 codes	Nega. Posi. Codes
1	I can't understand. (理解できない)	10	2	10	2	1	1	10	2
2	It's so frustrating that I still don't understand all the comparisons and verb usage. (比較とか動詞の使い方がまだまだ分からないことばかりで悔しい)	2,6,8,9	2	2,9	2	0	1	2,6,8	2
3	It was good to know the latest ratios of the types of power plants in Japan. (日本の発電所の種類の割合の最新のものを知れてよかった。)	2	1	1	1	0	1	1	1
4	There were many difficult words, and it was difficult to convey what I wanted to say in English only with my English ability. (難しい単語が多かったり、自分の持っている英語力だけでは伝えたいこと、いいたいことがうまく英語にできなくて大変だった。)	6,7	2	7,5,9	2	0	1	5,6,7	2
5	I now know how to use conjunctions. (接続詞の使い方が分かった。)	8	1	2	1	0	1	2	1
6	There were expressions of comparison that I don't usually use, so I have one more expression in my repertoire that I can use. (普段使わない比べ方の表現があったので、一つ使える表現のレパートリーが増えた。)	2	1	2	1	1	1	2	1
7	I could not understand the difficult parts of the text. (文章の難しいところが理解できなかった。)	3,10	2	3	2	0	1	3,10	2
8	Having learned about comparative expressions other than the comparison (-er, -est), I wanted to use comparative expressions other than the comparison when I write comparative sentences. (比較級以外にも比べる表現を知れたので、比べる文章を書くときには比較級以外の比べる表現も使いたいと思った。)	2,6,8	1	2	1	0	1	2,6	1
9	I didn't understand many of the words so everything seemed difficult. (多くの単語がわからなかったのですべてが難しく感じた)	7,10	2	7	2	0	1	7,10	2
10	There was a point where I didn't know what to do. (何をすればいいのか分からないところがあった。)	11	2	11	2	1	1	9	2
…	…	…	…	…	…	…	…	…	…

Appendix E: Values of Kurtosis, Skewness and Shapiro-Wilk Test for Each Data Set in Study 1

	Kurtosis	Skewness	Shapiro-Wilk test
pre_CS_accuracy	3.58	-0.81	$p = 0.17$
pre_G_accuracy	2.59	-0.27	$p = 0.90$
post_CS_accuracy	2.69	-0.49	$p = 0.38$
post_G_accuracy	3.44	-0.45	$p = 0.40$
pre_CS_# of words	1.91	0.15	$p = 0.34$
pre_G_# of words	2.40	0.47	$p = 0.25$
post_CS_# of words	1.80	0.27	$\boldsymbol{p = 0.03}$
post_G_# of words	1.98	0.09	$p = 0.08$
pre_CS_DC/T	5.50	1.48	$\boldsymbol{p < 0.001}$
pre_G_DC/T	2.19	0.08	$p = 0.73$
post_CS_DC/T	2.90	0.74	$\boldsymbol{p = 0.03}$
post_G_DC/T	3.62	0.65	$p = 0.09$
pre_CS_MTLD	2.76	0.61	$p = 0.15$
pre_G_MTLD	5.21	1.54	$\boldsymbol{p < 0.001}$
post_CS_MTLD	3.06	0.72	$p = 0.06$
post_G_MTLD	3.39	0.84	$p = 0.12$
pre_CS_MLT	6.85	1.84	$\boldsymbol{p < 0.001}$
pre_G_MLT	2.83	0.64	$p = 0.20$
post_CS_MLT	6.46	1.45	$\boldsymbol{p = 0.0029}$
post_G_MLT	3.48	0.52	$p = 0.58$
pre_CS_comprehensibility	3.54	-0.40	$p = 0.17$
pre_G_comprehensibility	3.62	0.01	$p = 0.14$
post_CS_comprehensibility	1.89	-0.23	$\boldsymbol{p = 0.01}$
post_G_comprehensibility	3.21	-0.02	$p = 0.69$
pre_CS_content	2.16	0.23	$p = 0.24$
pre_G_content	2.53	0.58	$\boldsymbol{p = 0.03}$
post_CS_content	2.91	0.34	$\boldsymbol{p = 0.02}$
post_G_content	2.66	0.33	$p = 0.27$
pre_CS_Task requirements	2.40	-0.58	$\boldsymbol{p = 0.05}$
pre_G_Task requirements	1.73	-0.05	$\boldsymbol{p = 0.02}$
post_CS_Task requirements	3.31	-1.04	$\boldsymbol{p < 0.001}$
post_G_Task requirements	3.03	-0.56	$p = 0.06$
pre_CS_coherence/cohesion	2.49	-0.11	$p = 0.15$
pre_G_coherence/cohesion	2.67	0.29	$\boldsymbol{p = 0.02}$
post_CS_coherence/cohesion	2.46	0.49	$p = 0.13$
post_G_coherence/cohesion	3.48	0.75	$\boldsymbol{p = 0.03}$

