JAAL IN JACET PROCEEDINGS

VOL. 2

JAAL in JACET Proceedings

Volume 2

The Second JAAL in JACET Conference (JAAL in JACET 2019)

30th November 2019, Takachiho University, Tokyo, Japan

JACET

Contents 目次

【Articles 論文】

Bridging the Gap Between EGAP and ESAP Through Student-Selection of Learning Materials

Akari Hirano* Noel H. Vincent** Yosuke Sasao*** Sachi Takahashi****

* *Kyoto University*
hirano.akari.24n@st.kyoto-u.ac.jp
***Kyoto University*
noelvincent@protonmail.com
****Kyoto University*
sasao.yosuke.8n@kyoto-u.ac.jp
*****Kyoto University*
takahashi.sachi.4n@kyoto-u.ac.jp

Abstract

Highly discipline-specific skills, such as abstract writing, are often omitted from English for General Academic Purposes (EGAP) curricula and left to the domain of English for Specific Academic Purposes (ESAP). To address the lack of discipline-specific content for EGAP courses, we explored how involving students in the selection of their own learning materials could serve to bridge the gap between the realms of EGAP and ESAP by providing authentic content that is relevant to learners of varied academic backgrounds. The present study reports on a method for fostering the academic writing ability of learners through their involvement in a materials-selection process. The method was taught through an online course in which Japanese undergraduates learned about research paper abstracts by selecting authentic and personally-relevant examples from a database of online journals. Students subsequently performed analyses of these abstracts to understand their move structure, followed by online discussions. The course culminated in students writing their own abstract through a final assignment. To evaluate the effectiveness of our method, we analyzed the work product and feedback of 21 participants. Our findings indicate that student-selection of writing examples contributed to learner motivation, as well as understanding of writing structures.

Keywords: EAP, abstract writing, move analysis, computer-mediated learning

1. Introduction

English for General Academic Purposes (EGAP) and English for Specific Academic Purposes (ESAP) represent two of the most common approaches to teaching English academic writing, and there has been some debate as to exactly what roles they should play within the domain of English writing instruction. Whereas ESAP focuses on writing concepts and tasks specific to a single academic discipline, EGAP seeks to impart writing skills that are more generic and broadly applicable. Despite their differing pedagogical foci, each approach is indispensable to learners, who require both the generic and transferable writing skills learned through EGAP (Leki & Carson, 1994; Wingate & Tribble, 2012), and instruction linked to specific academic disciplines (Bruce, 2011; Hyland, 2007).

Although both EGAP and ESAP skills are crucial to the academic success of learners, EGAP courses may be the only instructional option available (Flowerdew, 2016; Wette, 2014) in many higher educational contexts due to institutional constraints. Administrators, when confronted with a choice between implementing ESAP or EGAP, may struggle to resist the 'one size fits all' approach to academic writing offered by EGAP. Regrettably, these institutional decisions deprive learners of the discipline-specific writing skills that ESAP courses excel at teaching. The general focus of EGAP courses means that they can struggle to cater the needs of all students, and its generic learning materials can lack authenticity, as it is designed specifically for language learning purposes. Additionally, materials selected by teachers may feature examples that lack relevance to student needs and interests (Mishan, 2005).

To address the lack of authentic and discipline-specific materials in EGAP courses, we explored how involving students in the selection of their own learning materials could serve to bridge the gap between the realms of EGAP and ESAP by providing authentic content that is relevant to learners of varied academic backgrounds. Our approach aimed to increase exposure to discipline-specific content through student contributions, which are thought to raise genre-awareness (Hyon, 2016; Swales & Feak, 2000) and increase the motivation and engagement of students (Flowerdew, 2016).

2. Literature Review
2.1 EGAP and ESAP Writing

The EGAP versus ESAP debate has been a controversial

issue in the field of English for Academic Purposes (EAP). Although both are essential for students, it is important to understand the roles served by either approach. ESAP is thought to increase students' motivation, as students may feel more eager to study writing skills pertaining to their interests (Flowerdew, 2016). Furthermore, gaining familiarity with discipline-specific content is of equal importance to attaining generic writing skills, since the mechanics of writing cannot be divorced from their context. Learning academic writing skills should thus ideally be connected to the learning of discipline-specific content. Indeed, researchers have highlighted the importance of raising genre-awareness using corpus of varied disciplines (Hyland, 2014; Nesi & Gardner, 2012).

Because of the benefits associated with an ESAP approach, a hybrid method incorporating aspects of EGAP and ESAP is suggested in instances where standalone ESAP courses are unfeasible (Flowerdew, 2016). It has been proposed that EGAP and ESAP be united under a unified EAP framework (Tajino, 2004; Tajino, & Suiko, 2005). In such a context, teachers are advised to incorporate discipline-specific content shared by students, which may raise awareness of differences between text types across varied disciplines (Hyon, 2016; Swales & Feak, 2000).

Furthermore, it is thought that encouraging student contribution increases exposure to discipline-specific input and raises the motivation and engagement of students (Flowerdew, 2016). This way of thinking motivated our study, which examines whether such activities have similar effects on the motivation of students who lack prior exposure to a particular academic genre. Completion of certain genre-awareness raising tasks, such as move analysis, can help prepare learners for the actual production of academic writing (Hyon, 2016). For this reason, our approach incorporates tasks that raise awareness of textual patterns by utilizing exemplars chosen by students. Some literature on EAP teaching has recommended incorporating discipline-specific content through student contributions. However, these general recommendations did not describe specific methods of how such content might be effectively implemented within a particular context. Therefore, the present study examines the efficacy of this method on first-year undergraduate students lacking prior knowledge of academic genre distinctions.

2.2. Analyzing Move Structure in Research Paper Abstracts

Swales (1981) defined a "move" as a section of text, which serves a specific communicative function. The term and concept are often used in a genre analysis to identify rhetorical patterns. For example, there are certain moves commonly used in introductions of research articles, and certain linguistic patterns can be found within each move that serve specific communicative purposes (Swales, 1981, 1990). The communicative purpose of an abstract, for instance, is to convince potential readers that its contents are worth reading. The length requirements imposed on abstracts mean that they tend to serve as a "selective representation" of a paper, rather than its detailed description (Hyland, 2014, p. 64). What merits inclusion in an abstract is largely a function of the knowledge and expectations of its intended audience. For instance, medical paper abstracts can often forgo lengthy introductions when papers concern the treatment of well-known medical phenomena, such as heart disease or cancer. By contrast, papers regarding educational methods may require longer introductions in order to adequately frame the background and scope of an inquiry. In this way, structures can often differ greatly by discipline (Yang & Allison, 2004), which can make the instruction of abstract writing a challenge.

While abstracts can display a wide range of variation, Hyland (2014) puts forward a model for describing abstracts in terms of five constituent moves: 1. Introduction, 2. Purpose, 3. Method, 4. Product (or 'results'), and 5. Conclusion (or 'discussion'). In a typical abstract, an introduction section is included to explain the context in which the research was undertaken. This is followed by the purpose section, which describes the intention of the researcher(s). Next, the method describes the research design, experimental procedure, and means of data collection, followed by a product (or 'results') section which summarizes the paper's findings. Finally, the conclusion section presents the authors' interpretation of research findings and their implication for future research.

Another model for describing the composition of abstracts is known by the acronym IMRD or IMRAD. This model, which describes abstracts in terms of 'Introduction,' 'Methods,' 'Results,' and 'Discussion,' has become a widely adopted framework endorsed by the American Psychological Association (APA). Despite the popularity of IMRAD, however, the abstracts of many papers still diverge from this format. Yang and Allison (2004) note that it is common for papers in the field of applied linguistics to omit the 'Discussion' and 'Conclusion' sections from their abstracts. Due to these variations in move structure across disciplines, it has been advised that a generic model be employed for the sake of academic writing instruction. The curriculum in our study therefore adapted Hyland's (2014) model for the purpose of analyzing abstract structure. Although this generic model was employed for the sake of efficient instruction, students require exposure to various abstracts in order to determine which fit the model. As the structure of abstracts can vary by disciplines, it is crucial to understand the generic structure and how it differs from discipline-specific structure. Thus, we explored the ideal balance of generic and discipline-specific skills taught in lessons.

3. Purpose of Study

3.1 Incorporating Discipline-specific Skills in EGAP

While teachers of EGAP appreciate the importance of writing research abstracts, there is no consensus among practitioners as to exactly when this skill should be taught. Research regarding the perceptions of EGAP instructors in Japanese higher education indicated that 45 percent think abstracts should be taught through ESAP courses, while others think it should be a part of EGAP curricula (Maswana, Takahashi, Kanamaru, Sasao, & Tajino, 2018). Consequently, the skill of writing abstracts is taught inconsistently from institution to institution (or even from department to department), resulting in students with different levels of preparedness, and further compounding the difficulty of

teaching academic writing.

The range of views expressed by teachers' regarding this matter may be due to the difficulty of introducing discipline-specific skills through EGAP. One of these skills is abstract writing, which requires writers to learn and apply the writing conventions of specific discourse communities. Although EGAP courses aim to instill general academic writing skills, the use of inauthentic writing samples that have been specifically designed for learning purposes may lead to overgeneralization of writing concepts and fail to raise awareness of the differences between various disciplines and/or genres.

Thus, in order to address this problem, the present research explored the effectiveness of incorporating authentic discipline-specific content into an EGAP course by involving students in the material-selection process. Toward this end, we endeavored to design a course on abstract writing skills that was suitable for students of varied proficiency levels and academic backgrounds. By providing the course through an online learning management system (LMS), our hope was that students whose departments or institutions do not offer instruction on abstract writing could access and complete the course as a supplement to their existing education and independently study academic writing skills. Ultimately, however, the course design presented here was modified to incorporate some in-class instruction. Although the course was unable to provide a fully autonomous learning experience, the online platform allowed learners to share student-selected materials and comments, which seems to have increased student engagement in the learning process.

3.2 Assessing the Benefits of Student-Selected Materials

Our approach leverages content selected and or generated by students. These consisted of forum comments on the LMS and authentic writing samples that students selected from online journal databases.

The present study explored the effects of this approach on students' understanding of the rhetorical features and structural patterns of research paper abstracts, as well as its influence on student engagement and motivation. The following hypotheses regarding anticipated benefits were examined:

1. Student-selected, authentic writing samples can promote student engagement and motivation.
2. Exposure to a wide variety of abstracts can increase understanding of abstract structure and features.

4. Methods and Procedure

This research explored the effects of learner-selected materials on learner engagement, motivation and understanding of writing concepts in a 4-week course regarding the writing of research paper abstracts.

4.1 Participants

The participants were first year pharmacology majors at a university in Japan ($N=21$) enrolled in an English Writing & Listening class. The majority of students ($n=19$) responded to a post-study questionnaire regarding their learning experience.

Most of the respondents ($n=17$) did not have any prior experience writing research paper abstracts and lacked knowledge regarding the features of abstract move structure prior to taking this course.

4.2 Data Collection and Analysis

Both qualitative and quantitative data were collected through an online questionnaire ($n=19$), semi-structured interviews ($n=5$), students' written assignments, and online contributions, such as comments posted during student discussions. A four-point Likert scale was employed for the questionnaire (Strongly Agree, Somewhat Agree, Somewhat Disagree, Strongly Disagree). Students' motivation and engagement were measured through the questionnaire and interviews whereas submissions of essays and online comments were used to evaluate how the course helped to improve students' writing performance.

4.3 Course Description and Procedure

The present study was conducted within the context of an existing 15-week English writing and listening course which required students to attend a 90-minute class once per week. Content related to writing research abstracts was presented from week 10 through week 13 in tandem with standard course assignments and activities. The approach blended online (out of class) and face-to-face instruction (in class).

The lessons were designed around having students independently select and analyze learning materials through the learning management system (LMS). Students then offered insights and observations from their analysis in comment threads on the LMS.

Bilingual instructions and descriptions were provided on the LMS in English and Japanese.

The course was administered in six steps: 1. Prior-Knowledge Assessment, 2. Abstract Search and Discussions, 3. Recognizing Abstract Sections, 4. Writing Abstracts, 5. Revising Abstracts, 6. Final Submission and Grading.

Table 1. Overview of Abstract Writing Lessons

	Objectives		Activities
1	Evaluate Prior-Knowledge		Students:
		A)	Speculated about the function of abstracts
		B)	Shared thoughts with peers through online comments
2	Find and Discuss Abstracts	A)	Searched online databases
		B)	Posted abstracts related to personal research interests
		C)	Shared observations
3	Recognize Abstract Sections	A)	Read explanations
		B)	Took a quiz
		C)	Color-coded abstracts
		D)	Shared observations
4	Write an Abstract	A)	Learned how to use Academic Phrasebank
		B)	Drafted abstracts based on authentic materials
5	Revise an Abstract	A)	Learned about tense usage and hedging

		B)	Did anonymous peer review with a checklist
6	Submit the Final Assignment	A)	Submitted revised abstracts online
		B)	Awaited grades and feedback from instructors

In Step. 1, students speculated about the function of abstracts and shared thoughts with peers through online comments. This step was necessary for both students and teachers to understand students' prior knowledge and experience regarding research papers.

In Step. 2, students searched online journal databases, then posted abstracts related to personal research interests. Students then shared observations regarding the structure of abstracts and frequently occurring expressions.

In Step. 3, students read explanations regarding move structure, took a multiple-choice quiz to identify moves, color-coded abstracts (see Appendix A, Example of a Color-coded Abstract. [1]), and shared their observations regarding the structure abstracts. The course materials provided explanations of move structure adapted from Hyland's (2014) model (Introduction, Purpose, Method, Results (Product), and Conclusion). The term "Product" is changed to "Results" in our materials for clarity of instruction. Students were required to mark each section in its designated color (red for Introduction, orange for Purpose, yellow for Method, green for Results, and blue for Conclusion).

In Step 4., students drafted their own abstracts based on a sample research paper provided by the researchers. The provided sample was modified for instructional use by partially redacting its introduction and conclusion to discourage paraphrasing. Students were instructed to include all five sections (Introduction, Purpose, Methods, Results, Conclusion) introduced in the lessons. To assist in the writing process, students were introduced to an online corpus known as the Manchester Academic Phrase Bank (Morley, 2018a), which provides academic expressions sourced from authentic academic papers and categorized according to the rhetorical analysis approach of Swales (1990). Students were highly encouraged to use this resource, since the generic phrases contained within the corpus allow writers to use the expressions without risk of plagiarism (Morley, 2018b).

In Step. 5, students learned about tense usage and hedging strategies before completing a self-check and peer-review process to revise their drafts. The in-class peer-review activity was implemented using a checklist that explains assessment criteria (see Appendix B: Evaluation Checklist). Peer-review was conducted anonymously to elicit more honest feedback.

In Step. 6, students submitted their revised abstracts online. The researchers provided feedback and graded submissions based on the same checklist criteria that students used during peer-review (see Appendix C: Student Essay with Feedback).

5. Findings and Discussion
5.1 Benefits of Student-Selected Content
Our findings were consistent with previous studies in which student-generated learning materials promoted engagement and motivation in learning tasks (Lambert et al.,

2017). Our questionnaire (n=19) results indicated that student-selected writing samples were relevant to participants' academic needs, with 16 respondents reporting that they gained exposure to research papers related to their current university studies (Figure 1). In a follow-up interview (n=5), two students independently mentioned how a writing sample about Alzheimer's disease was relevant to their personal research interests.
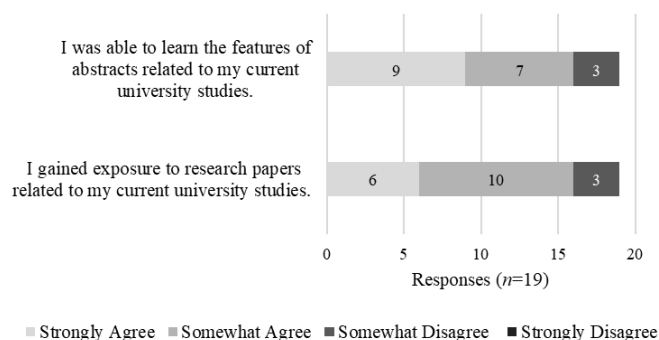
Figure 1. Discipline-specific input.

Student collaboration also seems to have contributed to students' learning. Two interviewees expressed that "active learning" aided their understanding. One interview participant indicated feeling "more actively involved in the learning process when taking an online-mediated class," indicating that online discussion contributed to student engagement.

Another indicator of increased student motivation is that nearly all questionnaire respondents (n=18) reported that the skills learned in the lessons will be of use to them in the future. This indicates that having students select personally-relevant writing samples may have allowed them to better imagine the future benefit of writing research paper abstracts.

5.2 Benefits of Varied Exposure
The findings echoed previous findings of genre awareness studies in that exposure to a wide variety of abstracts increases understanding of their structure and common features. It is suggested that abstract writing instruction include a wide range of model exemplars to help learners understand writing concepts such as genre (Hyland, 2007) as presenting a limited number of exemplars may lead to overgeneralization of rhetorical features (Wette, 2014).

When asked whether they were "able to learn the features of abstracts," a majority (n=17) of questionnaire respondents responded "Strongly Agree" or "Somewhat Agree," despite the fact that only 7 students indicated prior knowledge of research abstracts (Figure 2). The same number of respondents (n=17) indicated that viewing student-submitted abstracts helped them to complete their final assignment.
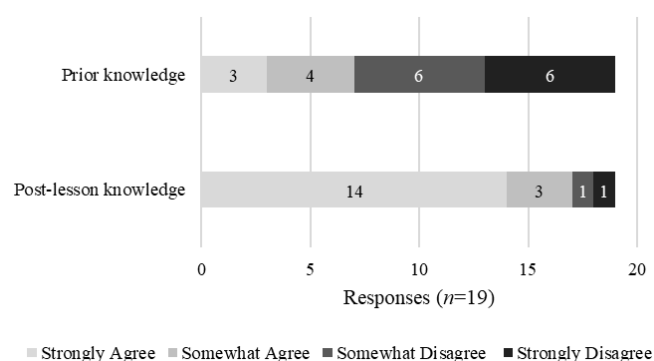
Figure 2. Knowledge regarding abstract sections.

In addition to the questionnaire results, student comments on discussion threads (Step. 3: Lesson 2, Task 2) demonstrated a structural understanding of patterns across the abstract writing samples. Students commented that "the order of the sections was mostly the same for all abstracts," and "the content may differ, but I felt that the [abstract] flow is consistent," demonstrating how students successfully learned the move patterns of abstracts from student-selected examples. The following comment, taken from the final assignment indicates that some students also referred to abstracts posted by other students when writing their abstracts:

> "During peer review, someone pointed out that the abstract should include a description of both of the tasks (employed in the study), but I think this will make the abstract overly lengthy and detailed. I also looked at abstracts posted during our second lesson, but there were not many abstracts with specific descriptions of experimental methods. For this reason, I would like to submit this without adding detailed descriptions of methodology."

This comment shows that abstracts and move analyses shared online informed the writing decisions made by students during their final assignment.

Additionally, the Manchester Academic Phrase Bank corpus served a vital role in increasing students' exposure to academic vocabulary and phrases. A majority of questionnaire respondents (n=15) reported that they used phrases from the corpus in their final abstract. Indeed, students' written work often included phrases cited from the website. Additionally, 13 respondents selected "strongly agree" to the following statement: "the Phrasebank helped me to write a better abstract."

5.3 Opportunities for Additional Intervention

On average, students performed well in the course, with an average final score of approximately 8.9 out of 10 (See Figure 3) indicating that most students benefitted from our course's approach to academic writing. However, analysis of several students' final writing assignments revealed that additional teacher intervention and scaffolding may be required for students to derive maximum benefit from student-selected examples.

Although the corpus and authentic materials seemed to have increased exposure to a wide range of abstracts, the wide range of examples shared by students may have made it difficult to distinguish appropriate examples from inappropriate ones. This is because learner-selected abstracts often include deviated from the model structure that they were taught in the lessons. As one interviewee explained, "I could not tell if the abstracts posted online were good examples or not."

In order to address this problem, teachers need to make it clear that there is a discrepancy between the models that students learn and discipline-specific writing structures found in authentic research paper abstracts. When using a wide variety of authentic materials, teachers need to help students evaluate the materials to determine whether they are relevant and consistent. Although the majority of students learned the skills successfully, some students may need such additional instructor support, making it difficult to implement a fully-autonomous writing course.

Submission of students' comments and assignments were analyzed to examine how students' selection of writing samples affected learning outcomes. Submission of students' final written tasks showed that many of them successfully completed the assignment using online resources and materials posted by their peers.

One student (see Figure 3, Student A), however, did not do well on his final assignment, having only scored 6.5 out of 10. The submitted final assignment did not include a purpose or results section, which lowered the total score by two points. The abstract sections were also not ordered properly and it included improper register (e.g. informal expressions, contractions, first and second person pronouns), leading to further point reduction (-1.5 points). Color-coded move analysis from Step 3 showed that the student chose an abstract with only three sections (introduction, purpose and conclusion) excluding methods and results sections.

As the abstract the student chose lacked the relevant sections, the student may have misunderstood the assignment at this critical stage and based his writing on the example he initially chose, which lacked criteria necessary for the final assignment. This hypothesis is further supported by Student A's comment from Step3: Task 3, Discussion 2, in which he wrote that "the abstracts do not necessarily include all of the sections (moves), but they had the same structure with introduction, purpose and methods, results, conclusion."

The analysis of other student submissions, however, revealed cases in which a student was not affected by the selection of an inappropriate abstract sample. Despite choosing an abstract which did not adhere to the five-move structure taught in the course, (Step3, Task 3 Discussion 2), Student B scored 9 out of 10, missing only one point for improper grammar. As she chose an abstract from a review paper, the abstract was missing two sections (a method and conclusion). Her choice of sample, however, did not seem to adversely affect her understanding of move structure. Perhaps this is because she made sure to carefully examine multiple exemplars shared by other students, rather than focusing on her own selection.

Student	A	B	Avarage
Length	1	1	0.952
Tense	1	1	0.928
Conjunctions	1	1	0.952
Register	0.5	1	0.928
Introduction	1	1	0.952
Purpose	0	1	0.904
Methods	1	1	1
Results	0	1	0.904
Conclusion	1	1	0.952
Implication	1	1	0.976
Grammar	0	-1	-0.119
Confusing	0	0	-0.19
Misordered	-1	0	-0.095
Incomplete	0	0	-0.071
Subtotal	7.5	10	9.452
Deductions	-1	-1	-0.476
Total	6.5	9	8.976

Figure 3. Scores of final assignments by sections.

In future implementations of this course, instructors should require that students view and analyze multiple submissions to reduce the influence of exemplars which may not adhere to the target model.

6. Conclusion

Although EGAP courses aim to provide general academic skills, exposure to a small number of inauthentic writing samples may lead to overgeneralization of writing concepts, and unawareness of the differences between discipline-specific writing styles. In order to address this need for authentic and discipline-specific content in EGAP, we had students select their own writing examples, analyze their structure and share them with peers in order to gain increased exposure to writing samples across disciplines. The sharing and structural analysis of abstracts through color-coding provided students with a visual means of recognizing move patterns.

Results from the student questionnaire and interviews indicated that student-selection of writing examples and student discussion contributed to learner motivation and engagement, as well as understanding of writing structures. However, analysis of student writing revealed that additional teacher intervention could help students to avoid forming misconceptions and derive more benefit from peer-selected exemplars. In order to address this issue, teachers should intervene as necessary to help students identify relevant examples by curating the abstracts chosen by students — making sure to present straightforward exemplars before introducing variations (Wette, 2014). Alternatively, teachers may also provide students with a carefully selected model abstract which adheres to the target structure in addition to the authentic samples selected by students (Hyland, 2007; Macbeth, 2010).

The current study echoed previous studies in that increased exposure to discipline-specific input by encouraging student contribution can raise the motivation and engagement of students (Flowerdew, 2016). Researchers have suggested that aspects of ESAP be integrated into EGAP curricula (Flowerdew, 2016) by incorporating discipline-specific content shared by students (Hyon, 2016; Swales & Feak, 2000). The present study explored the benefits of implementing this suggestion within a Japanese university classroom featuring students with limited experience writing academic papers in English.

The final results of our study indicate that while some learners may require additional scaffolding to identify and produce competent writing, the majority of participants seemed to benefit from the course method—with students indicating their understanding of structural patterns and increased motivation toward the subject matter. The present study demonstrates how methods which incorporate student-selected authentic materials have the potential to help even novice learners gain fundamental competency in academic writing skills.

Notes
[1] As the students used authentic research papers online, a sample abstract created by teachers is provided as an example instead of students' posts. The sample was partially redacted for copyright protection.

References
Bruce, I. (2011). *Theory and concepts of English for academic purposes.* Houndmills: Palgrave Macmillan.

Flowerdew, J. (2016). English for Specific Academic Purposes (ESAP) Writing: Making the case. *Writing & Pedagogy, 8*(1).

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing, 16*, 148–164.

Hyland, K. (2014). *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press. https://ebookcentral.proquest.com

Hyon, S. (2016). Using genre analysis to teach writing in the disciplines In J. Flowerdew, & T. Costley (Eds.), *Discipline-specific writing: Theory into practice* (pp. 77–94). https://ebookcentral.proquest.com

Lambert, C., Philp, J., & Nakamura, S. (2017). Learner-generated content and engagement in second language task performance. *Language Teaching Research, 21*(6), 665–680.

Leki, I., & Carson, J. G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly, 28*, 81–101.

Macbeth, K. (2010). Deliberate false provisions: the use and usefulness of models in learning academic writing. *Journal of Second Language Writing, 19*, 33–48.

Maswana, S., Kanamaru, T., & Tajino, A. (2015). Move analysis of research articles across five engineering fields: What they share and what they do not. *Ampersand, 2*(C), 1–11. doi:10.1016/j.amper.2014.12.002

Maswana, S., Takahashi, S., Kanamaru, T., Sasao, Y., and Tajino, A. (2018, December). *Eigo Gakujutsu Ronbun ni Hitsuyo na Gino to Chishiki* [Skills and Knowledge Required for EAP Writing: From Perspectives of Teachers]. A paper session presented at the 192nd East Asia English Education Association, Fukuoka, Japan.

Mishan, F. (2005). *Designing authenticity into language learning materials*. Bristol: Intellect.

Morely, J. (2018a). *University of Manchester Academic Phrase Bank.* http://www.phrasebank.manchester.ac.uk/

Morley, J. (2018b). *Academic PhraseBank: A compendium of commonly-used phrasal elements in Academic English.* The University of Manchester: Manchester.

Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education.* Cambridge: Cambridge University Press.

Swales, J. (1981). *Aspects of article introductions* (Aston ESP Research Report No. 1). Birmingham: Language Studies Unit: University of Aston.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. M., & Feak, C. (2000). *English in today's research world: A writer's guide.* Ann Arbor: Michigan.

Tajino, A. (2004). Nihon ni okeru daigakueigo kyouiku no mokuteki to mokuhyou ni tsuite [Regarding the objective and the purpose of university English education in Japan]. *Multi Media News*, (7), 11–22.

Tajino, A., & Suiko, M. (2005). Daigaku eigo kyouiku eno teigen: Karikyuramu kaihatsu eno shisutemu apuroochi [A proposal for university English language education: A systems approach to curriculum development]. *Korekara no daigaku eigo kyouiku [The future of university English language education]*, 1–46.

Wette, R. (2014). Teachers' practices in EAP writing instruction: Use of models and modeling. *System, 42*(1), 60–69. https://doi.org/10.1016/j.system.2013.11.002

Wingate, U., & Tribble, C. (2012). The best of both worlds? Towards an English for academic purposes/academic literacies writing pedagogy. *Studies in Higher Education, 37,* 481–495.

Yang, R., & Allison, D. (2004). Research articles in applied linguistics: Structures from a functional perspective. *English for Specific Purposes, 23*(3), 264-279. doi:10.1016/S0889-4906(03)00005-X

Appendix A: Example of Color-Coded Abstract

Appendix B: Evaluation Checklist

Assignment 1 (Abstract Writing) Evaluation

Appendix C: Student Essay with Feedback

The Reality of Peer, Self- and Teacher Assessments in an EFL Classroom

Sumie Matsuno*

*Aichi Sangyo University College
smatsuno@asu.ac.jp

Abstract

In a required English class, 38 university students gave a presentation and assessed not only their peers' presentations but also their own presentations. Furthermore, their teacher assessed their presentations. FACETS analysis was conducted in order to compare these peer, self- and teacher assessments. The results indicated that the correlation coefficients between the peer and teacher assessments were not as low as those between the peer and self-assessments. On the other hand, the teacher-assessor was the most peculiar. Moreover, the peer- and teacher-assessors were quite lenient, while the self-assessors were more severe than the other two. Regarding item difficulties, the teacher rated posture & eye contact strictly; meanwhile, peer- and self-assessors rated fluency & pronunciation strictly. As for bias analysis, 35 bias interactions between the peer-assessors and the presenters and no bias interactions between the teacher-assessor and the presenters were detected. Overall, this study gives some caution in utilizing peer and self-assessments as formal assessments to assess students' presentations in an EFL classroom.

Keywords: peer assessment, self-assessment, teacher assessment, presentation, Japanese university

1. Introduction

Peer and self-assessments are educational activities in which students assess their peers' or own work and have their own work assessed by their peers. Many researchers have proven that peer assessment is a good pedagogical instrument (Afitska, 2014; Ahmad & Majid, 2015; Bowman-Perrott, Mahadevan, & Etchells, 2016; Casallas & Samir, 2016; Cheng and Warren, 2005; Colby, 2010; Iraji, Enayat, & Momeni, 2016; Jingjing & Fengying, 2015; Sin, 2013), as is self-assessment (Andrade & Valtcheva, 2009; Babaii et al., 2016; Babaii, Taghaddomi & Pashmforoosh, 2016; Butler & Lee, 2010; Dann, 2002; De Saint-Leger, 2009; Little, 2005; Little, 2009). On the other hand, some researchers have claimed that peer and self-assessments cannot be considered an effective measurement tool (Alderson, 2005; Brantmeier & Vanderplank, 2008; Brantmeier, Vandeerplank, & Strube, 2012; Butler & Lee, 2006; Chen, 2008; Cheng and Warren, 1999; Dolosic, Brantmeier, Strube, and Hogrebe, 2016; Farrokhi, Esfandiari, and Schaefer, 2012; Mazloomi & Khabiri, 2016), while other studies have shown that peer and self-assessments could be good measurement tools (Azarnoosh, 2013; Birjandi and Bolghari, 2015; Jones and Alcock, 2014; Patri, 2002; Saito, 2008; Saito & Fujita, 2004; Matsuno, 2009; Topping, 2009).

In the present study, students assessed their peers' presentations, which may have been difficult for them because they had to assess their peers' performance within tight time constraints. Assessing presentations is totally different from assessing writing, wherein they can take more time. Using FACETS analysis and, partly, Excel, the following research questions are answered in the context of students assessing peers' presentations.

1. How could peer, self- and teacher assessments be measurable using FACETS analysis?
2. How much were peer, self- and teacher assessments correlated?
3. How were peer, self- and teacher assessments similar and different based on FACETS analysis?
4. Did peer- and teacher-raters have any biases toward presenters and items?

2. Method

Thirty-eight second-year national university students participated in this study. Their TOEFL scores were between 460 and 540. When they were freshmen, they took two English writing classes. Then, they took this required English class, wherein they learned how to give a 3-minute presentation in English. During the presentation period, they conducted peer and self-assessments. This presentation was conducted for 2 weeks, and two students (R1 and R8) did not attend either class; hence, they were not included as peer-raters. The assessment sheet included six items: posture & eye contact; voice; visuals & gestures; fluency & pronunciation; content & organization; and overall, which were scored on a five-point Likert scale. Moreover, their presentations were assessed by their teacher, who is Japanese and has been teaching this presentation class for more than 10 years.

FACETS analysis was conducted using the FACETS computer program, version 3.82.3 (Linacre, 2019). This analysis investigated the impact of rater severity, presenter ability, and item difficulty. In this study, FACETS was run four times: peer assessments only, self-assessments only, teacher assessments only, and peer and teacher assessments together. Self-assessors were assumed to have the same mean level of severity, and group anchoring was utilized in order to see the differences in presenters' abilities between peer, self- and teacher assessments.

3. Results

3.1 Research question 1

The first research question — "How could peer, self- and teacher assessments be measurable using FACETS analysis?" — was answered using Summary statistics in FACETS.

Table 1. Summary statistics of peer, self-, and teacher-assessments

	Peer assessment			Self-assessment		Teacher assessment	
	Presenters	Raters	Items	Presenters	Items	Presenters	Items
Logit-Scale							
Infit MSE							
M	1.00	1.02	1.00	0.91	0.89	1.00	0.92
SD	0.25	0.36	0.20	0.76	0.38	0.69	0.49
Std. Infit MSE							
M	-0.10	-0.20	-0.30	-0.30	-0.60	0.00	-0.50
SD	1.70	2.60	4.00	1.30	1.80	1.00	2.50
Outfit MSE							
M	0.99	1.01	1.00	0.89	0.86	0.99	0.99
SD	0.25	0.33	0.20	0.75	0.37	0.75	0.49
Std. Outfit MSE							
M	0.99	-0.20	-0.30	-0.40	-0.70	0.10	-0.10
SD	0.25	2.60	0.39	1.30	1.70	1.00	2.30
Reliability of Separation	0.97	0.98	0.97	0.91	0.85	0.68	0.95
Chi-Square	1286.30	2124.20	147.60	293.70	40.80	105.80	94.40

Based on Linacre (2012), values from 0.5 to 1.5 (the logit scale) of infit and outfit mean-square statistics are considered "productive for measurement" (Linacre, 2012, p.15). Moreover, the acceptable range of standardized-fit statistics is between -1.9 and 1.9, which indicates that "data have reasonable predictability" (Linacre, 2002, p. 878). In this study, all of the infit and outfit mean-square statistics and the standardized infit and outfit mean-square values were within the acceptable range; that is to say, the data in this study could be utilized for measurement.

The separation reliability statistic indicates how well individual elements within a facet can be differentiated from one another. In this study, high separation reliability statistics for presenters, raters and items were detected, meaning that each writer's ability, each rater's severity and each item's difficulty were different. In addition, a chi-square statistic determines whether or not the element within a facet can be exchangeable. In this study, each chi-square statistic indicated that the overall differences between elements within the writer, rater and item facets were significant and could not be exchanged.

3.2 Research question 2

Research question 2 — "How much were peer, self- and teacher assessments correlated?" — was answered using Correlation Statistics.

Table 2. Correlation coefficients between peer and teacher-assessments, and peer and self-assessments

Presenters	P x T	S x P	Presenters	P x T	S x P	Presenters	P x T	S x P
P1	0.55*	0.86*	P14	0.56*	0.32	P27	0.57*	0.84*
P2	0.68*	-0.27	P15	-0.22	-0.22	P28	0.09	0.24
P3	0.87*	0.54*	P16	-0.15	-0.91	P29	0.71*	-0.05
P4	0.67*	0.83*	P17	0.63*		P30	0.13	0.18
P5	-0.22	0.07	P18	0.70*		P31	-0.42	0.86*
P6	0.83*	0.47*	P19	0.87*	-0.54	P32	0.61*	0.54*
P7	0.53*	-0.20	P20	0.78*	0.72*	P33	0.70*	0.49*
P8	0.49*	0.16	P21	0.44*	0.85*	P34	0.11	0.47*
P9	0.73*	0.74*	P22	-0.08	-0.69	P35	0.73*	0.45*
P10	0.75*	-0.33	P23	0.76*	0.34	P36	0.29	0.23
P11	0.29	0.37	P24	0.91*		P37	0.54*	0.17
P12	-0.41	0.16	P25	0.85*	-0.07	P38	0.84*	0.69*
P13	0.90*		P26	0.20	-0.25			

Notes: P = peer assessments, T=teacher assessments, S= self-assessments

Some students did not assess their own performance; therefore, 38 peer assessments, 34 self-assessments and 38 teacher assessments were included. In Table 2, correlation coefficients above 0.40 were considered reasonable. Twenty-six correlations between peer and teacher assessments and 20 correlations between peer and self-assessments were found to be reasonable correlations. Presumably, the correlation coefficients between the peer and teacher assessments were not as low as those between the peer and self-assessments; however, considering four missing self-assessments, both correlations between P and T and those between P and S did not show good correlations. In addition, the point biserial correlation of raters in FACETS was 0.36, which also indicated that the raters were not sufficiently correlated with one another.

3.3 Research question 3

Research question 3 — "How were peer, self- and teacher assessments similar and different based on FACETS analysis?" — was answered using FACETS maps, rater measurement reports, and unexpected responses.

FACETS maps

FACETS maps display visual information on differences that might exist among different elements of a facet, such as differences in ability among writers. The scale in the first column is the logit scale; meanwhile, the one in the far-right column is the scale used in the scoring. The second column shows the variation in ability among presenters. The presenters are organized with the most able at the top and the least able at the bottom. The third column shows the variation in severity among raters. The most severe rater is at the top and the least severe rater is at the bottom. The fourth column shows the variation in difficulty among items. The most severely scored item is at the top and the least severely scored item is at the bottom. Figures 1, 2 and 3 show peer, self- and teacher assessments, respectively.

```
+----------------------------------------------------------------+
|Measr|+presenters|-raters |-items                        |Scale|
|-----+-----------+--------+----------------------------+--------|
|  4 +           +        +                             + (5) |
|     |          |        |                             |      |
|     | *        |        |                             |      |
|     | **       | *      |                             |      |
|  3 + **        +        +                             +      |
|     | *        |        |                             |      |
|     | *        |        |                             |      |
|     | ****     | *      |                             |  4   |
|     | *        | *      |                             |      |
|  2 + **        +        +                             +      |
|     | ***      | *      |                             |      |
|     | ***      |        |                             |      |
|     | ***      |        |                             |      |
|     | ***      |        |                             |      |
|     | ***      | *      |                             | ---  |
|     | ***      |        |                             |      |
|  1 + ***       + **     +                             +      |
|     | ***      | ***    |                             |      |
|     | *        |        |                             |      |
|     |          | ***    |                             |      |
|     |          | *      | Fluency                     |      |
|     |          | *      | Posture   Visuals           |      |
|  * 0 *         * ***    * Overall          Voice      * 3 * |
|     | *        | ***    |                             |      |
|     |          | *      |                             |      |
|     |          | ****** |                             |      |
|     |          | *      | Content                     |      |
| -1 +           + **     +                             +      |
|     | *        | *      |                             | ---  |
|     |          |        |                             |      |
|     |          |        |                             |      |
| -2 +           +        +                             +      |
|     |          | *      |                             |      |
|     |          | *      |                             |      |
|     |          | *      |                             |  2   |
| -3 +           +        +                             +      |
|     |          | *      |                             |      |
|     |          |        |                             |      |
|     |          |        |                             |      |
| -4 +           +        +                             + (1) |
|-----+-----------+--------+----------------------------+--------|
|Measr| * = 1     | * = 1  |-items                       |Scale|
+----------------------------------------------------------------+
```

Figure 1. FACETS map of peer assessments.

```
+------------------------------------------------------+
|Measr|+Presenters|-R2        |-items      |Scale|
|-----+-----------+-----------+-----------+--------|
|  8 +           +           +           + (5) |
|     |          |           |           |      |
|     | *        |           |           |      |
|  7 +           +           +           +      |
|     |          |           |           |      |
|     |          |           |           | ---  |
|  6 +           +           +           +      |
|     |          |           |           |      |
|  5 +           +           +           +      |
|     |          |           |           |      |
|  4 + *         +           +           +  4   |
|     |          | .         |           |      |
|  3 +           +           +           +      |
|     |          |           |           |      |
|     | *        |           |           |      |
|  2 +           +           +           +      |
|     |          |           |           | ---  |
|     |          |           | fluency   |      |
|  1 + **        +           +           +      |
|     | *        |           | posture   |      |
|     | ****     | .         | overall   |      |
|  * 0 *         * ********* . *          *      * |
|     | ****     | .         | content   |  3   |
|     |          |           | visuals   |      |
| -1 + **        +           +           +      |
|     | ******   |           | voice     |      |
| -2 +           +           +           +      |
|     | *        |           |           | ---  |
|     | ***      |           |           |      |
| -3 +           +           +           +      |
|     | **       |           |           |      |
|     |          | .         |           |      |
| -4 + *         +           +           +  2   |
|     |          |           |           |      |
| -5 +           +           +           +      |
|     |          |           |           |      |
| -6 + *         +           +           + ---  |
|     | *        |           |           |      |
| -7 +           +           +           + (1) |
|-----+-----------+-----------+-----------+--------|
|Measr| * = 1     | * = 3     |-items      |Scale|
+------------------------------------------------------+
```

Figure 2. FACETS map of self-assessments.

therefore, all raters were not equally severe.

Table 3. Rater measurement report

	Severity	Error	Infit		Severity	Error	Infit
R2	0.91	0.15	0.91	**R22**	-0.36	0.15	1.56
R3	0.46	0.15	0.91	**R23**	2.23	0.15	1.35
R4	-2.20	0.18	1.34	**R24**	-0.32	0.10	1.24
R5	0.41	0.15	1.46	**R25**	0.15	0.15	0.95
R6	-0.15	0.15	0.65	**R26**	-0.76	0.16	0.96
R7	1.00	0.15	1.15	**R27**	1.76	0.15	1.06
R9	-1.10	0.16	1.11	**R28**	-1.98	0.17	0.71
R10	0.00	0.15	0.64	**R29**	1.02	0.15	1.02
R11	-0.26	0.15	0.71	**R30**	0.89	0.15	0.87
R12	-0.97	0.16	0.57	**R31**	0.14	0.15	1.11
R13	-0.57	0.15	0.52	**R32**	0.42	0.17	1.37
R14	2.25	0.15	0.81	**R33**	-2.18	0.28	1.80
R15	-0.58	0.15	0.84	**R34**	3.32	0.15	0.81
R16	0.76	0.15	0.76	**R35**	-0.36	0.15	1.65
R17	-0.16	0.15	0.78	**R36**	-2.91	0.21	1.08
R18	-0.63	0.18	0.35	**R37**	-0.43	0.14	0.51
R19	0.14	0.15	0.76	**R38**	-0.24	0.15	0.73
R20/Teacher	-1.26	0.11	1.93	**R39**	0.64	0.15	1.00
R21	0.91	0.15	0.71				

Notes: Reliability of separation index = 0.97; fixed (all same) chi-square: 1295.7, df: 37, significance < 0.00

Linacre (2012) mentions that an infit mean-square value above 2.00 "distorts or degrades the measurement system" (p.15). In this table, fit values of all raters were below 2.00; all raters were fairly self-consistent in scoring. However, the infit mean square of R20, i.e., the teacher, was high (1.93), which indicated that the teacher was the most peculiar in her rating manners compared with peer raters.