Note. pre = pre-test, post = post-test, CS = content specific topic, G = general topic, DC/T = the number of dependent clause per t-unit, MTLD = lexical diversity, MLT = mean length of t-unit. The bold *p* values indicate violation of normality.

Appendix F: Teaching Materials Used in the First Class in Study 2

Academic Writing   Class 1 (2020/10/6)

What are the SDGs?
Watch an introduction video. Fill in the blanks to summarize the important points.

(The Sustainable Development Goals – Action Towards 2030 | CAFOD and SDGs
https://www.youtube.com/watch?v=9-xdy1Jr2eg) -1:26 まで

What are the SDGs?

1. When was it agreed?

2. How many countries agreed?

3. Why did they agree on this?

-    To eradicate (根絶する) extreme (   ) and (   )

-    To fight inequality

-    To tackle (   ) (   )

-    To achieve sustainable development for all   -0:26

4. What was agreed in 2000?

5. What progress was made?

6. What problems remain?

SDGs for Everyone

All human beings living on this planet should tackle the Sustainable Development Goals (SDGs). Let us consider why this is so from the perspective of environmental and social issues. Today, the Earth is facing problems such as global warming and natural disasters. These issues do not only affect one particular area, but they directly affect the safe lives of all of us who live on the planet. Therefore, it is important that each of us as inhabitants of the Earth takes responsibility for taking care of and passing on a better environment to the next generation. Let's also consider social issues. Although there are human rights violations such as poverty and inequality of educational opportunities in the world, respecting everyone's human rights will lead to a more peaceful world, benefiting all people on the Earth. Guaranteeing human rights in one country can lead to social stability, which in turn can positively affect other surrounding countries. In this way, the principles set forth in the SDGs are equally relevant to everyone in all regions of the world. Each one of us is required to act with a sense of ownership in order to achieve the SDGs.

Questions

1. What is the main point of this paragraph?

2. What are the reasons that support the main point above?

Paragraph format

①

Noriko Suzuki
October 6, 2020

②

③

SDGs for Everyone

④

All human beings living on this planet should tackle the SDGs. Let us consider why this is so from the perspective of environmental and social issues. Today, the Earth is facing problems such as global warming and natural disasters. These issues do not only affect one particular area, but they directly affect the safe lives of all of us who live on the planet. Therefore, it is important that each of us as inhabitants of the Earth takes responsibility for taking care of and passing on a better environment to the next generation. Next, let's look at social issues. Although there are human rights violations such as poverty and inequality of educational opportunities in the world, respecting everyone's human rights will lead to a more peaceful world, benefiting all people on the Earth. Guaranteeing human rights in one country can lead to its social stability, which in turn can positively affect other surrounding countries. In this way, the principles set forth in the SDGs are equally relevant to everyone in all regions of the world. Each one of us is required to act with a sense of ownership in order to achieve the SDGs.

①名前と提出日を右上に記入
②タイトルを真ん中に書く。単語の最初のアルファベットは大文字で書く。ただし、冠詞（a, an, the）、前置詞（in, at, from）は小文字で書く。
③段落の最初の文は、スペース５文字分インデントする。
④１文ごとに改行しない。改行をするのは、次の段落に移るとき。

Practice
Read the paragraph below and correct the format problems.

---

millennium goals

The Millennium Development Goals (MDGs) are eight goals that all 191 UN member states adopted in 2000 and agreed to try to achieve by the year 2015.
These eight goals mainly dealt with the problems evident in developing countries, such as hunger, illiteracy, and health.
These goals are added to the SDGs, expanding its scope. This is to cover not only more serious problems in developing countries but also problems that concern both developed and developing countries.

Appendix G: The Example Lesson Planning Concept Map Based on the Six Ts (Grabe & Stoller, 2017)

Overall syllabus based on the SixTs

Thread: SDGs and environmental protection

| Theme 1: What is SDGs? | Theme 2: Life below water | Theme 3: Affordable and clean energy | Theme 4: Life on land |

Topic 1: SDGs basic facts ⟷ Topic 2: Millennium Development Goals ⟷ Topic 3: 17 goals

Task 1 Listen and complete the introduction sentences	Text 1 SDGs/MDGs introduction video
Task 2 Read and find two key points why they are relevant	Text 2 Read a passage about the reason why the SDGs are relevant for everyone
Task 3 Introduce the basic paragraph structure by the teacher talk	Text 2 Teacher talk
Task 4 Raise awareness about the key paragraph format features	Text 2 Example paragraph

Task 1 Listen and write down key achievements of the MDGs	Text 1 SDGs/MDGs introduction video
Task 2 Read and correct a paragraph format	Text 2 Read a passage about the key features of the MDGs

Task 1 Classify the goals into similar groups to understand their aims in depth	Text 1 17 goals visuals
Task 2 Assign related goals to pictures showing problems	Text 2 17 goals visuals Pictures showing problems
Task 3 Practice writing a topic sentence	Text 3 Practice sentences related to one of the 17 goals
Task 4 Practice identifying and writing supporting sentences	Text 4 Practice sentences related to one of the 17 goals
Task 5 Practice identifying and writing a conclusion	Text 5 Practice sentences related to one of the 17 goals

236

Academic Writing Class 6 (2020/11/24)   Unit 7 Hot, powerful, and clean

Suppose you read this unit in your seminar class (ゼミ). You are going to write your opinion about the reading so that you can share it with other seminar members.

What can each of us do to reduce the amount of electricity use?

1. Write down as many ideas as you can. Why is this problem important to think about? Have you heard anything about this problem on TV or the internet?
Choose the ideas you want to expand on. Write down supporting details for each idea.
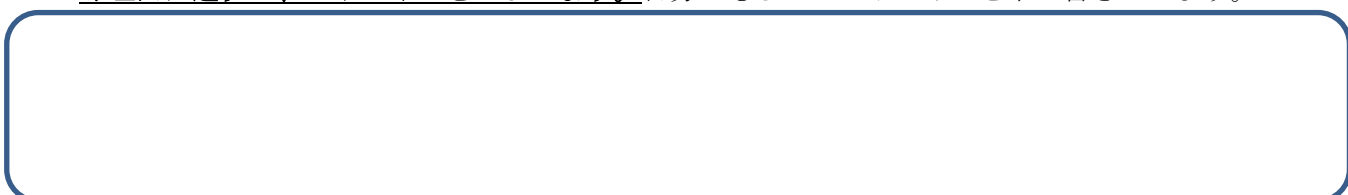
2. Think about the organization and make the essay outline below.
Topic sentence:

Supporting sentences:

Conclusion:

3. ペアのライティングプランを聞きましょう。構成は分かりやすいか、サポートで使用する例や理由は適切か、アドバイスをしましょう。自分がもらったアドバイスを下に書きましょう。

Title:

Appendix I: Interview Sheet Used in Study 2

インタビュー用紙　2021/2/　（　）　＿＿：＿～＿：＿　Name：＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

1. SDGs とライティングの仕方についての両方を学んだが、その経験はどうだったか？
（肯定的、否定的、ついた力、難しかったこと、英語学習（ライティング、文法、語彙の学習）への影響、トピックへの興味）
・（そういう風に学習していないので想像になると思うが、）ライティングのみ身近なランダムなトピックで学ぶのと比べてみて、学習に何か違いがでると思うか？