Unexpected responses
    Table 4 presents unexpected responses. FACETS identified 23 ratings as being unexpected.

Table 4. Unexpected responses

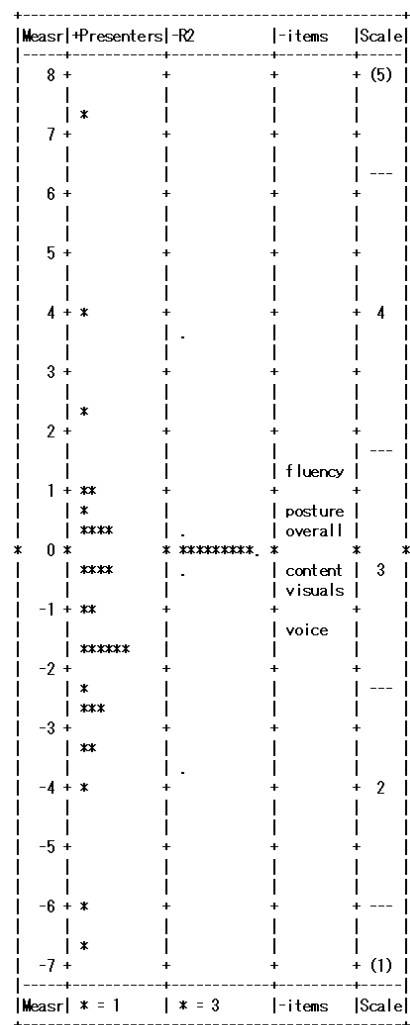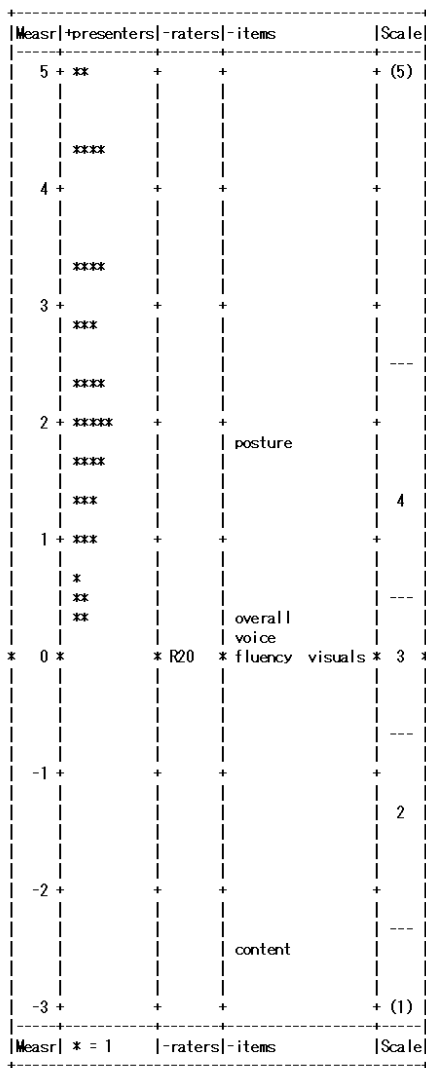| Observed Score	Expected Score	Residual	Standardized residual	Rater	Presenter	Item
2	4.6	-2.6	-4.7	R20	P8	Posture & eye
4	5.0	-1.0	-4.7	R36	P26	Content & Org
1	3.9	-2.9	-4.5	R20	P9	Posture & eye
1	3.7	-2.7	-4.2	R20	P34	Posture & eye
1	3.7	-2.7	-4.1	R20	P13	Posture & eye
2	4.3	-2.3	-3.9	R20	P1	Posture & eye
2	4.4	-2.4	-3.9	R20	P14	Posture & eye
1	3.3	-2.3	-3.6	R4	P3	Fluency
2	4.2	-2.2	-3.5	R24	P28	Voice
1	3.2	-2.2	-3.4	R14	P8	Posture & eye
2	4.1	-2.1	-3.4	R20	P17	Posture & eye
2	4.1	-2.1	-3.3	R20	P11	Posture & eye
2	4.1	-2.1	-3.3	R20	P24	Posture & eye
1	3.1	-2.1	-3.3	R30	P21	Voice
2	4.1	-2.1	-3.2	R20	P33	Posture & eye
5	2.9	2.1	3.2	R23	P36	Voice
1	3.0	-2.0	-3.1	R14	P5	Visuals
2	4.0	-2.0	-3.1	R20	P18	Posture & eye
3	4.6	-1.6	-3.1	R36	P21	Voice
5	3.1	1.9	3.0	R5	P19	Posture & eye
1	3.0	-2.0	-3.0	R22	P25	Fluency
5	3.1	1.9	3.0	R35	P25	Visuals
1	3.0	-2.0	-3.0	R35	P25	Fluency

Figure 3. FACETS map of teacher assessments.

Examining presenters' abilities in the maps, self-raters were more severe than teacher- and peer-raters because presenters' ability logits were often below 0.00 logit in the self-assessment, while those in peer and teacher assessments were often above 0.00 logit. Moreover, the teacher evaluated "posture & eye contact" the most severely, while peer- and self-raters assessed "fluency & pronunciation" the most difficult. The teacher and peer-assessors assessed "content" the most leniently; meanwhile, self-assessors assessed "voice" the most leniently.

Rater measurement report
    The following is the rater measurement reported by FACETS analysis, including both peer and teacher assessments. From the left, each column shows rater IDs, rater severity, error, and fit mean square. The most lenient rater was R36, whose severity was -2.91; meanwhile, the most severe rater was R34, whose severity was 3.32. The reliability of separation was 0.97, which was very high. Furthermore, the chi-square of 1295.7 with 37 df was significant at p<0.00;

The first column indicates the observed scores. The second column shows the expected scores based on the presenter and item estimated measure, and the third column, i.e., residual, indicates the difference between the observed and expected values. The fourth column presents the standardized residuals, which normally should be between -3.0 and 3.0. The fifth column shows the rater IDs, the sixth column shows the presenter IDs, and the last column shows the items. In this table, 11 out of 23 unexpected responses were those of the teacher. All of the teacher's unexpected responses were concerned with posture & eye contact. She probably assessed this item differently from her students.

3.4 Research question 4

Research question 4 was "Did peer- and teacher-raters have any biases toward presenters and items?" Table 5 shows the results of the bias analysis in terms of the interaction between rater severity and presenter ability. Each table lists only the rater–candidate interactions that displayed a significant bias, which means only those interactions in which the probability is below 0.05. There were a total of 35 significantly biased interactions. The first two columns show rater IDs (column 1) and presenter IDs (column 2). The next two columns show the total observed score of five items (column 4) and the total expected score of five items (column 5). Since each item had a range of 1 to 5, the total observed or expected score for the five items combined falls in the range of 6 to 30. For example, the total observed and expected scores by Rater 9 for Presenter 17 were 18 and 25.14 points, respectively, out of 30 possible points. This information suggests that the total score that the presenter received from R9 was lower than might have been expected. In other words, R9 scored P17 more harshly than expected. Columns 6 and 7 show a bias logit which represents the degree of difference between the observed scores and the expected scores (column 6) and the likely error of the bias estimate (column 7). In column 7, the bias estimates in column 5 are converted into z-scores. In this table, 20 raters scored the presenters more harshly compared to the way in which a particular rater scored the other presenters, since their z-scores were negative. Fifteen raters consistently scored the candidate more leniently than other presenters, since their z-scores were positive.

Table 5.   Bias calibration report: rater-presenter interaction

Rater	Presenter	Observed score	Expected score	Bias (logit)	Error	z-score	Probability
R9	P17	18	25.14	-2.86	0.63	-4.53	0.006
R4	P3	14	20.80	-2.71	0.64	-4.23	0.008
R33	P24	21	27.17	-4.22	0.63	0.00	0.008
R34	P28	10	15.75	-2.46	0.69	-3.55	0.016
R22	P31	15	20.92	-2.36	0.64	-3.71	0.013
R29	P25	10	15.16	-2.22	0.69	-3.20	0.023
R12	P4	18	23.28	-2.09	0.63	-3.30	0.021
R9	P16	20	25.04	-2.03	0.63	-3.23	0.023
R14	P8	15	19.93	-1.97	0.64	-3.10	0.026
R24	P28	20	24.86	-1.95	0.63	-3.10	0.026
R22	P34	16	20.87	-1.93	0.63	-3.05	0.028
R13	P3	12	16.67	-1.92	0.66	-2.92	0.033
R10	P7	20	24.68	-1.87	0.63	-2.98	0.030
R5	P3	10	14.25	-1.85	0.69	-2.67	0.044
R5	P18	16	20.50	-1.79	0.63	-2.83	0.036
R21	P27	16	20.22	-1.68	0.63	-2.65	0.045
R5	P2	16	29.21	-1.68	0.63	-2.65	0.045
R7	P7	18	22.23	-1.67	0.63	-2.64	0.046
R3	P16	17	21.19	-1.66	0.63	-2.62	0.046
R25	P29	21	25.00	-1.61	0.63	-2.58	0.049
R23	P30	20	15.95	1.61	0.63	2.57	0.049
R21	P21	23	18.64	1.72	0.63	2.72	0.041
R24	P32	26	21.77	1.76	0.68	2.58	0.049
R2	P18	24	19.22	1.89	0.64	2.95	0.031
R17	P3	21	15.66	2.12	0.63	3.39	0.019
R22	P27	28	23.43	2.18	0.82	2.65	0.045
R7	P8	28	23.10	2.32	0.82	2.81	0.037
R5	P10	28	22.74	2.46	0.82	2.99	0.030
R12	P3	24	17.69	2,51	0.64	3.90	0.011
R25	P27	28	22.15	2.69	0.82	3.27	0.022
R32	P27	28	21.47	2.96	0.82	3.60	0.015
R5	P5	29	23.10	3.18	1.08	2.95	0.031
R32	P28	29	23.08	3.18	1.08	2.96	0.031
R7	P14	29	21.49	3.81	1.08	3.54	0.016
R33	P25	30	23.24	4.00	1.54	2.60	0.048

As observed, 35 bias interactions between the peer-assessors and the presenters and no bias interactions between the teacher-assessor and the presenters were detected.

Table 6 shows the results of the bias analysis in terms of the interaction between rater severity and item difficulty. Each table lists only the rater–candidate interactions that displayed a significant bias, which means only those interactions in which the probability is below 0.05.

Table 6. Bias calibration report: rater-item interaction

Rater	Items	Observed score	Expected score	Bias (logit)	Error	z-score	Probability
R20	Posture	125	154.12	-1.87	0.25	-7.47	0.000
R26	Visuals	63	71.51	-1.16	0.37	-3.18	0.055
R31	Voice	57	65.63	-1.15	0.36	-3.15	0.059
R35	Fluency	59	67.14	-1.09	0.36	-2.98	0.084
R26	Content	67	72.81	-0.92	0.39	-2.38	0.029
R19	Fluency	52	58.10	-0.85	0.38	-2.28	0.036
R2	Posture	50	56.39	-0.85	0.37	-2.32	0.032
R22	Voice	63	69.30	-0.85	0.36	-2.33	0.032
R27	Voice	47	53.27	-0.84	0.37	-2.28	0.035
R17	Voice	61	67.05	-0.80	0.36	-2.21	0.041
R16	Content	59	64.93	-0.79	0.36	-2.16	0.045
R22	Visuals	74	68.14	0.83	0.38	0.83	0.046
R19	Posture	64	58.04	0.84	0.38	0.84	0.041
R26	Posture	76	69.83	0.90	0.40	0.90	0.035
R35	Visuals	77	68.82	1.20	0.40	1.20	0.082
R31	Content	81	70.23	1.69	0.43	1.69	0.001

Sixteen biases were detected, among which R20, i.e., the teacher, showed bias toward posture & eye contact. She

assessed this item more severely than expected.

4. Discussion

The first research question was "How could peer, self- and teacher assessments be measurable using FACETS analysis?" Summary statistics show that although all of the data were effective for measurement, the raters were significantly different in their severity, which was not ideal to be used as a formal assessment.

The second research question was "How much were peer, self- and teacher assessments correlated?" The correlation coefficients between the peer and teacher assessments were not as low as those between the peer and self-assessments. However, both correlations were not very strong, whereby indicating that peer assessments and self-assessments may be difficult to utilize as a formal assessment in class when assessing students' presentations.

The third research question was "How were peer, self- and teacher assessments similar and different based on FACETS analysis?" FACETS maps show that self-raters were the most severe raters, while both peer- and teacher-raters were lenient. On the other hand, peer- and self-raters rated "fluency & pronunciation" as the most difficult item, while the teacher-rater rated "posture & eye contact" as the most difficult. It indicates that the students and teacher assessed differently. The rater measurement report shows that the teacher was the most peculiar among teacher and peer assessments. Moreover, unexpected responses indicated that 11 out of 23 were produced by the teacher. All of hers were concerned with "posture & eye contact," showing that her rating manner was different from that of her students. Considering the teacher assessment to be a formal assessment, since big differences were detected, it might be difficult to use peer assessments as a formal assessment. Moreover, self-assessment was different from the peer and teacher assessments. Hence, it is also difficult to include as a formal assessment.

The fourth research question was "Do peer- and teacher-raters have any biases toward presenters and items?" Bias analysis showed 35 bias interactions between the peer-assessors and the presenters and no bias interactions between the teacher-assessor and the presenters. Although the teacher was peculiar among them, she did not have any biases toward the presenters. On the other hand, 35 bias interactions were produced by peer-assessors. Regarding bias interactions between raters and items, one teacher bias, i.e., "posture & eye contact," and 15 peer-raters' biases were detected. Again, this showed that the teacher assessed "posture & eye contact" differently from her peers. What is more, bias analysis concluded that using peer assessments as a formal assessment may cause some worry.

Overall, this study indicates that self- and peer assessments may be inappropriate to be used as formal assessments when students assess their peers' or own presentations. However, as other researchers demonstrate, those assessments can be good pedagogical instruments to improve their own performance, enhance autonomy, and assess their own skills. (Afitska, 2014; Ahmad & Majid, 2015; Bowman-Perrott, Mahadevan, & Etchells, 2016; Casallas & Samir, 2016; Cheng and Warren,

2005; Colby, 2010; Iraji, Enayat, & Momeni, 2016; Jingjing & Fengying, 2015; Sin, 2013). Therefore, instead of using peer and self-assessments as formal assessments, they should be utilized only for educational purposes.

5. Limitations of the present study

Some limitations require consideration concerning this study. Since only 38 students participated in this study, more participants are needed in order to generalize this result. Moreover, only one teacher participated in this study and she became peculiar among peers. If more teachers had participated, the results would have been different.

Despite having some limitations, it showed the reality of peer, self- and teacher assessments in an EFL presentation class. Peer and self-assessments seemed to be different from the teacher assessment. In this case, it may be difficult to use them as one of the formal assessments. In particular, assessing peers' presentations may be difficult because they have to assess quickly, and may not understand the presentations well enough to evaluate, but cannot watch them again. This is totally different from evaluating writing, wherein they can take more time. Hence, this study concludes that peer and self-assessments cannot be reliable in the context of evaluating presentations in an EFL classroom.

References

Afitska, O. (2014). Use of formative assessment, self and peer-assessment in the classrooms: some insights from recent language testing and assessment research. *i-Manager's Journal on English Language Teaching, 4*(1), 29–39. doi:10.26634/jelt.4.1.2640

Ahmad, Y., & Majid, M. (2015). Portfolio assessment, peer assessment and writing skill improvement. *Theory and Practice in Language Studies, 5*(12), 2504–2511. doi:10.17507/tpls.0512.10

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice, 48*(1), 12–19. doi:10.1080/00405840802577544

Azarnoosh, M. (2013). Peer assessment in an EFL context: attitudes and friendship bias. *Language Testing in Asia, 3*(11), 1–10. doi:10.1186/2229-0443-3-11

Babaii, E., Taghaddomi, S., & Pashmforoosh, R. (2016). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing, 33*(3), 411–437. doi:10.1186/2229-0443-3-11

Birjandi, P., & Bolghari, M. S. (2015). The relationship between the accuracy of self- and peer-assessment of Iranian intermediate EFL learners and their learning styles. *Theory and Practice in Language Studies, 5*(5), 996–1006. doi:10.17507/tpls.0505.15

Bowman-Perrott, L., Mahadevan, L., & Etchells, M. (2016). Assessing the academic, social, and language production outcomes of English language learners engaged in peer tutoring: A systematic review. *Education and Treatment of Children, 39*(3), 359–388. doi:10.1353/etc.2016.0016

Butler, Y., & Lee, J. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal, 90*, 506–518. doi:10.1111/j.1540-4781.2006.00463.x

Casallas, U. D. F. & Samir, F. P. C. (2016). Argumentation skills: a peer assessment approach to discussion in the EFL classroom. *PROFILE Issues in Teachers' Professional Development*, *18*(2), 111–123. doi:10.15446/profile.v18n2.53314

Cheng, Y. M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research, 12*, 235–262. doi:10.1177/1362168807086293

Cheng, W., & Warren, M. (1999). Peer and teacher assessment of the oral and written tasks of a group project. *Assessment & Evaluation in Higher Education, 24,* 301–314. doi:10.1080/0260293990240304

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, *22*(1), 93–121. doi:10.1191%2F0265532205lt298oa

Colby, D. C. (2011). Using "Assessment for Learning" practices with pre-university level ESL students: A mixed methods study of teacher and student performance and beliefs. ProQuest Dissertations Publishing.

De Saint-Leger, D. (2009). Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals, 42*, 158–178. doi:10.1111/j.1944-9720.2009.01013.x

Dann, R. (2002). *Promoting assessment as learning: Improving the learning process.* London: Routledge.

Dolosic, H. N., Brantmeier, C., Strube, M., & Hogrebe, M. C. (2016). Living language: Self-assessment, oral production, and domestic immersion. *Foreign Language Annals; Alexandr*ia, *49*(2), 302–316.

Engelhard, G., & Winds, S. A. (2016). Exploring the effects of rater linking designs and rater fit on achievement estimates within the context of music performance assessment. *Educational Assessme*nt, *21*(4), 278–299. doi:10.1080/10627197.2016.1236676

Farrokhi, F., Esfandiari, R., & Schaefer, E. (2012). A many-facet rasch measurement of differential raters severity/leniency in three types of assessment. *JALT Journal, 34*(1), 79–102.

Harrington, D., & LeBeau, C. (2009). *Speaking of speech. Macmillan Language House*: Tokyo.

Iraji, H. R., Enayat, M. J., & Momeni, M. (2016). The effects of self- and peer-assessment on Iranian EFL learners' argumentative writing performance. *Theory and Practice in Language Studies, 6(*4), 716–722. doi:10.17507/tpls.0604.08

Jafarpur, A. (1991). Can naïve EFL learners estimate their own proficiency? *Evaluation and Research in Educatio*n, 5, 145–157.

Jingjing, X., & Fengying, M. (2015). A study into students' views on guided peer feedback in group work. *Chinese Journal of Applied Linguistics*, *38*(2), 188–203. doi:10.1515/cjal-2015-0011

Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education, 39*(10), 1774–1787. doi:10.1080/03075079.2013.821974

Linacre, J. M. (2012). Many-facet Rasch measurement: Facets tutorial. Retrieved June 6th, from http://winsteps.com/tutorials.htm.

Linacre, J. M. (2017). FACETS: computer program for many faceted Rasch measurement (version 3.80.0). Chicago, IL: Mesa Press.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, *16*(2), 878.

Little, D. (2009). Language learner autonomy and the European language portfolio: two L2 examples. *Language Teaching: Surveys and Studies, 42,* 222–233. doi:10.1017/S0261444808005636

Mazloomi, S., & Khabiri, M. (2016). The impact of self-assessment on language learners' writing skill. *Innovations in Education and Teaching International*, *55*(1), 91–100. doi:10.1080/14703297.2016.1214078

Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, *25*(4), 553–581.

Saito, H., & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching, 8*, 31–54. doi:10.1191%2F1362168804lr133oa

Shin, R. (2013). Effect of using Facebook to assist English for business communication course instruction. *The Turkish Online Journal of Educational Technology*, *12*(1), 52–59.