2. （学生のライティングを見せながら）タスクの差をどう感じたか（難易度、英語学習への影響）
・どちらのタイプのタスクが書きやすかったか、もしくは難しかったか。
・opinion を書くときに、reading materials は参考にしたか？どのように使ったか？
・summary を書くときに、reading materials は参考にしたか？どのように使ったか？
・reading materialsを読んでから書くことによって、自分のライティングや英語学習にどのような影響があったと思うか？

3. それぞれのタイプのライティングを 2 回ずつ行ったが、回を重ねるごとに自分のライティングの質について何か違いを感じたか？

4. 1 回目のライティングを完成させるのに大事だったステージ（最初の Reading / understanding, individual planning, ペアでの Planning）はどれか？

・授業以外での英語に触れる量と頻度（Reading, writing 両方、TOEIC の勉強していたか）

Appendix J: Values of Kurtosis, Skewness and Shapiro-Wilk Test for Each Data Set in Study 2

	Kurtosis	Skewness	Shapiro-Wilk test
With integration _accuracy	5.68	-1.53	**$p = 0.003$**
Without integration _accuracy	2.05	-0.16	$p = 0.36$
With integration_# of words	2.16	0.07	$p = 0.25$
Without integration_# of words	3.36	0.63	$p = 0.53$
With integration _DC/T	2.67	0.52	$p = 0.56$
Without integration _DC/T	5.75	1.69	**$p < 0.001$**
With integration _MTLD	2.79	0.56	$p = 0.32$
Without integration _MTLD	3.40	0.52	$p = 0.29$
With integration _LS2	2.48	-0.16	$p = 0.85$
Without integration _LS2	2.06	0.35	$p = 0.12$
With integration _MLT	2.87	0.67	$p = 0.19$
Without integration _MLT	2.06	0.32	$p = 0.26$
With integration _comprehensibility	2.46	-0.10	$p = 0.49$
Without integration _comprehensibility	2.52	-0.15	$p = 0.53$
With integration _content	2.63	-0.06	$p = 0.29$
Without integration _content	2.47	0.15	$p = 0.61$
With integration _Task requirements	3.05	-0.37	$p = 0.27$
Without integration _Task requirements	2.01	-0.17	$p = 0.18$
With integration _coherence/cohesion	2.62	-0.20	$p = 0.82$
Without integration _coherence/cohesion	2.34	-0.59	$p = 0.06$

Note. The bold p values indicate violation of normality.

Appendix K: Themes for With/Without Integration Tasks with English Translation
Themes for with integration task

Theme	Categories	Definition	Example descriptive codes	Example quotes
1. Understanding of reading	1-1. Deeper processing of the reading materials	Deeper reading and understanding materials for task completion	For the with integration task, I had the content of the textbook in mind for writing Depth of a reading material is important	The content of the textbook was in my head when I was writing (student B) I can't include the information in my writing unless I understand the meaning of the sentence. So, the with integration writing is better, even though it took time to look up (the meaning of words). (student E)
	1-2. The level of understanding affects the writing	Better understanding of reading positively affects writing and vice versa	I can't write because I can't read a reading material The ease of writing for the with integration task depends on the understanding of a reading material	I maybe didn't include (textbook information). I just don't recall citing the information from the textbook. (student K) If I understand the reading, that makes writing easier. However, if I couldn't really understand it, I need to try to understand it fully. Otherwise my citation makes little sense even if I try to integrate the source reading. (student C)
2. Reading as a resource for writing	2-1. Resource for writing content	The use of ideas from reading in writing	The role of citation as a starting point for writing Information in a source reading can be used as a concrete example → easy to write	Having read the textbook, my opinions are formed based on the knowledge I gained in the textbook (student G) Concrete information is present, and I can show some examples by writing "according to the source…" So it was easier to write. (student J)
	2-2. Source of difficulty or support for organization	Reading materials can both support organization and make it difficult	Refer to the source reading's logical development → can write in a coherent way For the with integration task, it is more difficult to make the connection between the quoted passage and students' own thoughts	If I organize my writing like this (like a source reading), it won't be exactly the same but I was able to create a writing that made sense. (student I) The process of integrating my own ideas with the source text information was difficult to some extent. If I write with my own ideas only, I could write smoothly from the beginning, but for this one, I need to align with the cited sentence, which increased the time on thinking. (student M)
3. Influence on language learning	3-1. Processing language through paraphrasing and summarizing	Difficulty of paraphrasing and summarizing and learning through them	For the with integration task, it is difficult to paraphrase and summarize Paraphrasing is a good way to learn English	I need to write things differently for citation, and that was difficult (student A) (Paraphrasing can become a) learning for myself because I need to express it in different words (student I)
	3-2. Positive impact on L2 learning	Positive impact on fluency vocabulary, grammar, and	I can write longer with citations	If I am going to complete writing, I think it is probably better to be able to write a lot, like this one (the with integration task) (student H)

| | | | reading/writing skills | For the with integration task, the ability to read and write improves | When I cite, I need to read (the source text). Then my reading skill will improve, and if I can cite, I can improve my writing skill to write about it. (student L) |

Themes for without integration task

Theme	Categories	Definition	Example descriptive codes	Example quotes
1. Variability in the use of reading	1-1. Reading used for writing	Confirming the use of ideas in reading for writing	Referenced a reading material even without a requirement for citation	I tended to use the textbook (information) as my own opinion in my writing from the beginning (student D)
	1-2. Reading not used for writing	Using students' own ideas rather than ideas from reading	Did not use a reading material so much for the without integration task Wrote from my own experience and knowledge	I did not use (the reading material) so much, but wrote based on my experience and knowledge (student G)
2. Impact on organization	2-1. Smooth organization	Writing flows smoothly by using students' own ideas	The argument flows more smoothly if I write with my own opinion without source integration	I can start writing with my previous knowledge. So, it was easy for me to connect ideas for writing. I knew what to write next. (student F)
	2-2. Detrimental impact on organization	Increased organizational difficulty	It takes time to organize writing	It was necessary to search for information in my memory to use it for my writing, and it took time to organize the content before writing (student H)
3. Impact on content	3-1. Freedom in choosing the content	Less regulation positively perceived by students	I like the freedom to write without source integration	For the without integration task, there were fewer requirements. (For the with integration task) I shouldn't use the same words (from the reading material), but there was no requirements like that. So, this task type was easier for me to write than the other one. (student B)
	3-2. Difficulty in creating content	Difficulty coming up with their own ideas for writing	Hard to think about what to write More to think about (structure, vocabulary, content)	When I need to write my opinion, I can't write without background knowledge (student A) Because I need to think and write it myself, the without integration task has a lot more to think about compared to the with integration task. (student D)

Themes concerning both with and without integration tasks

Theme	Categories	Definition	Example descriptive codes	Example quotes
1. Reading material and the use of vocabulary / expressions in it	1-1. The use of vocabulary / expressions in reading for writing	Referring to reading for vocabulary / expressions for both types of writing	Find new expressions and grammar by having a source text Pick up and use words from a reading material	(I found) expressions, grammar and different ways of writing, maybe (student L) I picked up vocabulary from the textbook and was able to write up my essay (student M)
	1-2. Vocabulary / grammar learning unaffected by reading	Vocabulary / grammar learning in writing unaffected by reading regardless of task types	Doubtful that grammar skills have increased and reading does not make any difference Did not think about vocabulary learning	About grammar, reading the material did not simply lead to the use of that grammar in the reading for writing (student K) Vocabulary was not influenced so much (because of the use of the reading material). I can only use vocabulary that I know how to use. So, I didn't think or deeply consider (the influence on vocabulary). (student A)
2. Perceived task similarities and differences	2-1. Similarities between the two tasks	Similarities across tasks in terms of the writing process, difficulty, and use of reading	I don't feel much of a task difference because it's the same for both tasks in terms of reading to write Equal difficulty for both tasks	After all, (both tasks required) reading and understanding of it, and then I wrote about what I thought about. So, I think I performed the task in the same way (student D) Both were difficult (student I)
	2-2. Differences between the two tasks	Differences across tasks in terms of the writing process and skills to be gained	The processes of writing for with and without integration tasks are very different Different language skills are trained for both tasks and both are good	I wrote differently for the two tasks. It was completely different (student M) Maybe both have good points. It's not either one is good, but aspects or skills to be trained are different. (student J)