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, *26*(1), 75–100. doi:10.1177%2F0265532208097337

Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, *19*, 109–131. doi:10.1191%2F0265532202lt224oa

Topping, K. J. (2009). Peer assessment. *Theory in Practice*, *48*(1), 20–17. doi:10.1080/00405840802577569

Japanese English Learners' Responses to English Debate: Insights from a Pilot Class

Masakazu Mishima* and Yuka Yamamoto**

*Rikkyo University
mmishima@rikkyo.ac.jp
**Rikkyo University
yukayamamoto@rikkyo.ac.jp

Abstract

This study presents the results of course evaluation survey administered to university students in a pilot debate class. The results suggest that student participants were highly content with the pilot course, and learning to debate in English enhances their ability to think logically and communicate effectively. While a debate consists of a series of conventionalized communicative situations, participants reported rebuttal is the most challenging aspect of debate as it requires them to not only to identify weaknesses of the opponents' claims, but also spontaneously deliver counterarguments. In terms of debate preparation, participants highly valued recurrent opportunity to work in groups in and out of the classroom. They also mentioned several strategies to facilitate their debate preparation processes. The paper presents select examples of such strategies and discuss the findings from a perspective of socially shared regulation of learning. The study highlights socially and materially situated regulatory strategies used by participants to achieve a shared goal.

Keywords: Teaching debate, curriculum development, critical thinking, argumentation, and socially shared regulation of learning

1. Introduction

The benefits of learning debate have been well-recognized in various educational contexts (Brown & Brown, 2014). In the face of changing landscape of English language education in Japan, learning to debate in English encapsulates many important qualities required to become a competent language user. In particular, advocates of debate education claims that learning to debate facilitates critical thinking, argumentation, communication, and research skills of the learner (e.g., Kennedy, 2007).

The development of the aforementioned skills may have been on the margins of traditional English education in Japanese higher education. However, the tide of the nation-wide English language curriculum reform initiated by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) places the previously compartmentalized skills in the central part of English language education (MEXT, 2011). Given the ongoing curriculum reform movement across the nation, the present research was initiated as part of new English language program development at Rikkyo University, a private university situated in Tokyo, Japan. The study examined students' evaluative responses to a pilot debate class, which is to be launched for all freshman year students in 2020. This paper, albeit briefly, presents select results of the study and discusses some important implications for curriculum developers and/or instructors who are interested or currently engaged in debate class development in Japanese higher education.

2. Research Context

The introduction of the new Course of Study Guidelines (see Tahira, 2012 for a brief historical overview) by MEXT revolves around the development of Japanese students' English ability to appropriately function in global society. The MEXT proposals, for instance, provided an overview of English language proficiency required in global society as follows:

> Foreign language proficiency required in global society can be defined as capability of smooth communication with people of different countries and cultures using languages as a tool. The capability of smooth communication implies, for example, confident and active attitude toward communication with people of different countries and cultures as well as accurate understanding of partner's thoughts and intentions based on his/her cultural and social background, logical and reasoned explanation of one's own views, and convincing partners in course of debates.
>
> (MEXT, 2011, p. 3)

In line with the MEXT's initiatives, there was a strong need to adapt our curriculum to prospective students' needs. The current English language curriculum at Rikkyo University was designed on the basis that students were mainly receiving receptive English instructions (i.e., reading and listening skills) during their high school years and lacking in productive skills (i.e., speaking and writing). However, due to the ministry's new curriculum guidelines, our future students

will be exposed to communicative language skills from elementary school and onward. Accordingly, students are expected to learn comparatively more sophisticated productive and cognitive skills in higher education.

In 2014, Rikkyo University was selected for the government funding, Top Global University Project (Type B: Global Traction Type) to play a leading role to facilitate the global competitiveness of higher education in Japan. As one of the university's reforms, there is a growing number of English-medium instruction (EMI) courses offered by the university.

In order to make a smoother transition to its EMI courses, one of our main missions is to develop a strong first year mandatory English courses and elective courses (i.e, Content and Language Integrated Learning (CLIL) programs) by 2024. These courses are expected to foster students' essential academic English language skills to study in EMI classes. Figure 1 shows the prospective model of classes that students are to take at different stages based on their English proficiency level.
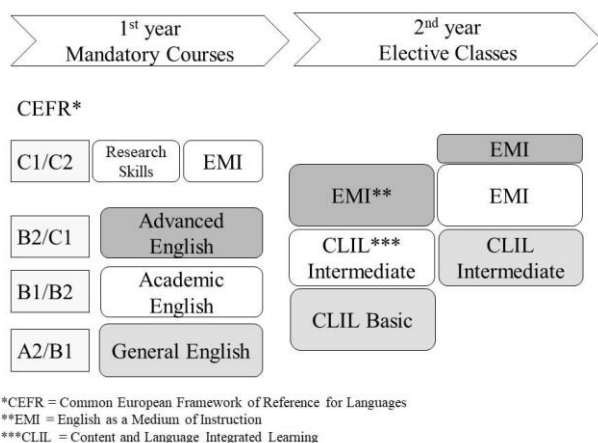

Figure 1. Ideal Progress and Stages of 2024 English Curriculum.

Currently, we are in the process of strengthening the first year mandatory courses listed in Figure 2. From 2020, the discussion class will be reduced to one semester and instead, all freshman year students will be taking a debate class to develop their critical thinking and research skills (Figure 2). Our two-semester discussion classes (i.e., English Discussion Class 1 & 2), have significantly contributed to the development of general communicative English ability of our students over the years. However, in order to implement the 2024 curriculum, students need to develop academic English language skills in terms of research, critical thinking, and speaking ability with rich academic content, which is, we believe, in good alignment with the MEXT's required English proficiency description presented earlier.
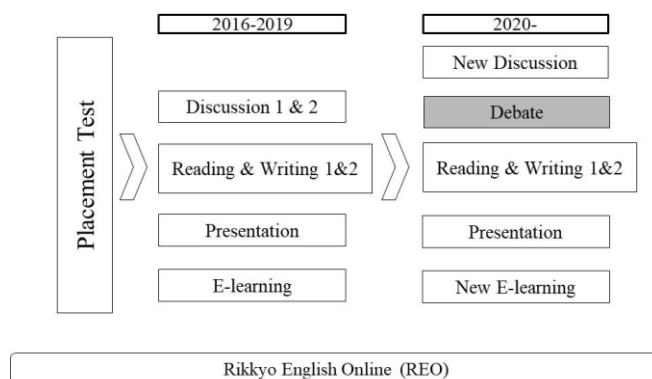

Figure 2. Changes to 2020 First-year English Mandatory Courses.

The present study examined the students' evaluations of a pilot debate class, which lasted once a week for 14 weeks. The aims of the course are to understand the nature and structure of debate in English, to develop critical thinking skills by analyzing and formulating arguments on issues from multiple perspectives, and to learn how to respond to questions through the development of research skills. In order to facilitate student learning, the instructor employed a number of instructional approaches such as active learning, collaborative learning, and student-centered.

Major activities in class were conducted in three formats: individual, pair, and group work. This was to practice and familiarize themselves with the debate structure, formulate and organize their arguments logically, search for information to support their arguments, and share and discuss relevant sources to be prepared for three official debates.

The official debates were assessed and performed in front of the class. To ensure the quality of the unified curriculum, a rubric was developed and used to assess learners' debate performance based on the quality of organization, arguments, evidence from different types of sources, rebuttal, and presentation style. The official debates followed a specific debate format as shown below in Table 1.

During the official debates, three matches/debates were conducted consisting of three or four members in each team. They debated on an assigned topic for approximately 30 minutes following the format.

Table 1. Debate Format

Affirmative Team (AT)	Negative Team (NT)
Affirmative Speech (3 min)	
	Question Time (2 min)
	Negative Speech (3 min)
Question Time (2 min)	
Preparation Time for Rebuttal (5 to 10 min)	
	Rebuttal (2 min)
Rebuttal (2 min)	
	Summary (2 min)
Summary (2 min)	

3. Methods

3.1 Data collection

Data collection took place in a one-semester 100-minute weekly elective debate class. The researchers visited a class session to explain the nature and procedures of the present study and asked students for their voluntary participation. Out of 25 students enrolled in the class, 21 students signed a consent form and formally participated in the study. Table 2 and Table 3 below summarize the basic demographic information of participant including participants' sex, university year, and majors.

Table 2. Sex and University Year of Participants

		University Year			
		Sophomore	Junior	Senior	
Sex	Male	3	5	1	
	Female	4	7	1	
Total		7	12	2	21

Table 3. Participants' Majors

Major	Frequency	Percent
Economics	1	5.0
Education	1	5.0
English Literature	4	20.0
Finance	1	5.0
French Literature	1	5.0
Intercultural Communication	2	10.0
International Business Law	4	20.0
Life Science	3	15.0
Modern Culture	1	5.0
Social Science	1	5.0
World History	1	5.0
No Response	1	5.0
Total	21	100.0

Aside from the demographic information, participants reported their TOEIC® IP examination score. The mean TOEIC® IP score was 705.40 (SD = 91.87) ranging from 495 to 890, which indicated that their overall English proficiency was intermediate to low-advanced level. All but one student had no experience taking a debate course.

On the last day of class, the researchers administered an online course evaluation survey to debate class students to elicit information on their overall response to the course, the challenging aspects of conducting a debate and how they resolved them, and the process of preparing for the class. The questions were given in Japanese and the students responded to them in Japanese. All 21 participants responded to the survey.

In addition to the course evaluation survey, the researchers collected various teaching materials (i.e., artifacts) used in the pilot class. The materials include, for instance, the course syllabus, worksheets, and debate structure sheet. These teaching materials were not formally analyzed but used to better inform the researchers of the instructional context to help analyze and interpret the course evaluation survey.

3.2 Data analysis

The survey (see Appendix A) contained four seven point Likert-scale statement questions. Participants responded to these questions to indicate their extent of agreement/disagreement. The questions were designed to elicit students' response to the following themes:
a. Course satisfaction
b. Confidence to conduct a debate
c. Understanding of debate structure
d. Importance of research before debating

Responses to the categories above were all statistically analyzed to examine their overall patterns. In the interest of space, we present the results of analysis on the two categories—course satisfaction (a) and importance of research before debating (d).

In addition to the Likert-scale questions above, seven text response questions were qualitatively analyzed via the open-coding method. Open-coding prototypically begins with listing any potentially representative categories to summarize qualitative data without any top-down prefabricated framework (see Yin, 2003). The researchers first individually coded participants' responses and then repeatedly compared the emerged codes and discussed the meanings of the codes to create core categories—repetitively emerged themes in the participants' responses. Upon reaching agreement, the emerged themes were summarized as follows:
1. Challenging aspects of conducting a debate
2. Useful resources to prepare for a debate

In the final stage of analysis for the two core categories, the researchers screened all participants' textual responses and selected exemplifiers linked to the core categories to present the results.

4. Results and Discussion

4.1 Course satisfaction

Figure 3 shows the frequency distribution of participants' response to the statement, "I am satisfied with the debate course." 7 indicates the strongest agreement and 1 indicates the strongest disagreement. The overall distribution of course satisfaction indicates the vast majority of students were content with the debate course—19 out of 21 participants selected 5 or above, which accounts for approximately 90% of all participants' responses. From our separate textual response questions, participants noted, for example, the instructor's explanation on debate format was very clear and useful. They also liked group or pair preparation activity for a debate such as impromptu practice debate. Furthermore, the instructor frequently changed students' grouping arrangements. To which, participants made positive comments since they were able to work with new group members. Although we cannot readily conclude that those are the reasons for the high level of course satisfaction, they seem to be significant contributing factors.
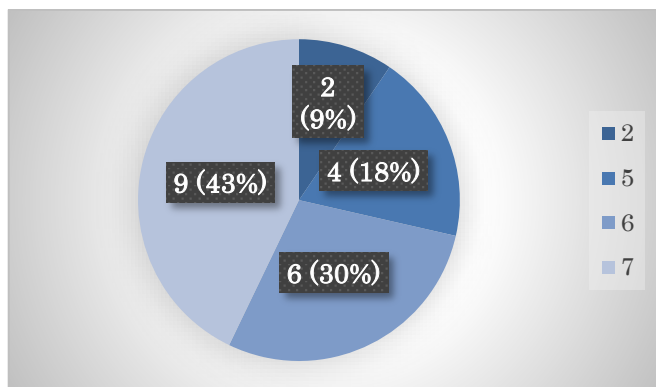
Figure 3. Pilot debate course student satisfaction.

4.2 Importance of research before debating

Figure 4 shows the frequency distribution of participants' response to the statement, "It is important to search and study topic-related reading materials before conducting a debate." 18 out of 21 participants indicated 5 or above to the statement. It indicates that 95% of participants perceived that searching and studying written works relevant to a given debate topic was important. The pilot debate class aimed to help students not only to develop their speaking skills in the specific communicative situation—debate, but also to develop research skills and critical thinking ability in the process of preparing for a debate. Based on the result, for the vast majority of participants, we believe that learning to debate was conducive to cultivating their understanding of the importance of research and critical thinking skills in debate as intended.
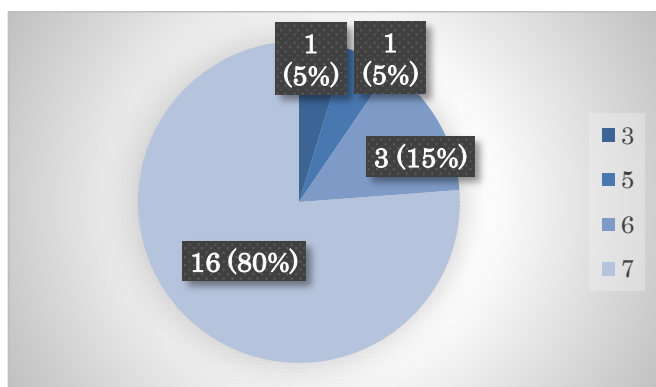

Figure 4. Importance of research before debating.

4.3 Debate: Challenging aspects

One challenging aspect of conducting a debate was connected to a specific aspect of debate—*rebuttal*. Rebuttal is a phase of debate during which debaters must offer counterarguments to the opponents' team. See the excerpt below:

Excerpt 1
To think up a plan to rebut. Effective rebuttal requires a clear understanding of opponents' claims and data presented. Along with that, I needed to come up with a

way to put my ideas into words [within a limited time]. That was hard.

The rebuttal phase is almost always present regardless of debate format adopted (e.g., Lincoln-Douglas) (Phillips, Hicks, & Springer, 2005). In a rebuttal phase, debaters are required to deliver counterarguments to the opponent team by identifying weaknesses of their arguments and/or evidence presented. Compared to other phases of debate (e.g., affirmative speech), the rebuttal phase is highly interactive as participants have to carefully listen to their opponents' claims, assess their arguments and evidence, and identify potential weaknesses. Furthermore, they are required to challenge their opponents' ideas on the spot. Although some debate formats such as policy debate include a cross-examination period during which debaters can discuss and coordinate how to rebut (Phillips, et al., 2001), this was not the case for students in the pilot debate class—they had 5 to 10 minutes to prepare.

Debate in a categorical sense may seem to be one coherent seamless speech event. However, it is comprised of various speech events, for each of which debaters must follow well-defined rules and perform their respective roles. Many of these roles can be well performed after careful planning and preparation as they are premeditated speech events. Rebuttal in the case of the present study, however, was a highly spontaneous speech event, for which debaters have to deliver counterarguments with little preparation. From a perspective of L2 speaking skills, we view that the perceived difficulty reported by participants was associated with the construct of L2 speaking fluency. Although the definition of L2 speaking fluency is rather elusive (Freed, 2000), Segalowitz (2012) succinctly defines L2 speaking fluency as follows:

[F]or most, the qualities that make speech fluent include fast speech, and the relative absence of undue hesitations, pausing, repetition, and repairs. (p. 240)

Given the definition of L2 speaking fluency, delivering counterarguments associated with academic content fluently is conceivably a more challenging task unless it is premeditated. In fact, many researchers reported that the development of L2 speaking fluency requires, for instance, planning time (i.e., premeditation) or task repetition (Bygate, 2001; Meahnert, 1998; Seifoori & Vahidi, 2012; Tavakoli & Skehan, 2005). Foregrounded in the previous studies, we interpreted that the interactive and unpredictable nature of rebuttal task in the pilot debate class may have led to the student perceived difficulty presented in Excerpt 1.

4.4 Useful resources to prepare for a debate

In our analysis on the qualitative portion of data, one of the most recurrent themes was collaborative strategies used for debate preparation. While the extent of reported strategies and their details differed, participants' strategic approach to debate preparation was intertwined with the presence of others—debate team members who worked together to prepare for a debate. For instance, participants reported the use of SNS such as LINE to regularly communicate with team

members and to coordinate logistics for a debate presentation (see Excerpt 2).

Excerpt 2
I used LINE to communicate with my team members and decide what to present.

Regular communication with team members was not aimlessly conducted by participants. Rather, participants communicated with their respective team members to exchange their ideas, which helped developing their arguments (see Excerpt 3).

Excerpt 3
I learned a lot from exchanging ideas with my group members. Through that, I learned different ways to argue.

The use of information and communication technology (ICT) tools is inseparably linked to the strategies used by participants. In addition to LINE, participants used the Internet to find topic-related materials and discuss what they learned with their team members (see Excerpt 4)

Excerpt 4
Use the internet to research and then discuss my findings with my team members.

As shown in Excerpts 2, 3, and 4, participants communicated with each other to share ideas and/or discuss their findings in the process of preparing for their debate. These strategic approaches to debate preparation was made possible by the presence of ICT tools—LINE and the Internet, which is now commonplace whether it is in classrooms or outside. To this point, Ghasemi and Hashemi (2011) justly states that ICT has clearly alternated the ways in which we communicate, teach, and learn. What is worth noting at this point is that participants' strategy use in those specific instances was not entirely dependent on social others (i.e., debate team members), but also material others as their strategy use was clearly mediated or afforded by material resources (i.e., the Internet and LINE). Hence, we view that participants' strategy use is deeply embedded in the specific social relationship and material conditions—team-based activity and the availability of the ICT tools.

4.5 Strategy use and self-regulation

In educational psychology, learners' use of strategies to facilitate their learning is referred to as self-regulated learning (SRL) (Zimmerman, 1990). Self-regulated learning focuses on individual autonomous behaviors/actions directed towards achieving a goal in learning via self-monitoring and evaluation (Paris & Paris, 2001). One other theoretical framework relevant to the current study is co-regulated learning (CRL). CRL is grounded in Vygotsky's (1978) extensive work on higher psychological processes in learning. CRL views individual's regulatory behaviors are cultivated through social interaction. Thus, CRL is contrastive to SRL in that social interaction is prerequisite to the emergence of

regulatory strategies in individual learning (McCaslin & Hickey, 2001; McCaslin, 2009).

Although the notion of SRL/CRL is useful in examining individual learning behavior, process, and strategy use, it is deeply rooted in the cognitivist paradigm of research and thus exclusive focus is placed upon explaining an individual's cognitive function and resulting decision/action including his/her strategy use.

While our findings show some relevance to SRL/CRL, the strategies reported by participants were all socially and materially afforded—it is not individually regulated learning. Rather, their regulated learning behaviors/strategy use were collectively coordinated and displayed by groups of learners. Hence, theoretical tenets of SRL/CRL cannot readily offer explanatory value to our findings. In fact, the epistemological assumption of SRL is tightly connected to *ontological individualism* (see Epstein, 2007, for an extensive discussion). In short, SRL/CRL treats individuals as the primary subject of interest and all other external conditions are typically treated as accessories to the individual cognition and behavior.

In response to the limitation of SRL/CRL, some suggested a more socially situated view on self-regulation—socially shared regulation of learning (SSRL) (e.g., Järvelä, Järvenoja, Malmberg, & Hadwin, 2013). SSRL views learning behaviors/strategies as a collective enactment of actions/action plans within a group of individuals to achieve a common goal (Winne, Hadwin, & Perry, 2013). Hence, unlike SRL or CRL, in SSRL the emergence of regulatory strategy is not ascribed to an individual but rather to a group of individuals. This shift in perspective is a significant one particularly when examining collective learner strategy, which is not aimed toward regulating or monitoring one's own learning but socially shared learning processes. In relation to the present study, our participants' strategy use can be seen as a socially and materially constructed shared regulatory state to achieve a unified goal.

4.6 Implications for Research

In the process of debate preparation, participants used a variety of strategies and resources to regulate their learning. SRL/CRL has been receiving much attention from researchers particularly in relation to motivation research (Dornyei, 2001). However, our findings point to regulatory strategy use which cannot be contained in the self as our participants' strategy use was found in social interaction and collective enactment of actions. Regulatory behaviors/strategy use in collaborative learning environment is different from individual learning. Hadwin, Järvelä and Miller (2011), for instance, claims that the conventional notion of self-regulation and/or co-regulation primarily focuses on examining individual learning rather than collaborative learning situations. Suffice it to say that how learning unfolds is inseparably linked to the context of teaching and learning.

If so then, it is perhaps a worthwhile attempt to shift our focus from individual learners to groups of learners as a unit of study to further our understanding of their learning processes and associated regulatory strategy use. Researchers, for example, may examine how a specific activity is

conducted and what social and material resources are present in a given pedagogical environment to map out both social and material conditions from which socially shared regulatory strategy use may emerge.

Such shift in perspectives can help us examine student learning processes in instructional contexts wherein collaborative/active learning is emphasized, as in the case of our pilot debate class. From both theoretical and empirical points of view, SSLR can be a more suited alternative framework to examine collective learner strategy as it encapsulates regulatory learning behaviors coordinated and negotiated via social and/or material interaction commonly observed in group/collaborative learning situations (e.g., Järvelä & Järvenoja, 2011).

4.7 Potential Pedagogical Implications

Overviewing the findings of the present study, there are a few potential pedagogical implications, which we believe to be valuable to those who are currently teaching debate or involved in debate class development.

Our study participants reported that the rebuttal phase of debate was challenging. As we noted earlier, the rebuttal phase is largely a spontaneous speech event and it requires students to present counterarguments within a limited amount of time. Our participants were mostly at advanced level of English language proficiency; there were little or no problem for them to speak English as long as the topic is within the realm of daily conversation. However, debate is entirely different from regular conversational exchange from content to style. Given the fact that some of our participants who were at advanced level of English expressed difficulty with rebutting, it is likely to be even more challenging for less proficient learners. Hence, providing a sufficient amount of time for students to prepare for their rebuttals may well be beneficial to the smoother execution of a rebuttal task. For example, adopting a debate format containing cross-examination can be helpful as it contains a period of time for rebuttal preparation. In addition, students may benefit from predicting opponents' claims at debate preparation stage and ready possible counterarguments before debating since premeditating strategies can enhance speech fluency (Bygate, 2001; Meahnert, 1998; Seifoori & Vahidi, 2012; Tavakoli & Skehan, 2005).

One other important implication was that the high level of course satisfaction indicated by participants may be attributable to the clarity of instruction on debate format and abundant pair/group in-class activities. For the former, participants highly valued the instructor's clear explanation on how a debate must be conducted. Every phase of the debate was clearly defined, and rules and roles were set. This helped students to clearly picture what a debate entails and what expected performances are. As for the latter, debate is a team-based speech event, which naturally requires every student to be actively involved in planning and preparing for a debate. Because of the nature of the activity, it is quintessential for students to work together.

Due to the aspects of debate presented above, tasks/activities provided in the debate class were heavily geared toward reinforcing collaborative learning. Successful collaborative learning is largely dependent upon positive group dynamics, and when collaborative learning is properly executed, learners' overall satisfaction with their learning naturally increases (Beckman, 1990; Chickering, 1991). Thus, the instructor provided many collaborative tasks for students not only to be better prepared for a debate but also to help establish a good rapport amongst students in the classroom, which then can extend over when they work together outside of the classroom and positively affect their learning experience.

5. Conclusion

The present research examined and discussed the results of a course evaluation survey administered to Japanese university students in a pilot debate class. Our findings indicated the high level of student satisfaction with the pilot class. However, they experienced some difficulties with the rebuttal phase of debate due to its specific nature; it is largely spontaneous and requires students to offer counterarguments within a highly limited time frame. The study also identified participants' collective use of regulatory strategies to prepare for a debate. Their strategy use was interlinked to the presence of social and material others as theorized in SSRL.

On a last note, the present research is limited due to its small sample size. Since this is an institution-based case study, our intention is not to claim any generalizability or replicability of the findings or implications put forth. Rather, following the convention of the interpretivist paradigm of research, the applicability of our findings or implications should be judged based on the notion of transferability—if the findings can be relevant to the respective readers in similar instructional contexts.

References
Brown, T., & Brown, J. (2014). *Teaching Advanced Language Skills Through Global Debate: Theory and Practice.* Washington, DC: Georgetown University Press.
Beckman, M. (1990). Collaborative Learning: Preparation for the Workplace and Democracy. *College Teaching*, 38(4), 128–133
Bygate, M. (2001). Effects of task repetition on the structure and control of language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Task-based learning: Language teaching, learning and assessment* (pp. 23–48). London, England: Longman.
Chickering, A. W. (1991). Applying the Seven Principles for Good Practice in Undergraduate Education. In Z. F. Gamson (Eds.), *New Directions for Teaching and Learning*, no. 47.
San Francisco: Jossey Bass.
Dornyei, Z. (Ed.). (2001). *Motivational strategies in the language classroom.* New York: Cambridge University Press.
Epstein, B. (2007). Ontological individualism reconsidered. *Synthese*, *166*(1), 187–213.
Freed, B. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243–265). Ann Arbor: University of Michigan Press.
Ghasemi, B., & Hashemi, M. (2011). ICT: Newwave in English language learning/teaching. *Procedia Social and Behavioral*

Sciences, 15, 3098–3102

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 65–84). New York: Routledge.

Järvelä, S., & Järvenoja, H. (2011). Socially constructed self-regulated learning and motivation regulation in collaborative learning groups. *Teachers College Record, 113*(2), 350–374.

Järvelä, S., Järvenoja, H., Malmberg, J., & Hadwin, A. F. (2013). Exploring socially-shared regulation in the context of collaboration. *Journal of Cognitive Education and Psychology 2*(3), 267–286.

Kennedy, R. (2007). In-class debates; Fertile ground for active learning and the cultivation of critical thinking and oral communication skills. *International Journal of Teaching and Learning in Higher Education*, *19*(2), 183–190.

Lan, W. Y., Bradley, L., & Parr, G. (1993). The effects of a self-monitoring process on college students' learning in an introductory statistics course. *Journal of Experimental Education, 62*(1), 26–40. doi:10.1080/00220973.1996.9943798

Leslie, P., William, S. H., & Douglas, R. S. (2005). *Basic debate student edition.* McGraw-Hill Education.

McCaslin, M. (2009). Co-regulation of student motivation and emergent identity. *Educational Psychologist, 44*(2), 137–146. doi:10.1080/00461520902832384

McCaslin, M., & Hickey, D. T. (2001). Educational psychology, social constructivism, and educational practice: A case of emergent identity. *Educational Psychologist, 36*(2), 133–140.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*, 83–108. doi:10.1017/S0272263198001041

MEXT. (2011). Five proposals and specific measures for developing proficiency in English for international communication. Retrieved from https://www.mext.go.jp/component/english/__icsFiles/afield file/2012/07/09/1319707_1.pdf

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist*, *36*(2), 89–101.

Segalowitz, N. (2012). Fluency. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 240–244). London, England: Routledge.

Seifoori, Z., & Vahidi, Z. (2012). The impact of fluency strategy training on Iranian EFL learners' speech under online planning conditions. *Language Awareness*, *21*(1-2), 101–112. doi:10.1080/09658416.2011.639894

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford, England: Oxford University Press.

Strong, G. (2019). EAP as a bridge to EMI: Learning from the UK. *The Language Teacher, 43*(6), 3–9.

Tahira, M. (2012). Behind MEXT's new Course of Study Guidelines. *The Language Teacher*, *36*(3), 3–8.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–277). Amsterdam, the Netherlands: John Benjamins.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Winne, P., Hadwin, A. F., & Perry, N. E. (2013). Metacognition and computer-supported collaborative learning. In C. E. Hmelo-Silver, A. O'Donnell, C. Chan & C. Chinn (Eds.), *International handbook of collaborative learning* (pp. 462–489). New York: Taylor & Francis.

Yin, K. R. (2003). Case study research: Design and methods (3rd ed.). Thousand Oaks: Sage Publications.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3–17.

Appendix A: Course Evaluation Survey (English Version)

Q1. What is your sex?

☐Male ☐Female

Q2. What is your university year?

☐Sophomore ☐Junior ☐Senior

Q3. What is your major?

Q4. What is your latest TOEIC IP score? When did you receive the score? (e.g., 580, March, 2018)

Q5. Have you ever learned how to debate in class before taking this class?

☐Yes ☐No

Q6. If answered yes, can you briefly describe your experience learning to debate? If answered no, skip this question.

Please indicate the extent of your agreement/disagreement to the following statements.

Q7. I am satisfied with the debate course.

Disagree ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 Agree

Q8. I understand how to conduct a debate in terms of its format.

Disagree ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 Agree

Q9. I gained confidence to debate through learning in debate class.

Disagree ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 Agree

Q10. It is important to search and study topic-related reading materials before conducting a debate.

Disagree ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 Agree

Q12. Please briefly describe good and bad points of the debate course. What would you suggest to improve the quality of the course?

Q13. Please explain how you prepared for debate.

Q14. What was difficult in conducting a debate? Explain.

Q15. What were useful resources/materials when learning to debate? List some examples.

Q16. What were unuseful resources/materials when learning to debate? List some examples.

Q17. Did you experience any problem when learning to debate? If you did, how did you attempt to solve it? Explain.

Effects of Extensive Reading on Reading Fluency Development for Japanese False-Beginner University EFL Learners

Reina Nakamura* and Mika Nishizawa**

*Dokkyo University
reina_nkmr@dokkyo.ac.jp
**Dokkyo University
mikani@dokkyo.ac.jp

Abstract

The present study empirically investigated the effect of extensive reading (ER) on reading fluency development for Japanese false-beginner university EFL learners. Currently, many false beginners attend universities in Japan, and ER has received attention as a potentially effective instructional approach for such learners. However, few empirical studies exist on the cognitive effects of ER on false beginners including the learners' reading fluency development. Simultaneous reading rate and comprehension measurements demonstrated that the participants in the ER treatment group improved their reading fluency more than those in the control group as revealed by several statistical analyses.

Keywords: extensive reading, reading rate, reading fluency, false beginner, EFL reading

1. Introduction

1.1 Extensive Reading

Extensive reading (ER) is an activity in which "the leaner reads huge amounts of very simple text so that she can read smoothly, confidently and pleasurably" (Waring, 2000, p. 6). ER has been receiving attention as an approach to L2 acquisition worldwide (e.g., Cha, 2009; Lao & Krashen, 2000; Lin, 2010). Likewise, there has been an increasing trend of adopting ER in university English classrooms in Japan (e.g., Beglar & Hunt, 2014; Beglar, Hunt, & Kite, 2012; Yamashita, Tsurii, & Herbert, 2011).

One incentive for adopting ER in university English courses in Japan may be that ER can provide the learner large quantities of language input that is generally otherwise difficult to acquire especially outside classrooms. Comprehensible input is essential for language learning (Krashen, 1985, p. 2), and ER can provide large quantities of comprehensible input to the learner when appropriate reading materials are used (Nation, 2007). In addition, ER can provide the EFL learner in Japan a significant opportunity to cultivate his or her competence to communicate in English. Although it is gradually changing, the focus of English education in secondary school in Japan has been to prepare students for college entrance examinations based on the traditional grammar-translation method, and there has been a limited emphasis on cultivation of the learner's communicative competence. The gap between the learner's knowledge about English language features and how these features are used in communication may be moderated significantly by ER, which uses a large amount of authentic and semi-authentic English texts.

A number of positive effects of ER on the development of various aspects of the Japanese university EFL learners' English proficiency have been empirically shown including effects on vocabulary (e.g., Hayashi, 1999; Yamamoto, 2011), reading comprehension (e.g., Hayashi, 1999; Yamashita, 2008), and reading rate and fluency (e.g., Beglar & Hunt, 2014; Beglar et al., 2012). In addition to these cognitive effects, previous studies (e.g., Takase, 2012; Tsurii, 2016) have shown that ER also provides affective effects to learners including improvement of the learners' attitudes toward learning English.

Because of the potential positive cognitive and affective effects of ER, ER has been adopted in Japanese EFL university classes at various proficiency levels including those for false-beginner students. In Japanese universities, false beginners are students who failed to make sufficient progress during their six-year secondary English education to acquire the basic English skills that are necessary for university English education.

When ER is introduced to false beginners, leveled readers (LRs) and/or graded readers (GRs) are typically used. LRs are designed to help young native speakers cultivate reading skills, and they are available in different levels. Unlike LRs, GRs are designed for language learners. They are also available in different levels, and different grammar structures and numbers of headwords are used at each level.

In terms of the empirical evidence of the effects of ER on false beginners, previous studies (e.g., Joichi, 2014; Tsurii, 2016) provided questionnaire-based evidence that ER has positive affective effects including increased satisfaction, interest, and motivation. However, few empirical studies have been conducted on the cognitive effects of ER on these learners, and these effects remain largely unknown.

Currently, many false beginners attend universities because with the decreasing number of children in Japan, students with relatively lower ability can now enter universities. While there has been some evidence for positive affective effects of

ER on these learners as discussed above, some teachers may remain reluctant to adopt ER for their false-beginner students without additional empirical evidence on its cognitive effects. Accordingly, the present study will examine the effect of ER on false beginners in reading fluency development.

1.2 Automaticity and Reading Fluency

Fluency is a complex notion (Grabe & Stoller, 2002, p. 110), and it can be characterized by speed, accuracy, and fluidity (Segalowitz, 2000, p. 200). Because of the intricate processes that are relevant for reading fluency, no simple method exists for rigorous assessment of a learner's reading fluency. With this in mind, since reading rate is one characteristic of reading fluency (Samuels, 2006), previous studies measured the reading rates of learners for evaluating their reading fluency.

To probe into the essence of reading fluency, processes of reading are first reviewed. These processes are frequently classified into two categories: lower- and higher-level processes (e.g. Grabe, 2009, Chapter 2; Grabe & Stoller, 2002, Chapter 1). The lower-level processes include word recognition, syntactic parsing, and semantic proposition formation. These processes can be highly automatized and are highly automated among fluent readers. In contrast, the higher-level processes include cognitive and metacognitive processes that represent what is typically perceived as reading comprehension. These processes include integration of information within a text, activation and utilization of background knowledge in text meaning construction, inferencing, monitoring comprehension, and strategic processing.

Cognitive resources that are available for information processing are limited within an individual. Therefore, if the lower-level processes are not sufficiently automated and consume a large portion of the reader's cognitive resources, the remaining cognitive resources are insufficient to carry out the higher-level processes as rapidly as otherwise. In extreme cases, the higher-level processes may not even be carried out sufficiently for accurate comprehension. Thus, fluent (i.e., rapid and accurate) text comprehension is possible only with automatization of the lower-level processes. For beginning readers, many of the lower-level processes are not yet automatized, thus, automatization of these processes is crucial for improving their fluent reading.

Reading fluency development requires a large amount of reading as discussed in Grabe and Stoller (2002, p. 47), and the use of ER for reading fluency development is supported by second language (L2) authorities (Beglar et al., 2012). However, there are few studies in the literature that empirically examined the effects of ER on reading fluency development on L2 learners including those on Japanese university EFL learners.

1.3 The Present Study

The present study extends previous research on the effect of ER on reading fluency development among Japanese university EFL learners in two ways. First, to the best of the authors' knowledge, of the studies on the effect of ER on

reading fluency development among Japanese university EFL learners, none focused on false-beginner learners except for Taguchi, Takayasu-Maass, and Gorsuch (2004) and Nakamura (2017). Even these two studies were limited in some aspects. In Taguchi et al. (2004), the participants were a mix of false and upper beginners although most of them were likely false beginners. In addition, the pre-tests and post-tests were determined to be inequivalent for measuring the participants' reading rate and comprehension. It was also the case that the participants did not adequately comprehend the testing passages. In Nakamura (2017), since no control group was included, it was not possible to strictly conclude whether the statistically significant reading fluency gain found in the study was attributable to ER. Second, many of the previous studies lacked rigor in evaluating reading fluency development, which calls for new studies with more rigorous assessment of reading fluency development. Specifically, as was the case for Nakamura (2017), some of the studies lacked control groups (e.g., Keith, 2015; Yamashita & Kan, 2011; Yukimaru & Pennington, 2013) although control groups are critically important for investigating the effect of a treatment (Nakanishi, 2015). Some other studies did not accompany concurrent assessment of comprehension (e.g., Robb and Susser, 1989; Yukimaru & Pennington, 2013). Since reading fluency can be characterized by speed, accuracy, and fluidity of processing (Segalowitz, 2000), reading fluency development requires reading rate gains to be accompanied by high levels of comprehension (Beglar & Hunt, 2014). Without concurrent comprehension measurements, it remains unclear whether the learners comprehended texts that were used for the reading rate measurements and also whether reading rate gain occurred at the expense of comprehension.

Given the background discussed above, the following two research questions are addressed in the present study:

1) Do false-beginner learners' reading rates increase after they engage in ER for one semester?
2) If the false-beginner learners' reading rates increase with ER, does it occur without adversely affecting their reading comprehension?

In the present study, increasing reading fluency is operationalized as increases in reading rate while maintaining adequate levels of comprehension similarly to Beglar and Hunt (2014).

2. Method
2.1 Participants

Participants were 50 first-year Japanese undergraduate students who were majoring in economics. They were enrolled in a required reading course and participated in the study with consent on a voluntary basis. Twenty-two participants attended a class taught by the first author (treatment group (TG)) and twenty-eight participants attended a class taught by the second author (control group (CG)). However, since four TG participants were not able to appropriately complete the reading tests administered for the study, they were excluded from the analysis.

In addition to the reading course, the TG participants were enrolled in three other required English courses, which were in writing, listening, and speaking. Similarly, the CG participants were enrolled in two other required English courses, which were in listening and speaking. All of these required courses met once a week as was the case for the reading course.

The English proficiency levels of all the participants were at the false-beginner level as indicated by their TOEIC test scores (Table 1.).

Table 1. Summary of two groups

Group	Number of participants included in analysis (N)	TOEIC scores
TG	18	285–320
CG	28	240–265

2.2 ER Materials

The following graded readers (GRs) were adopted as ER materials: Cambridge English Readers, Oxford Bookworm Library, Dominoes (publisher: Oxford University Press), Penguin Readers (publisher: Pearson), and MacMillan Readers. The headwords of these books ranged between 250 and 800, and a total of 58 GR titles were available. None of the GRs was in comic-book style.

2.3 Design of the Study

The present study was conducted over 12 calendar weeks during a semester. The TG participants read a GR as homework on a weekly basis except for one occasion. On this occasion, they read only a single GR during a two-week period since two class sessions were separated by two weeks because of a holiday.

The GRs were made available in the classroom so that the participants were able to select a book to read at the end of each class session. The participants were instructed not to read the same book more than once during the ER treatment period.

One characteristic of ER is that the learner chooses the reading material to read (Waring, 2000). To assist the participants in choosing a GR that they can read smoothly, confidently and pleasurably, they were advised to choose a GR for which they were able to answer "yes" to all of the following questions from Waring (n.d.):

Q.1   Can I read the book without a dictionary?
Q.2   Can I read it quite quickly, without pausing?
Q.3   Can I understand most of the book?
Q.4   Is the book enjoyable?

In relation to Q.1, the participants were repeatedly informed that they should choose a GR in which the percentage of unknown vocabulary was about 2%, based on the finding of Hu and Nation (2000).

Two types of ER-related activities were incorporated into the class activities: submission of book reports and in-class oral reports on the GRs the participants read. These activities were adopted because such activities may encourage the participants' participation in ER (Sheu, 2003) and also ensure that the participants read the contents of the GRs closely.

The book report asked the participants questions that are difficult to answer without reading the content closely. These questions were created and adopted based on students' responses to book reports from classes in previous years. The grades from the book report were incorporated into the participants' course grades.

In every class session held immediately after reading a GR, the participants gave oral reports about the GR in pairs. One participant gave a two-minute report about the GR to the other participant, who asked two or more questions after the report. The roles were then switched.

3.   Materials

3.1 Reading Test

A reading test was administered twice to measure the participants' reading rate and comprehension in both CG and TG. The first test (pre-test) was administered during the class period at the end of which the TG participants chose their first GR to read after class. The second test (post-test) was administered during the class period which was held immediately after they finished reading their last GR. These two tests were administered to CG on the same days as they were administered to TG.

An identical set of passages and comprehension questions were used for pre- and post-tests since use of such tests ensures that they are objectively comparable (Lin, 2010). Use of an identical reading test may raise the concern that the participants' memory of the passages and comprehension questions from the pre-test affects the results of the post-test. However, such effects were considered minimal for the following two reasons based on the discussions by Lin (2010). First, administrations of the pre-test and post-test were separated by twelve weeks, which was considered sufficiently long to minimize the effects of the participants' memory of the pre-test on their post-test performance. Second, the participants were never informed of the correct answers for the comprehension questions on the pre-test, which made it highly unlikely that administration of the pre-test affected the participants' comprehension question scores on the post-test. In addition, the participants did not have an incentive to remember the content of the pre-test as they were never informed that the content of the post-test would be identical to that of the pre-test, which further justifies the use of identical pre- and post-tests.

Two passages, Elephants and Reading Books from *Reading for Speed and Fluency 1* (Nation & Malarcher, 2007) were used for the reading test. The Flesch-Kincaid (F-K) Grade Levels of the two passages were 4.5 and 4.2, respectively. Each of the passages was accompanied by eight multiple-choice (3-answer) questions. When the reading test was prepared, care was taken so that the test takers would not be able to choose the correct answer without reading the passage, that is, with the use of their common sense or general knowledge alone.

Both the pre- and pos-tests were administered with the

following procedure. First, the participants silently read one of the passages and recorded the time they needed to finish reading the passage. While the time record was kept with a 1-second increment in TG, it was kept with a 10-second increment in CG for technical reasons. Next, the participants answered eight comprehension questions on the back side of the passage without referring to the passage. Then, they repeated these steps for the other passage. Before taking both pre- and post-tests, they were instructed to read each passage carefully at their regular reading speed. This instruction was given to ensure that they could comprehend the passages adequately without being hasty and answer the comprehension questions correctly. It was also explained to the participants that they were allowed to jump back during the reading as long as it was a sentence or two.