Appendix L: Themes for Students' Perceived Learning Experiences in the Current CBI Course with English Translation

Theme	Categories	Definition	Example descriptive codes	Example quotes
1. Difficulty due to content, language, or both	1-1. Difficulty arising from novelty of the content	Difficulty due to the newness of content, even in Japanese	The SDGs themselves are quite social and difficult content	The SDGs themselves have a social aspect and are difficult (Student G)
			It is difficult to write about the SDGs even in Japanese	I found it difficult to write about the SDGs in Japanese (Student L)
	1-2. Difficulty arising from writing and reading skills	Lack of experience in writing and difficulty of reading due to unfamiliar words	Not being used to writing in English affects difficulty rather than content	I think the difficulty simply came from writing in English rather than the theme (Student J)
			Forgetting the grammar for writing	English writing was difficult because I didn't remember grammar. It took a long time to write one (Student B)
			Words in a reading are difficult	There were many specific terms and I felt difficulty at first (Student C)
	1-3. Difficulty coming from the combination of content and language learning	Increased difficulty of learning about new content, and writing about it in an unfamiliar language	CBI is highly challenging	It was difficult to learn a new content and write about it at the same time (Student J)
			Reading was difficult because I didn't know about the SDGs	I didn't know about the SDGs at all at the beginning, so I didn't understand the content by reading about it in English. It was very difficult. (Student F)
			The concept of what proper nouns represent is difficult	I could find the words when I looked them up, but the proper nouns were too difficult to understand and the phrases were difficult to understand (Student K)
2. Positive effects of having content focus for language and affect	2-1. Positive effects on writing and reading by having the SDGs as content and a comparison with familiar topics	Connection between content learning and writing/reading development and comparison of learning with the use of familiar topics	After gaining new content knowledge, writing becomes easier	I think if there is that kind of theme, that makes writing easier, even in English. Learning about a new thing and keeping it in mind while writing makes writing easier. (Student J)
			There are connections between themes → leads to ease of writing	I started from no knowledge about the SDGs. When I studied about deforestation, the content reappeared in the ocean topic, for example. They share common topics, so it was a little easier to write. (Student D)
			Read by mood rather than reading accurately (tolerate ambiguity)	For grammar, I didn't completely understand why the sentence was as it was. But when I was reading, I tried to feel the nuance, and I looked at words and guessed the meaning. (Student D)
			Familiar topics tend to elicit monotonous	

			sentences that cannot be expanded upon	I think topics like my favorite food will end up plain sentences. I can't dig into these topics deeply. It may be difficult to develop writing with these topics. (Student E)
	2-2. Wide range of vocabulary input	Various vocabulary input and learning of it through repeated exposure	The SDGs can provide more word input Same words appear repeatedly and that makes memorization easy	I had learned only a few words before. But by reading various texts, I found that there are various new words. (Student C) There were so a lot of grammar and new words, and some of the vocabulary were naturally learned by using them repeatedly (Student D)
	2-3. Value of learning about the SDGs	Expressing values of learning about the SDGs and appreciating the connection between the students' previous learning and the SDGs	It was good to learn about a new content The current class connected with the past experience of learning a little about the SDGs	I didn't know much about the SDGs, so I think there were many positive aspects (Student H) I had learned only a little about the SDGs in the first technical course of my department in the fourth year, so I didn't know much about it. Then, I felt I learned a lot in the English class which incorporated the learning about the SDGs. (Student I)
	2-4. A sense of achievements	A sense of achievement learning about the SDGs and writing about them	It was difficult, but I gained language skills The feeling that I did my best	The SDGs itself is a social and difficult topic. In addition, I don't usually write my opinion in English, so I thought the combination of these made the course very hard. But, I think I gained a lot of skills. (Student G) I made use of my background knowledge and tried hard (to write), so it was good in the end (Student A)