To minimize the effect of the participants' unfamiliarity with the test format on their performance, a practice reading test in the format of the pre- and post-tests was given in both groups one week prior to the administration of the pre-test.

3.2 Questionnaire

A post-questionnaire was administered in Japanese to both CG and TG after TG's ER treatment period. The purpose of the questionnaire was to identify the reading activities of the participants during the ER treatment period.

4. Results
4.1 ER Reading Amount

Since some of the TG participants missed classes and/or reading a GR as their homework, the number of GRs read by the participants ranged from 8 to 11 with a mean of 10.5. Similarly, the total number of standard words read by the participants ranged from 11,963 to 41,955 with a mean of 21,091. A standard word is six character spaces as defined by Carver (1976), in which letters, spaces, and punctuation marks count as character spaces. The average running word length and the difficulty of a text are correlated (Carver, 1976), and the use of standard words allows for more rigorous comparisons across various studies (Beglar et al., 2012; Huffman, 2014). The number of standard words in each GR was calculated using the method adopted by Huffman (2014). A summary of the difficulty level of the GRs read by the participants in terms of the number of headwords is shown in Table 2.

Table 2. The number of headwords of the GRs read by the TG participants. The numbers in parentheses indicate the percentage of GR checkouts with respect to the total, i.e., 191.

Headwords	Number of GR checkouts
250−300	150 (78.5%)
400	25 (13.1%)
600	15 (7.9%)
800	1 (0.5%)

4.2 Descriptive Statistics of Reading Rates

To investigate Research Question 1) (Section 1.3), several statistical analyses were performed using the pre-test and post-test reading test data. Prior to these analyses, the time records from TG were adjusted as though it had been recorded in 10-second increments to match with the time records from CG. The reading rates reported hereafter are in standard words per minute (swpm).

The descriptive statistics for the participants' reading rates from the pre- and post-tests are summarized in Table 3. The mean reading rates of TG prior to and immediately after the ER treatment were 74.8 swpm and 95.8 swpm, respectively. The corresponding mean reading rates of CG were 62.7 swpm and 67.8 swpm, respectively. In both groups, the mean reading rates on the post-test were higher than those on the pre-test.

Table 3. Statistics of the participants' reading rates in standard words per minute (swpm) calculated from pre- and post-tests. *M*: mean, *SD*: standard deviation

Group	N	Pre-test		Post-test		Difference	
		M	SD	M	SD	M	SD
TG	18	74.8	14.6	95.8	21.5	21.0	15.4
CG	28	62.7	10.2	67.8	14.9	5.0	11.0

4.3 ANCOVA on Post-test Reading Rates

To examine the effect of the ER treatment on the participants' reading rates, the statistical significance of the difference in the pre-test reading rate between CG and TG was first examined. Since the reading rate data met all of the assumptions required for the Mann-Whitney U test but did not meet those required for the independent *t*-test, the Mann-Whitney U test was selected for the analysis. The statistical significance level of the analysis was set at $p < 0.01$ for all the statistical tests conducted in the present study. Visual inspection revealed that the distributions of the pre-test reading rates for the two groups were dissimilar. The pre-test reading rates for TG (mean rank = 30.28) were statistically significantly higher than those for CG (mean rank = 19.14), $U = 374$, $z = 2.748$, $p = .006$.

Because the difference in the pre-test reading rate between CG and TG was statistically significant, a one-way analysis of covariance (ANCOVA) was performed to examine the effect of ER on the post-test reading rate after controlling for the pre-test reading rate. Before performing the analysis, it was confirmed that the reading rate data met the required assumptions for ANCOVA. The results of the ANCOVA are summarized in Table 4. After adjusting for the pre-test reading rates, the analysis revealed that post-test reading rates from TG were statistically significantly higher than those from CG, $F(1,43) = 12.99$, $p = 0.001$, *partial* $\eta^2 = 0.232$. Since there was a statistically significant difference in the pre-test reading rate between CG and TG, the effect size of ER on the reading rate, d_1, was calculated using the mean reading rate gains and the standard deviation of the reading rates of CG (Glass, McGaw, & Smith, 1981, pp. 114-119):

$$d_1 = \frac{M_t - M_c}{SD_{pc}}$$

where

M_t: the mean reading rate gain for TG,
M_c: the mean reading rate gain for CG,
SD_{pc}: the standard deviation of the post-test reading rate of CG

The calculation yielded $d_1 = 1.07$ for the effect size of ER on the reading rate. (The effect size of ER on the reading rate, d, is evaluated using three different definitions in this paper; a subscript 1, 2, or 3 is added to indicate which definition it refers to.)

Table 4. Results of one-way ANCOVA

Source	SS	Df	MS	F	P	partial η^2
Group	2200.19	1	2200.19	12.99	0.001	0.232
Error	7284.98	43	169.42			

While ANCOVA has been used in applied linguistics research to control for the effects of a covariate, e.g., the pre-test reading rate in the present study, it remains somewhat controversial in that the effects may not be fully controlled for (Loewen & Plonsky, 2016, p. 4). Accordingly, additional analyses were performed to examine the effect of the ER treatment on the participants' reading rates.

4.4 Paired-samples t-Test for Within-group Reading Rate Gains

A paired-samples t-test (paired t-test hereafter) was performed to examine the statistical significance of the mean reading rate gain within each group. The required reading rate data for both groups met all the requirements for the paired t-test, and the following results were obtained: the mean reading rate gain for CG, $M = 5.0$, was statistically insignificant, 99% CI [-0.74, 10.8], $t(27) = 2.41$, $p = 0.023$ (two-tailed), $d_2 = 0.46$ and that for TG, $M = 21.0$, was statistically significant, 99% CI [10.5, 31.4], $t(17) = 5.78$, $p = 0.000022$ (two-tailed), $d_2 = 1.4$. Here, d_2 refers to the effect size for paired-observations (Cohen 1969, p. 46):

$$d_2 = \frac{M_z}{SD_z}$$

where

M_z: the mean reading rate gain within each group,
SD_z: the standard deviation of the reading rate gain within each group

in the present study.

4.5 Independent-samples t-Test for Between-group Reading Rate Gain Comparison

Subsequently, the reading rate gains from CG and TG were compared using the independent-samples t-test (t-test hereafter). Prior to the analysis, it was confirmed that the data met all the assumptions required for the t-test. The analysis

indicated that there was a statistically significant difference in the reading rate gain between CG and TG. The reading rate gain from TG ($M = 21.0$, $SD = 15.4$) was higher than that from CG ($M = 5.0$, $SD = 11.0$), with a statistically significant difference, M =16.0, 99% CI [-26.4 -5.46], $t(44) = -4.098$, $p = 0.000176$ (two-tailed), $d_3 = 1.2$. Here, d_3 represents the effect size as in Cohen (1969, p 64):

$$d_3 = \frac{M_t - M_c}{SD_{pooled}}$$

where

$$SD_{pooled} = \sqrt{\frac{(n_t - 1)SD_t{}^2 + (n_c - 1)SD_c{}^2}{n_t + n_c - 2}}$$

in which

M_t, M_c: the mean reading rate gains for TG and CG,
SD_{pooled}: the pooled standard deviation of the reading rate gains,
SD_t, SD_c: the standard deviations of the reading rate gains for TG and CG,
n_t, n_c: the numbers of TG and CG participants

in the present study.

4.6 Descriptive Statistics of Comprehension Scores

The descriptive statistics for the participants' comprehension scores from the pre- and post-tests are summarized in Table 5. The maximum possible comprehension score was 16 for answering all 16 comprehension questions correctly. The mean comprehension scores of TG prior to and immediately after the ER treatment were 12.5 and 13.5, respectively. The corresponding mean comprehension scores of CG were 12.2 and 12.9, respectively. In both groups, the mean comprehension scores on the post-test were higher than those on the pre-test.

Table 5. Statistics of the participants' comprehension scores from pre- and post-tests

Group	N	Pre-test		Post-test		Difference	
		M	SD	M	SD	M	SD
TG	18	12.5	2.3	13.5	1.8	1.0	2.3
CG	28	12.2	2.3	12.9	2.3	0.7	2.8

4.7 The Wilcoxon Signed-rank Test for Within-group Pre- and Post-test Comprehension Scores

To examine the effect of the ER treatment on the participants' comprehension scores, statistical significance of the difference in the mean pre-test comprehension scores between CG and TG was first examined with a t-test. The analysis revealed that the difference in the mean pre-test comprehension score between the two groups, i.e., 0.32, was not statistically significant, 99% CI [-2.20, 1.56], $t(44) = -0.461$, $p = 0.647$ (two-tailed). Prior to the t-test analysis, it was confirmed that the data used for the analysis met all the

assumptions required for a *t*-test.

Subsequently, the statistical significance of the difference between the pre- and post-comprehension scores in Table 5 was determined for each group. While the comprehension score data for TG that were required for the analysis met all the requirements for the paired *t*-test, those for CG did not. However, the data for both groups met all the requirements for the Wilcoxon signed-rank test. Therefore, the Wilcoxon signed-rank test was performed for both groups, and the paired *t*-test was additionally performed for TG only. The Wilcoxon signed-rank test revealed that the differences in the pre- and post-test comprehension scores for both groups were statistically insignificant, $z = 1.233$, $p = 0.218$ for CG and $z = 1.727$, $p = 0.084$ for TG. The difference between the pre- and post-test comprehension scores for TG was also determined to be statistically insignificant by the paired *t*-test, 99% CI [-0.54, 2.54], $t(17)=1.886$, $p = 0.076$ (two-tailed).

5. Discussion

5.1 Do the false-beginner learners' reading rates increase after they engage in ER for one semester?

After controlling for the pre-test reading rate, ANCOVA on the post-test reading rate revealed that the post-test reading rates from TG were statistically significantly higher than those from CG (Section 4.3). As mentioned earlier, some argue that ANCOVA does not fully control for the effects of a covariate, e.g., the pre-test reading rate in the present study. However, since the other two statistical tests on the reading rate, i.e., 1) the paired-samples *t*-test for within-group reading rate gains and 2) the independent-samples *t*-test for between-group reading rate gain comparison, indicated that the reading rate gain of TG was statistically larger than that of CG, it can be safely concluded that the reading rates of the participants in TG increased more than those in CG.

The mean reading rate gain in TG, 21.0 swpm (28.1% gain), was roughly the same as that from Nakamura (2017), in which the participants were false-beginner learners with TOEIC scores of 275 – 310. Specifically, the mean reading rate gain of the participants who received ER treatment in that study was 15.8 swpm (26.2% gain) (unpublished data from Nakamura (2017); the mean reading rate gain and percentage gain reported in Nakamura (2017) was based on running words per minute). Although the reading tests adopted in Nakamura (2017) differed somewhat from those adopted in the present study, the English proficiency of the participants and the semester-long ER treatment were nearly identical between the two studies. The exact number of words read by the participants was not reported in Nakamura (2017); however, the participants read a total of 9.6 GRs on average. Since the participants from the present study read 10.5 GRs on average and the levels of the GRs read were comparable in the two studies, the results from these two studies combined reveal the robustness of the mean reading rate gain measured in the present study.

Subsequently, the reading rate gain for TG from the present study, i.e., 21.0 swpm, is compared to those assessed for Japanese university EFL learners by Beglar et al. (2012) and Huffman (2014), which adopted study designs that enabled more rigorous assessments of reading rate gains than those from preceding studies. In Beglar et al. (2012), Pleasure Reading Group 3 attained a mean reading rate gain of 16.85 swpm after reading, on average, 200,170 standard words in two semesters. While all three of the pleasure reading groups in Beglar et al. (2012) read a combination of GR and unsimplified books designed for native speakers, the percentage of the GR text read by Pleasure Reading Group 3 was 87.5% and was the largest of all the groups in that study. This justifies the use of the reading gain data from this group over the other two groups for comparison to the reading gain of TG from the present study. In Huffman (2014), the ER group attained a mean reading rate gain of 20.73 swpm after reading 80,201.74 standard words on average and engaging in six timed readings in one semester. Although the English proficiency of the participants in the two studies are not explicitly specified, the levels of the GR and other reading materials read by the participants imply that they were likely a mix of beginner- and intermediate-level learners.

Given that TG from the present study read an average of 21,091 standard words, which is much less than the amounts read by the above-mentioned participants in Beglar et al. (2012) and Huffman (2014), the mean reading rate gain of 21.0 swpm from the present study may appear to be somewhat large. One possible explanation is that, compared to learners with higher proficiency, it takes false beginners a smaller amount of ER reading to increase a certain amount of reading rate. In the case of false beginners, the number of words, phrases, and grammatical structures that they are familiar with is smaller compared to the case of higher-proficiency learners. Thus, at the false-beginner level, in comparison to higher proficiency levels, many of the words, phrases, and sentence structures that learners are familiar with are likely recycled at a higher frequency in GRs that they can read smoothly, confidently and pleasurably for the purpose of ER. Thus, at the false-beginner level, a certain amount of ER likely elicits more rapid overall automatization of lower-level reading processing, e.g., word recognition, syntactic parsing, and semantic proposition formation (Section 1.2), involving the words, phrases, and grammatical structures that learners are familiar with than is the case for students at higher proficiency levels.

Finally, the effect sizes, *d*, for the effect of ER treatment on the reading rate development in Sections 4.3 –4.5, i.e., $d_1 = 1.07$, $d_2 = 1.4$, $d_3 = 1.2$, represent large effects according to the operational definitions set by Cohen (1969, pp. 22-25). These effect sizes may be easily visualized by examining their implications: the average participant from TG would be in the 86th, 92nd, and 89th percentiles in the CG distribution, respectively.

5.2 If the false-beginner learners' reading rates increase with ER, does it occur without adversely affecting their reading comprehension?

The mean comprehension rates of TG were 78.1% and 84.4% on the pre- and post-tests, respectively, and this pre-post change was statistically insignificant (Section 4.7). Given that scores of seven or eight out of ten on a

comprehension test are considered appropriate for careful silent reading (Nation, 2005), it can be said that the participants read the passages on both tests with adequate comprehension rates. Therefore, the participants' mean reading rate increased without loss of comprehension, indicating that the TG participants' reading fluency increased over the period of the ER treatment.

The reading fluency development in TG is likely attributable to the development of automatization in word recognition and subsequent lower-level processes during the ER treatment. Furthermore, this development likely benefited from the implementations of the two sets of guidelines by Waring (*n.d.*) and Hu and Nation (2000) for selecting a GR to read for the ER treatment (Section 2.3). These two guidelines ensured that many, if not all, participants in TG would read GRs that they were able to read with the vocabulary and grammar they were already familiar with, creating conditions conducive to the development of automatization in lower-level reading processes.

5.3 Pedagogical implications

The empirical evidence in the present study on the effect of ER on false beginners' reading fluency development has practical pedagogical implications. First, ER can be adopted as an efficient method to develop false-beginner learners' reading fluency given that weekly reading of a GR over a semester results in a significant reading rate gain, i.e., by nearly 30% (Section 5.1). Second, since ER elicits positive affective effects among these learners including increased satisfaction, interest, and motivation (Section 1.1), it can be said that false-beginner learners can benefit from ER in both the cognitive and affective domains, giving more incentives for teachers to adopt ER in classrooms with these learners than earlier. To this end, it is important to recall that reading fluency development likely results when teachers ensure that learners choose and read GRs from appropriate levels that they can read smoothly, confidently and pleasurably.

5.4 Limitation of the study

There are several limitations of the study, some of which are attributable to the use of two intact classes. First, the participants' exposure to English outside the reading course differed between TG and CG. However, the influence of this factor on the evaluation of the effect of ER on the reading fluency was likely minimal given the following circumstances. As mentioned earlier, the participants in TG were additionally enrolled in required listening, speaking and writing courses, while the participants in CG were additionally enrolled in required listening and speaking courses only. However, none of these courses were reading intensive as confirmed by the post-questionnaire (Section 3.2). As for voluntary reading the participants engaged with outside their enrolled courses, the questionnaire revealed that only two participants, both in CG, read additional reading materials on a voluntary basis.

Another limitation of the study is that the initial English proficiencies of the participants in the two groups were somewhat different. The English proficiency of the participants in TG was higher (TOEIC: 285-320) than that of the participants in CG (TOEIC: 240-265). This difference might have caused a small but non-negligible overestimation of the effect of ER on the reading fluency development for the following reason: When two groups of learners with a relatively small difference in their English proficiency (as was the case for TG and CG) are compared, the degree of overall automatization of lower-level processes that they possess for reading and comprehending level-appropriate text is likely already higher in the group with higher proficiency. Accordingly, the learners in that group may need less exposure to print for a certain degree of further automatization of these processes than the learners in the other group.

Finally, the number of participants in the present study was small, $N = 18$ for TG and $N = 28$ for CG. Although the robustness of the reading rate gains measured in the present study was supported by a comparison with those measured in a previous study (Section 5.1), future studies with a larger number of participants are desirable for more robust assessments of the effect of ER on the reading fluency development of false-beginner learners.

6. Conclusion

The present study investigated the effects of ER on the reading fluency development of Japanese false-beginner university EFL learners. Simultaneous reading rate and comprehension measurements demonstrated that the participants in the ER treatment group improved their reading fluency more than those in the control group as revealed by several statistical analyses. This finding is significant given that there are few studies in the literature which empirically examined the reading fluency development of L2 learners with rigor, let alone that of Japanese false-beginner university EFL learners. Several limitations of the present study were discussed, which supports the need for an improved study design for a more refined estimate of the effect of ER on reading fluency development for false beginners.

References

Beglar, D., & Hunt, A. (2014). Pleasure reading and reading rate gains. *Reading in a Foreign Language*, *26*, 29–48.

Beglar, D., Hunt, A., & Kite, Y. (2012). The effect of pleasure reading on Japanese university EFL learners' reading rates. *Language Learning*, *62*, 665–703. doi:10.1111/j.1467-9922.2011.00651.x

Carver, R. P. (1976). Word length, prose difficulty, and reading rate. *Journal of Reading Behavior*, *8*, 193–203.

Cha, J.-E. (2009). The effects of extensive reading on enhancing vocational high school students' L2 vocabulary & reading rates. *English Teaching*, 64, 3–30. doi:10.15858/engtea.64.3.200909.3

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York, NY: Academic Press.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publication.

Grabe, W. (2009). *Reading in a second language: Moving*

from theory to practice. New York, NY: Cambridge University Press.

Grabe, W., & Stoller, F. L. (2002). *Teaching and researching reading.* Harlow, UK: Pearson Education.

Hayashi, K (1999). Reading strategies and extensive reading in EFL classes. *RELC Journal*, *30*, 114–132. doi:10.1177/003368829903000207

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–430.

Huffman, J. (2014). Reading rate gains during a one-semester extensive reading course. *Reading in a Foreign Language*, *26*, 17–33.

Joichi, M. (2014). *Rimedhiaru kyouiku ni okeru eigotadoku: yomukoto eno iyoku wo takameru youin* [Extensive reading in remedial English class: Variables influencing the development of students' favorable attitudes]. *Language Education*, *12*, 27–40.

Keith, B. (2015). Measuring the effects of reading fluency practice and extensive reading on L2 reading fluency. In P. Clements, A. Krause, & H. Brown (Eds.), *JALT 2014 Conference Proceedings* (pp. 511–521). Tokyo, Japan: JALT.

Krashen, S. (1985). *The input hypothesis: Issues and implications.* New York, NY: Longman.

Lao, C. Y., & Krashen, S. (2000). The impact of popular literature study on literacy development in EFL: More evidence for the power of reading. *System, 28*, 261–270. doi:10.1016/S0346-251X(00)00011-7

Lin, L-F. (2010). Senior high school students' reading comprehension of graded readers. *Journal of Language Teaching and Research*, *1*, 20–28. doi:10.4304/jltr.1.1.20-28

Loewen, S., & Plonsky, L. (2016). *An A - Z of applied linguistics research methods.* New York, NY: Palgrave Macmillan.

Nakamura, R. (2017). *Gijishoshinsha (false beginners) taishou no tadoku no kouka to yakuwari ni tsuite* [The effects and roles of extensive reading for false-beginner university students]. *Dokkyo Journal of Language Learning and Teaching*, *6*, 59–77.

Nakanishi, T. (2015). A meta-analysis of extensive reading research. *TESOL Quarterly*, *49*, 6–37. doi:10.1002/tesq.157

Nation, P. (2005). Reading faster. *PASAA, 36*, 21–37.

Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, *1*, 2–13. doi: 10.2167/illt039.0.

Nation, P., & Malarcher, C. (2007). *Reading for speed and fluency 1.* Seoul, Korea: Compass Publishing.

Robb, T. N., & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, *5,* 239–251.

Samuels, S. J. (2006). Reading fluency: Its past, present, and future. In T. Rasinski, C. Blachowicz, & K. Lems (Eds.), *Fluency instruction: Research-based best practices* (pp. 7–20). New York, NY: The Guilford Press.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200–219). Ann Arbor, MI: University of Michigan Press.

Sheu, S. P.-H. (2003). Extensive reading with EFL learners at beginning level. *TESL Reporter, 36*, 8–26.

Taguchi, E., Takayasu-Maass, M., & Gorsuch, G. J. (2004). Developing reading fluency in EFL: How assisted repeated reading and extensive reading affect fluency development. *Reading in a Foreign Language*, *16*, 70-96.

Takase, A. (2012). The impacts of extensive reading on reluctant Japanese EFL learners. *The European Journal of Applied Linguistics and TEFL*, *1*, 97–113.

Tsurii, C. (2016). *Tadoku kyouzai (graded readers) wo shiyou shita yomi no shidou ga ataeru eikyou: hieigosenkou no shokyuu reberu gakushuusha wo taishou to shita baai* [The effects of extensive reading instruction with graded readers on EFL learners: questionnaire surveys of non-English-major university students]. *Journal of Humanities Research, St. Andrew's University*, *4*, 37–77.

Waring, R. (n.d.). [How to select graded readers to read]. Retrieved from http://www.robwaring.org/er/Graded_readers/Using_GRS.htm

Waring, R. (2000). *Guide to the 'why' and 'how' of using graded readers.* Tokyo, Japan: Oxford University Press.

Yamamoto, Y. (2011). Bridging the gap between receptive and productive vocabulary size through extensive reading. *The Reading Matrix*, *11*, 226–242.

Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency. *System*, *36*, 661–672. doi:10.1016/j.system.2008.04.003

Yamashita, J., & Kan, K. N. (2011). Examining effects of L2 extensive reading in the cognitive and affective domains. In British Association for Applied Linguistics (Ed.), *Proceedings of the 43rd annual meeting of the British Association for Applied Linguistics* (pp. 375–385). London, UK: Scitsiugnil Press.

Yamashita, M., Tsurii, C., & Herbert, H. (2011). Exploring the relationship between extensive reading and EFL learners' reading proficiency. *Kwansei Gakuin University Humanities Review*, *16*, 73–86.

Yukimaru, N., & Pennington, W. (2013). Introduction of extensive reading to Japanese university students: A case study. *Center for Language Education: Annual Review of Language Learning and Teaching*, *3*, 1–13.

Benefits and Difficulties with CLIL Physics for University Students in Japan

Yoko Maekawa*

Okayama University of Science
y_maekawa@dls.ous.ac.jp

Abstract

The globalized world of today has compelled Japanese English education, especially in the field of science and engineering, to enhance its curricula to prepare students for their future careers. As a result, universities have started offering English for specific purposes (ESP) courses or subject courses taught only in English. Content and language integrated learning (CLIL) may be a suitable approach for students because it enables them to learn the content of their majors as well as the technical English used in such a field. Moreover, CLIL promotes higher order thinking skills such as analysis, evaluation and creation. The author herein planned and conducted a three-week CLIL physics course for second-year university students as well as a survey to measure the effectiveness of the course. The results showed that students improved their English physics vocabulary and English comprehension, and formed a deeper understanding of the content. It was observed, however, that the students lacked English expression skills even though many of them commented that they learned how to explain in English after taking the course. This study revealed that the English skills of students made a difference in their reaction to the course.

Keywords: CLIL, physics, science and engineering students, attitudes and motivation

1. Introduction

The globalized world of today has imposed a strong demand for global human resource development in Japanese universities and colleges. This development has prompted changes to English education in Japan (MEXT, 2003, 2012a). First, practical and pragmatic English education has become a requirement for science and engineering students because they have to acquire the knowledge and skills to communicate in English for various purposes. On this account, studies have been conducted in the field of English for specific purposes (ESP) to evaluate the needs of students and design English courses that are more authentic and practical (e.g., Miyama, 2007; Miyama & Nitta, 2003; Tsuda, 2006; Yamauchi, 2005). ESP researchers, however, have faced difficulties in designing appropriate curricula since they mainly teach first- or second-year students, whose English proficiency and content knowledge are inadequate. Students in these courses are unsure of how their future career will unfold, which makes it difficult to formulate "authentic" English curricula (Anthony, 2009; Anthony, Noguchi, & Orr, 1998; Gally, 2009; Noguchi, 2010). With these struggles and persisting demands for practical English education, English-medium instruction (EMI) courses were introduced in several universities for students to learn both the subject matter in their specialized field and the English needed in such fields (Bradford & Brown, 2018; Kojima & Yashima, 2017). The class was principally taught by content professors, which also created difficulties for both teachers and students. These courses were especially challenging for the students whose English proficiency level was low (Kojima & Yashima, 2017). In recent times, the Japanese government has emphasized the importance of active learning as a way to better promote students' more in-depth learning of each subject, active participation in the learning process, and self-regulated or autonomous learning (MEXT, 2012b). In fact, many universities have started to incorporate active learning in their syllabi. The author posited that Content and Language Integrated Learning (CLIL) can be a solution for the struggles in practical English education and may address the demands for active learning in English. This study briefly introduces CLIL, the practice of CLIL Physics for university students, and the problems therein.

2. Content and Language Integrated Learning (CLIL)

Content and Language Integrated Learning (CLIL) is "a dual focused educational approach in which an additional language is used for the learning and teaching of both content and language" (Coyle, Hood, & Marsh, 2010, p. 1). According to Morton and Llinares (2017), the main original objectives were to enhance multilingualism and multicultural citizenship and improve foreign language learning and teaching methodologies. In the Japanese educational context, CLIL has been compared to and discussed with Content Based Instruction (CBI) and EMI where it is difficult to distinguish the differences (Hanzawa, 2017; Kojima & Yashima, 2017; Watanabe, Ikeda, & Izumi, 2011). The main characteristic of CLIL is the key concepts of the 4Cs (content, communication, cognition, and community/culture) and their intertwined roles in constructing an adequate curriculum. According to Watanabe et al. (2011), content denotes the curricular subjects. In CLIL courses, instructors need to present the content in a comprehensible manner. Communication, both written and oral, is the language aspect that the learners must accomplish. For the communication aspect, the curriculum design includes the plan of "language of learning," "language for learning," and "language through

learning." Cognition may be the central aspect of CLIL and aims at promoting students' cognitive engagement (Coyle et al., 2010). Based on Bloom's taxonomy, the CLIL curriculum is supposed to cover both lower-order thinking skills (LOTS) and higher-order thinking skills (HOTS) (Coyle et al., 2010). LOTS consist of remembering, understanding, and applying while HOTS include analyzing, evaluating, and creating. The last C is difficult to define. Since the CLIL started in Europe where multilingual and multicultural understanding is crucial and more familiar to students, the fourth C was originally defined as culture. However, Japanese people find it difficult to interact with the culture behind the language. Accordingly, the concept of community may be better (Watanabe et al., 2011). Using the 4Cs concept, the author designed a three-week CLIL physics course for university students.

3. Class Design and Practice

This study was conducted in collaboration with the CLIL project in Okayama University of Science. The project aimed to prepare faculty members for a newly required specialized English class for third-year students, which was to be taught by the content professors of each department. The author joined the project as a language teacher who could also teach science content because of her engineering background. The author chose the content that she would be comfortable teaching. The goals of this CLIL Physics course were for students to understand the fundamental concept of dynamics (Newton's second law of motion), explain the basic mechanism of dynamics in English, and solve physics questions written in English. The content for each week was planned according to CLIL's 4Cs and was developed in consultation with the project team. Table 1 introduces the first-week course based on the 4Cs. The details of how the author conducted each class and the other class contents are discussed elsewhere (Maekawa et al., 2017). To reduce students' anxiety especially in lower level classes, the author allowed students to use Japanese in pair or group work but instructed them to produce their end-results in English.

Table 1 First-Week Course Design Based on CLIL's 4Cs

4Cs		
Content		Understanding "Newton's second law of motion" and "normal force"
Communication	Language of Learning	the second law, normal force, equilibrium, frictionless, exert, magnitude, object, tension
	Language for Learning	Do you know the second law? Who made the law? Do you know the normal force? Find the symbol indicating the normal force from the figure. What value do you have to obtain? Can you think of any examples of the second law in your daily life? What are the examples of the normal force in your daily life? Discuss in your group.
Cognition	LOTS	Memorize new vocabulary Understand "the second law of motion" and "normal force" Use (apply) new vocabulary
	HOTS	Find (analyze) the examples of "the second law of motion" and "normal force" in daily life Explain (create) the examples
Community		Pair and Group work (matching words and discussion)

The CLIL physics class was conducted as a part of five classes of the author's Integrated English IV in 2017 and 2018. In this university, Integrated English III and IV are required for second-year students. Students are appointed to different levels—S (the highest), A, and B (the lowest)—which are based on the scores from the Visualizing English Language Competency Test (VELC test). The students enrolled in the author's classes were from Biosphere-Geosphere Science, Information Science, Engineering, and Science. The engineering students were from the departments of Mechanical Engineering, Information and Computer Engineering, and Architecture. The science students were from the departments of Applied Mathematics, Chemistry, and Life Science. Students from the faculty of Engineering and departments of Applied Mathematics and Chemistry seemed to be familiar with the content, while the others showed some anxiety when the class content was announced. The number of students enrolled in each class is shown in Table 2. The class size of Biosphere and Information Science in 2018 was much larger than the average-English classes in this university. There were more repeaters in addition to the large number of first-time enrollments.

Table 2 Number of Students Enrolled in Each Class

	Students' Major	Year		Total	Level
		2017	2018		
Class	Biosphere / IS	41	55	96	B
	Engineering	34	38	72	B
	Science	35	0	35	A
Total		110	93	203	

Since this class was a part of the CLIL project, all faculty members were allowed to observe during the first week.

4. Survey

To examine the effects of the intervention given, the author conducted a questionnaire survey at the end of the third week. When conducting the survey, the author explained to students that this survey is anonymous and that the result will not affect their grade. The number of participants is shown in Table 3.

Table 3 Number of Participants

	Students' Major	Year		Total
		2017	2018	
Class	Biosphere / IS	26	41	67
	Engineering	25	29	54
	Science	24	0	24
Total		75	70	145

The research questions were as follows: 1) What do students perceive they have learned through CLIL Physics class? 2) What causes the difference in students' reactions? 3) What are the problems encountered in practicing CLIL?

The questionnaire was developed in consultation with the project team and was written in Japanese. The questionnaire consists of four parts: 1) English learning / Attitudes towards learning English (10 items, 5-point scale), 2) Content learning / Attitudes towards learning content (10 items, 5-point scale), 3) Attitudes towards the teaching style (5 items, 5-point scale), and 4) Open-ended questions about the perception of students on what they learned (5 items).

To analyze the quantitative data gathered in the first three parts, IBM SPSS Statistics 25 was used, whereas MAXQDA Analytic Pro 2018 was used for the qualitative analysis of the open-ended questions.

5. Result

To answer the first research question, the descriptive statistics for all items were collected. Table 4 shows the items, mean scores, and statistical deviations. The items with an asterisk (*) are negative statements. The item with the highest mean was "I couldn't explain well in English." The next highest was "It may increase my ability to think." Among the items related to English learning, "I learned the meaning of new vocabulary" and "My English vocabulary increased" showed higher scores. With regard to content learning, none of the items showed scores as high as English learning. Among them, "I could understand the theme more deeply," and "I could have my own opinions and perspectives" were higher. All items related to attitudes towards the teaching style scored rather high compared to other questions.

Table 4 Items, Means, and Standard Deviations (N = 145)

		Item	M	SD
English learning / Attitudes towards learning English	E1	My English vocabulary Increased.	3.51	0.83
	E2	I learned how to make conversation in English.	3.17	0.80
	E3	I learned the meaning of new vocabulary.	3.56	0.87
	E4	My English comprehension improved.	3.05	0.82
	E5	I couldn't explain well in English.*	3.72	0.87
	E6	My listening comprehension improved.	3.06	0.80
	E7	It didn't affect my English skills.*	2.61	0.84
	E8	I want to tackle English learning tasks with more difficult content.	2.74	0.99
	E9	I got interested in English learning.	2.98	0.91
	E10	I got more anxious toward learning English.*	3.01	0.93
Content learning / Attitudes towards learning content	C1	I could understand the theme more deeply.	3.46	0.82
	C2	I could understand the content.	3.17	0.87
	C3	I can apply what I learned to my future study.	3.16	0.89
	C4	I didn't know what I was studying.*	2.76	1.03
	C5	I'm not sure if I understood.*	2.94	1.03
	C6	I could have my own opinions and perspectives.	3.21	0.78
	C7	I learned a new way to study.	2.97	0.81
	C8	I want to study more about this subject.	2.84	0.84
	C9	I learned the major points of this subject.	2.97	0.84
	C10	I don't want to study this subject any more.*	2.81	0.94
Attitudes towards the teaching style	T1	The handouts were easy to understand.	3.41	0.75
	T2	The content was interesting.	3.17	0.90
	T3	I enjoyed pair work and group work.	3.16	1.11
	T4	Pictures and visual aids were good.	3.24	0.83
	T5	It may increase my ability to think.	3.61	0.83

With the expectation that engineering students are more familiar with physics and would find the class more interesting, an analysis of variance overview (ANOVA) based on the students' major was conducted. Table 5 shows the means, standard deviations, and ANOVA results. The numbers in parentheses are standard deviations. The results show that engineering students mostly scored in the middle of the three classes. There were some items showing significant differences among classes. Among the English learning items, "I learned how to make conversation in English" and "My listening comprehension improved" showed significant differences between science and Biosphere / IS students. For

content learning, science students scored significantly higher than the other majors for the item "I didn't know what I was studying." There were significant differences between the science students and other majors in three teaching style items: "The handouts were easy to understand," "I enjoyed pair work and group work," and "It may increase my ability to think." Also, Biosphere / IS students scored significantly lower than the others in "Pictures and visual aids were good."

Table 5 Means, Standard Deviations, and ANOVA Based on Major

		Bios/IS	Engineer	Science	F	p	
	N	67	54	24			
English learning / Attitudes towards learning English	E1	3.55 (0.91)	3.37 (0.81)	3.71 (0.62)	1.53	.220	
	E2	3.01 (0.77)	3.20 (0.83)	3.54 (0.72)	4.04	.020	S>B
	E3	3.55 (0.93)	3.56 (0.88)	3.58 (0.72)	0.01	.988	
	E4	2.94 (0.74)	3.07 (0.87)	3.29 (0.91)	1.68	.189	
	E5	3.76 (0.91)	3.70 (0.86)	3.67 (0.82)	0.13	.881	
	E6	2.94 (0.80)	3.02 (0.79)	3.46 (0.72)	3.98	.021	S>B
	E7	2.51 (0.82)	2.78 (0.77)	2.54 (0.98)	1.69	.188	
	E8	2.76 (1.02)	2.61 (0.96)	3.00 (0.98)	1.30	.276	
	E9	3.03 (0.90)	2.80 (0.98)	3.25 (0.68)	2.30	.104	
	E10	3.09 (0.95)	2.96 (0.91)	2.92 (0.93)	0.43	.650	
Content learning / Attitudes towards learning content	C1	3.42 (0.80)	3.41 (0.86)	3.67 (0.76)	0.97	.383	
	C2	3.03 (0.95)	3.24 (0.80)	3.42 (0.72)	2.05	.133	
	C3	3.16 (0.93)	3.15 (0.90)	3.17 (0.76)	0.01	.994	
	C4	2.88 (1.08)	2.89 (0.99)	2.13 (0.74)	5.75	.004	S<B,E
	C5	3.01 (1.09)	2.98 (0.96)	2.63 (0.97)	1.35	.262	
	C6	3.13 (0.76)	3.15 (0.81)	3.57 (0.73)	2.92	.057	
	C7	2.88 (0.79)	2.93 (0.84)	3.29 (0.75)	2.42	.093	
	C8	2.87 (0.80)	2.80 (0.92)	2.88 (0.80)	0.12	.884	
	C9	2.87 (0.85)	3.02 (0.88)	3.17 (0.70)	1.27	.285	
	C10	2.88 (0.91)	2.83 (0.97)	2.54 (0.98)	1.17	.312	
Attitudes towards the teaching style	T1	3.31 (0.72)	3.35 (0.73)	3.83 (0.76)	4.77	.010	S>B,E
	T2	3.01 (0.88)	3.28 (0.81)	3.33 (1.09)	1.80	.168	
	T3	2.97 (1.09)	3.09 (1.05)	3.83 (1.09)	5.87	.004	S>B,E
	T4	2.93 (0.88)	3.39 (0.63)	3.79 (0.72)	12.88	.000	B<E,S
	T5	3.49 (0.79)	3.54 (0.82)	4.08 (0.83)	5.08	.007	S>B,E

The results above did not show clearly significant differences among the students' specializations. Nonetheless, it suggested that there might be significant differences among the students' English proficiency levels. To examine possible differences, ANOVA was conducted based on the class level. Table 6 shows the result. Level A scored significantly higher than level B on "I learned how to make conversations in English" and "My listening comprehension improved." For content learning, "I didn't know what I was studying," "I could have my own opinions and perspectives," and "I learned a new way to study" showed significant differences between level A and level B. For teaching style, except for "The content was interesting," all items showed significant differences according to the English proficiency level.

Table 6 Means, Standard Deviations, and ANOVA Based on English Level

		A	B	F	p
	N	24	121		
English learning	E1	3.71(0.62)	3.47(0.87)	1.63	.204
	E2	3.54(0.72)	3.10(0.80)	6.32	.013
	E3	3.58(0.72)	3.55(0.90)	0.02	.880
	E4	3.29(0.91)	3.00(0.80)	2.57	.111
	E5	3.67(0.82)	3.74(0.88)	0.12	.724
	E6	3.46(0.72)	2.98(0.79)	7.69	.006
	E7	2.54(0.98)	2.63(0.81)	0.21	.645
	E8	3.00(0.98)	2.69(0.99)	1.92	.168
	E9	3.25(0.68)	2.93(0.94)	2.58	.110
	E10	2.92(0.93)	3.03(0.93)	0.31	.576
Content learning	C1	3.67(0.76)	3.41(0.82)	1.94	.165
	C2	3.42(0.72)	3.12(0.89)	2.29	.132
	C3	3.17(0.76)	3.16(0.91)	0.00	.961
	C4	2.13(0.74)	2.88(1.04)	11.58	.001
	C5	2.63(0.97)	3.00(1.03)	2.69	.103
	C6	3.57(0.73)	3.14(0.78)	5.87	.017
	C7	3.29(0.75)	2.90(0.81)	4.77	.031
	C8	2.88(0.80)	2.83(0.85)	0.05	.831
	C9	3.17(0.70)	2.93(0.86)	1.54	.217
	C10	2.54(0.98)	2.86(0.93)	2.29	.133
Teaching	T1	3.83(0.76)	3.33(0.72)	9.51	.002
	T2	3.33(1.09)	3.13(0.86)	1.01	.318
	T3	3.83(1.09)	3.02(1.07)	11.40	.001
	T4	3.79(0.72)	3.13(0.81)	13.86	.000
	T5	4.08(0.83)	3.51(0.80)	10.14	.002

For the open-ended questions, the author coded each comment and counted how frequently it occurred. Answers to four questions were coded. The questions were as follows: 1) What was the most interesting activity for you? 2) What was the most difficult activity for you? 3) What do you think you have learned through CLIL Physics class? 4) What skills do you think you could acquire through CLIL Physics class? Table 7 shows the codes with the most recurrent comments and their examples for each question. As the most interesting activity, many students answered "cooperation in group." Answers for the most difficult activity were "making English explanations," "cooperation in group," and "solving problems." There were also comments about physics content as the most difficult activity. The most selected answers for what students thought they learned were: "physics,"

"explaining physics in English," "English explanation skills," and "vocabulary." As for the skills they acquired, students answered "communication," "physics," "reading comprehension," and "explain physics in English."

Table 7 Code of Students' Answers for Open-Ended Questions

	Code	Example	# of coding
The most interesting activity	Cooperation in group	Working in groups / Thinking in teams	54
	Solving physics problems	Calculating / Solving problems	6
	Vocabulary activity	Vocabulary games	6
	Presentation	Presentation because we can learn many different opinions	3
The most difficult activity	Making English explanations	To explain answers for questions in English / It was difficult to explain in English	35
	Cooperation in group	To discuss in a group and solve problems / To communicate my opinion to members	27
	Solving problems		22
	Presentation	Preparing for presentation	7
	Physics	The concept of physics	6
What students believe they learned	Physics		17
	Explain Physics in English	How to explain physics in English	17
	English explanation skill	I can express my idea in simple English and can communicate well.	14
	Vocabulary	Physics and mathematics vocabulary in English	12
	Communication	Discussions with other people.	6
	General English skills		6
Skills Students believe they acquired	Communication	Communication skills / Skills to discuss physics with others	12
	Physics		11
	Reading comprehension	Skill to grasp the meaning of English written text	11
	Explain physics in English	How to explain physics in English	10
	English explanation skill	Skill to explain my own idea in English	9
	Vocabulary		8

6. Discussion

The descriptive statistics of all the participants showed that "I couldn't explain well in English" had the highest score. The qualitative data also showed that most students commented that preparing English explanations was the most difficult activity. Therefore, these results may suggest that students found their weakness to be in expressing opinions in English. Moreover, two items related to vocabulary showed higher scores among items related to English learning. Since the two

items were similar, students may have interpreted them as being the same question. However, several comments about vocabulary, especially technical terms, were also given as something students perceived they had learned or skills they had acquired. Thus, the result may suggest that students perceived that they had learned technical terms. For content learning, all items did not show as high scores as English learning. Among them, two items involving understanding the content scored higher. Like the two items related to vocabulary in English learning, those two items may have been interpreted as being the same question. However, the qualitative data also suggest that several students perceived that they had learned physics or how to explain physics in English through the course. The results may show some effectiveness of the class in learning physics content. Since all items related to teaching style received higher scores than other questions, students may have enjoyed the class, although some of them were not familiar with physics knowledge. The qualitative data also showed that many students answered that group work was the most interesting activity. Moreover, gaining thinking skills scored the second highest of all items. This result showed students' perception of improving their thinking skills. Based on these results, the answers to the first research question "what do students perceive they've learned through CLIL Physics class?" are that students perceived that they had gained great ability in thinking (cognition) and that they had also learned both vocabulary and content. Since cognition is a very important aspect of CLIL, this study showed the possibility of achieving the aim of CLIL. However, students still experienced difficulty explaining in English, and thus it may take time for them to gain confidence in English speaking.

The ANOVA based on the students' major showed that engineering students, who seem to be more familiar with physics, scored in the middle of three classes in most items. Moreover, there were significant differences between science students and the other students with regard to several items. Considering the fact that students from the Life Science department showed anxiety in learning physics, the result was rather a surprise. At the same time, the author noticed a better atmosphere in the science class when students worked together. When a group member could not understand the concept of dynamics, the other members tried to explain with gestures and drawings, as well as in English and Japanese. There were similar interactions in other classes, but peer help was seen more often in the science class. Therefore, the significant difference between science and Biosphere / IS students in "I learned how to make conversation in English" and "My listening comprehension improved" may have been influenced by the classroom atmosphere. For visual aids, Biosphere and Information Science students scored significantly lower than the students in other majors. This particular result may be due to the size of the class and classroom. The projector in the classroom also had low resolution. The ANOVA based on English proficiency level showed significant differences between levels A and B. These results addressed the second research question "What causes the difference in students' reactions?" The students' English

proficiency level caused more noticeable differences than the students' major. Furthermore, the items showing significant differences indicated that students with higher English proficiency enjoyed group work more and interposed that the CLIL class cultivated their higher order thinking skills. This may mean that students with higher English proficiency could achieve the goals of CLIL more effectively.

The comments for the open-ended questions revealed that many students enjoyed group activity even though it was also difficult for some of them. Although it was an English class, students stated that they learned physics rather than English. This may indicate that the class content was not familiar for many students even though it was fundamental content. The fact that some students rated physics content as the most difficult may indicate that some students struggled primarily because of the content. In addition, many of them felt that they acquired communication skills and English explanation skills, which shows the effectiveness of the class in fostering important social skills. In view of the comments and results above, the third research question, "What are problems of practicing CLIL Physics?" may be addressed as the students enjoying group activities, although it may still be difficult for some of them. If the students are not interested in either the content or English, a CLIL class may actually demotivate students. Therefore, the instructor should choose appropriate and interesting content for students and have enough knowledge of both content and English.

7.   Conclusion

This study discussed the practice of CLIL physics for university students and how the students reacted to the class. During the classes, the author noticed a large number of students struggled with their social skills such as communicating with others or collaborating in groups. The tendencies are especially conspicuous among those who are not proficient in English. The survey results, however, suggested the possible favorable effects of CLIL physics even though it was only a three-week course. It revealed that students, especially with higher English proficiency, enjoy group work and acquire communication skills through this practice. The results also suggested that students are able to learn content knowledge even through a language teacher. Moreover, students perceived that they improved their cognition, which is the central aspect of CLIL. There is also a big demand for cognitive abilities in Japanese educational settings. CLIL, therefore, is applicable to train students' cognitive abilities and to help them link their content knowledge and English.

There are, however, some limitations in this study. Since this course was a three-week course, it may not show sufficient effect of CLIL classes. Second, the survey was conducted only after the course; so it does not show the effects clearly; it may therefore be better to check students' content knowledge beforehand. The author would like to design a year-long or a semester-long CLIL course and see the effect with pre-post surveys. To design effective CLIL classes for university students, instructors can choose fundamental content knowledge and easy vocabulary related to the content.

If the class facilitates a lot of discussion and presentation, it could cultivate the students' research skills and higher order thinking skills.

For future research, the author would also like to study how learning occurs in group cooperation as well as the problems of designing a CLIL class from the instructor's viewpoint.

References

Anthony, L. (2009). ESP at the Center of Program Design. In K. Fukui, J. Noguchi, & N. Watanabe (Eds.), *ESP-teki bilingual o mezashite: Daigaku eigo kyoiku no saiteigi [Towards ESP bilingualism: Redefining university English education in Japan]* (pp. 18–35). Osaka: Osaka Daigaku Shuppan-kai.

Anthony, L., Noguchi, J., & Orr, T. (1998). *Applying ESP to reading,writing,listening and speaking.* Paper presented at the The JACET 37th Annual Convention, Shujitsu Women's University, Okayama.

Bradford, A., & Brown, H. (2018). ROAD-MAPPING English-Medium Instruction in Japan. In A. Bradford & H. Brown (Eds.), *English-Medium Instruction in Japanese Higher Education: Policy, Challenges and Outcomes* (pp. 3–13). Bristol: Multilingual Matters.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and Language Integrated Learning*. Cambridge: Cambridge University Press.

Gally, T. (2009). Specific in general: Scientific writing and presentation within a liberal arts curriculum. In K. Fukui, J. Noguchi, & N. Watanabe (Eds.), *ESP-teki bilingual o mezashite: Daigaku eigo kyoiku no saiteigi [Towards ESP bilingualism: Redefining university English education in Japan]* (pp. 120–129). Osaka: Osaka Daigaku Shuppan Kai.

Hanzawa, K. (2017). CBI/CLIL/EMI no sai-teigi (Redefining CBI/CLIL/EMI). In I. f. A. S. i. E. Waseda University (Ed.), *Eigo de kyoka naiyo ya senmon o manabu (Learning Content and Specialized Fied in English)* (pp. 31–45). Tokyo: Gakubunsha.

Kojima, N., & Yashima, T. (2017). Motivation in English Medium Instruction Classrooms from the Perspectives of Self-determiantion Theory and the Ideal Self. *JaCET journal, 61*, 23–40.

Maekawa, Y., Okunishi, Y., Maruyama, I., O'Halloran, J., Takashima, K., Hamaya, Y., & Yanagi, K. (2017). Academic Eigo Noryoku Syutoku o Kano ni suru Naiyo-Gengo-Togogata-Kyoiku no Jissen: 4tsu no C ni Motoduita CLIL Curriculum no Kochiku (Practice of CLIL (Content and Language Integrated Learning) Instruction to Acquire Academic Language Skills: Constructing CLIL Curriculum Based on 4Cs) *Okayama University of Science Educational Practice Research, 1*, 155–164.

Ministry of Education, Culture, Sports, Science and Technology. (2003). Regarding the Establishment of an Action Plan to Cultivate"Japanese with English Abilities". Retrieved from http://www.mext.go.jp/b_menu/houdou/15/03/03033101/001.pdf

Ministry of Education, Culture, Sports, Science and Technology. (2012a). Project for promotion of global human resource development. Retrieved from http://www.mext.go.jp/english/highered/1326713.htm

Ministry of Education, Culture, Sports, Science and Technology. (2012b). Towards a Qualitative Transformation of University Education for Building a New Future - Universities Fostering Lifelong Learning and the Ability to Think Independently and Proactively (Report). Retrieved from https://www.mext.go.jp/en/publication/report/title01/detail01/1380275.htm

Miyama, A. (2007). Genre bunseki ni motozuita ESP approach no jissen [Practice of ESP approach based on genre analysis]. *Current English Studies, 2007*(46), 1–15. doi:10.11293/jaces1962.2007.46_1

Miyama, A., & Nitta, K. (2003). Riko-kei gakubu ni okeru ESP Coursebook no Bunseki to ESP Kyozai Kaihatsu [Analyising of ESP Coursebooks and Developing ESP Teaching Material in Engineering Department]. In S. Sasajima (Ed.), *Developing Practical Teaching Methods in ESP Based on a Survey of Teaching   English in Tertiary Education* (pp. 40–57): Report of Grants-in-Aid for Scientific Research (Research number 13680326).

Morton, T., & Llinares, A. (2017). Content and Language Integrated Learning (CLIL): Type of programme or pedagogical model? In A. Linares & T. Morton (Eds.), *Applied Linguistics Perspectives on CLIL* (pp. 1–16). Amsterdam: John Benjamins Publishing Company.

Noguchi, J. (2010). What can ESP do ? *Journal of JSEE, 58*(3), 3_9-3_11. doi:10.4307/jsee.58.3_9

Tsuda, A. (2006). Career kaihatsu no tame no daigaku eigo: Digaku hisshu eigo ni okeru jugyo-rei kara [College English for career development: Lesson plans on the English core course]. *ESP no Kenkyu to Jissen [ESP Research and Practice], 5*, 34–43.

Watanabe, Y., Ikeda, M., & Izumi, S. (2011). *CLIL(Content and Language Integrated Learning) New Challenges in Foreign Language Education at Sophia University Volume1: Principles and Methodologies*. Tokyo: Sophia University Press.

Yamauchi, H. (2005). Kogaku-kei ESP no tame no syllabus to kyozai kaihatsu ni muketa needs bunseki; Kogaku-kei (Kurume Kogyo Daigaku) no ba-ai [Needs analysis for developing engineering ESP syllabus and teaching material: Case of engineering (Kurume Institute of Technology)] *ESP no Kenkyu to Jissen [ESP Research and Practice], 4*, 70–90.

Taking Charge of Student Feedback:
Using Narrative Frames for Course Evaluation

Sarah Mason* and Kaori Sakakibara**

Mejiro University
mason@mejiro.ac.jp
***Mejiro University*
k.sakakibara@mejiro.ac.jp

Abstract

The administration of student questionnaires and the implementation of feedback from such questionnaires have been shown to improve teacher efficacy (Boice, 2000). The acquisition of feedback from students via quantitative questionnaires is common practice in many countries, but empirical studies on evaluations are insubstantial and alternatives are inadequately researched (Hiratsuka, 2018). In an effort to address this gap this paper describes the use of an innovative tool for gathering feedback termed a 'narrative frame' (Barkuizen & Wette, 2008) was implemented by the researchers. This paper describes narrative frames and presents findings concerning the feedback. The study concludes that narrative frames are an effective method of eliciting feedback from English language students in a writing course, and furthermore, that narrative analysis complements thematic analysis as an analytical method in this type of qualitative study.

Keywords: narrative frame, TESOL, qualitative research, data collection

1. Introduction

Research has shown course evaluations to improve teacher efficacy (Boice, 2000), but, as Hiratsuka (2018) writes, "empirical studies on the topic are not substantial enough, nor are alternatives adequately considered". Course evaluations are generally controlled by school administrative departments and tend to be administered at the conclusion of courses. Their purpose is to confirm teaching quality and course value. Although they may include optional open-ended questions, they tend to be largely quantitative in nature, composed of Likert scales with feedback communicated to teachers in the form of radar charts and the like. This type of feedback facilitates comparisons between individual teacher performance and institutional averages, but it does not provide detailed and specific information concerning classroom activities, course topics, assignments, or challenges encountered by students during the course. They also do not convey highly personal, individual stories of student experience. The present study is an attempt to explore narrative frames which are an alternative to traditional, quantitative style course evaluations. The purpose of narrative frame evaluations, in contrast to the more traditional style evaluation, is to gather detailed data which can then be used by the teacher involved to revise curriculum content and teaching. This paper describes the use of narrative frames to facilitate effective feedback from students at a tertiary institution in Japan. It also presents empirical findings concerning the data gathered and evaluates narrative frames as a tool for collecting data. It answers the question: What is the effectiveness of narrative frames as a course evaluation instrument in a first-year English writing class for English language majors at a Japanese university?

Narrative frames are essentially story templates in the form of incomplete paragraphs. They consist of sentence starters and transition words. Respondents (in this case English language learners) are provided with a frame and are asked to complete it in their own words in order to compose a story of their experience (see appendix A for the frame used in the present study).

Not all feedback obtained from student questionnaires is comprehensible or helpful even when written in their native language, and the students involved had a low English level which was a cause of concern, therefore a narrative frame was selected as the most appropriate data collection tool. It was hoped that due to its supportive structure the narrative frame would enable the students to produce comprehensible narrative-writing in English, writing that would also function to effectively evaluate the course and report the students' experiences.

Whilst traditional quantitative course evaluation instruments seek to evaluate teachers' competence in comparison with other teachers and meet the needs of various stakeholders such as parents, administrators and employers, it was our intention to gather rich data that would illuminate the students' classroom experiences and furthermore help us to shape the course moving into the following academic year. The frame used was therefore designed to gather data pertaining to various issues, including reflections on the curriculum, class activities, assignments, homework, suggestions for change, learning difficulties and students' feelings.

2. Literature Review

Narrative frames as a data collection tool in the field of teaching English to speakers of other languages (TESOL) were first developed by Barkhuizen & Wette (2008) who argued that the frames had particular advantages when used in qualitative research. Namely, narrative frames could provide scaffolding to beginner English language learners who might otherwise not have the necessary skills to clearly express themselves in written form; the researchers could obtain detailed feedback in contrast to the "snapshot" style of feedback obtained through quantitative evaluations; the data collected through the frame would be limited to the information requirements of the person administering it; and finally, timing and design could be controlled by the teacher. Since Barkhuizen and Wette (2008) research using narrative frames has rapidly expanded. It has focused on four main areas, language teacher development (Barnard & Nguyen, 2010; Nguyen, 2017; Shelly, Murphy & White, 2013; Wette & Barkhuizen, 2009; Xu, 2014), needs analysis (Hiratsuka, 2016; Macalister, 2012), study abroad (Swenson & Visgatis, 2011; Ryan, Rabbidge, Wang, & Field, 2019), and course evaluation, (Hiratsuka, 2014, 2018). Research in all of these areas has been unanimous in its praise of narrative frames, concurring with Barkhuizen & Wette's (2008) summary of the advantages of the frames as a data collection tool in qualitative research.

Of particular relevance to the present study are studies by Swenson and Visgatis (2011) and Hiratsuka, (2018) which focus on English language learners in Japan. Swenson and Visgatis (2011) is a study of four university students' study abroad experiences. The researchers used the frames to help the students begin to write their own stories of the study abroad experiences. The students had a relatively high level of English ability, with the minimum required TOEIC score for participation in the internship program set at 610. Swenson and Visgatis reported that the use of the narrative frame was particularly effective at revealing *commonalities* of participant experience.

Hiratsuka (2018) is a study of the use narrative frames as a course evaluation instrument. The course was an undergraduate English teaching methods course held at a Japanese university. Hiratsuka collected data from 26 students. He concluded that use of the narrative frame, "facilitated the quality of their [the students] reflection and increased the reliability of the information upon which to evaluate the course" (p.6). He also reported that use of a narrative frame let to critical reflection on teaching and learning by the teacher involved.

Despite these positive observations concerning narrative frames, it is worth noting that they inherently involve some limitations. For example, some participants may be frustrated by the lack of opportunity to write about topics not included in the frame, and, if the frame is distributed on paper, by the physical limitation to the length of possible responses. (Barkhuizen & Wette, 2008). Moreover, Barkhuizen (2014) notes that depersonalization of individual participant experience is possible, "especially when combined and analyzed together with many other completed frames. The individuals and their stories get lost in the amalgamation of data and their reduction into themes and categories, and possibly even numbers (e.g. frequency counts and statistics, if a quantitative analysis is done)" (p. 14). A further limitation observed by Nguyen and Bygate (2012) is that, as with all self-report research, the truth value of what is reported by the participants is open to debate. In their study of 23 teachers working at three suburban schools in central Vietnam, they also note that due to the small size of the respondent group it is not possible to generalize beyond the particulars of their study, a feature one can assume is common to all small scale narrative frame studies.

3. Methodology

Nguyen and Bygate (2012) write about the truth-value of narrative frames as an instrument for data collection. They note that while trustworthiness cannot be guaranteed, because the truth value any self-report method of data collection is potentially questionable, nonetheless the procedure can make it easier for participants to write what they want to say while reducing threats to their 'face' when compared, for example, with interview style qualitative data collection. In his commentary Bygate notes the procedure is implicitly comparatively high in trustworthiness (Nguyen & Bygate, 2012, p. 58). Moreover, they observe that the inherent difficulty of establishing truth-value in qualitative research approaches may be overcome by consistently and rigorously carrying out data collection and analysis, and by transparently and honestly reporting the findings. In the present study we attempted to do so and accordingly outline our procedures for data collection and analysis in detail next.

Creswell (2008), in his seminal work on conducting qualitative research, writes that qualitative researchers "will not use someone else's instrument as in quantitative research" (p. 213), instead qualitative researchers collect data with a tool tailored to the research context and question. Therefore, the style and content of the narrative frame used in the present study differs somewhat from Hiratsuka (2018), an earlier study using narrative frames for course evaluation. The frame in the present study was composed with the specific contextual details of the writing course in mind. Barkhuizen (2014) reviews the usage of narrative frames in recent years and clarifies various essential design features of narrative frames: the purpose of the study should be to collect qualitative data, that the topic should be clearly defined at the outset, that the frame should have story elements (time, place, characters), and that the frames structure should require reflection from the participants on their personal experience in order to be completed. Furthermore, the frame should have coherence as a narrative and that it should be formatted as an incomplete paragraph. When composing the frame used in the present study these criteria were adhered to as closely as possible (refer to

Appendix A for the frame used in the study). The frame itself specifies the location for the story, the classroom. The temporal dimension is also provided with the story beginning at the start of the academic year in April and concluding in January in the final class of the course. The main character appearing in the story is "I" (the student completing the frame), although in some cases the student completing the frame also made references to the teacher or the other students.

The narrative frame was administered at the end of a year-long university EFL writing course taught at a university in Tokyo. The participants consisted of 35 first year English language major students. The class met twice a week for a total of 30 weeks over the course of a year. A communicative and process-oriented approach was taken to teaching the writing course. Students were given the opportunity to write drafts of their work, receive feedback (both peer review and direct feedback from the teacher) and revise it accordingly. Because of the communicative approach taken students also had many opportunities to read and listen to the work of other students, to share their writing and to discuss various topics covered in class. The classes generally began with free writing, followed by use of a textbook connected to the students' future study abroad experience (all second-year students have an opportunity to participate in a study abroad program in their fall semester). The second half of each class was devoted initially to paragraph writing skills and later to essay writing. Various genres (email, essay etc) and topics were covered in this second half of each lesson.

The language choice (English or the students' native language, Japanese) was an important consideration when designing the frame. The students participating in the present study were undertaking an English language writing course and therefore it was decided that a course evaluation written in English would be appropriate. A further consideration was that the students would be required to complete another (Japanese language) course evaluation administered by the university's academic affairs section at roughly the same time, and the researchers were wary of inducing student fatigue if they were to request the frames be completed in both languages.

At the time the narrative frame was administered the average TOEIC score of the students was 369 and it was therefore deemed necessary to prepare the students for the frame in order to maximise their ability to competently complete it. Therefore, one week prior to the administration of the frame the administering teacher reviewed the relevant grammar used in the frame. Also, at the start of the lesson in which the frame was administered the year's classroom activities and assignments were reviewed so that they would be fresh in the students' minds. Finally, the narrative frame was photocopied and distributed to the students who completed it in pencil over the course of forty minutes. Students were advised not to write their names on the frame and reassured that it was anonymous.

As with Swenson and Visgatis (2011) and Hiratsuka (2018) narrative frame studies generally involve qualitative analysis approaches in order to identify themes. Findings are frequently presented in table from with numbers representing frequency counts of the occurrence of each theme with the narrative frames. In the present study frames were analyzed both thematically and narratively. Narrative analysis was consciously undertaken in an effort to overcome the potential depersonalization of the narratives identified above by Barkhuizen (2014) as a potential limitation of narrative frames.

Thematic analysis involves repeated reading of the data to identify recurring themes (Bogdan & Biklen, 2007). Both authors read through all the individual frames in their entirety twice to get a sense of over-arching categories. Both authors listed emergent themes during the second reading. The researchers then met, agreed on two primary categories that the data fell into, and also agreed on a preliminary list of themes. Each frame was then coded, sentence by sentence, and each sentence was classified under the relevant thematic heading by the first author. At this stage new themes emerged and were added to the two main categories. Not all sentences were able to be classified at this point due to ambiguity. Following this process, the researchers met again. The second researcher checked the themes and the sentences classified under them. In some cases, after discussion took place, sentences were reclassified under different themes at this point new themes developed. Ambiguous sentences were then looked at in the context of the overall frame, discussed between researchers and either classified or discarded. (Two were discarded due to ambiguity). The number of sentences per theme was then calculated to arrive at a frequency count.

Although it had initially been assumed that the sentence starters would correspond to the themes occurring in the frames this was found not to be the case. For example, it was initially assumed that all sentences that began with the starter "Now in January the class is almost over and I feel …" would be thematically classified under the category "Students" and then under the theme "Emotions" (and then under the relevant sub-theme). However, one student wrote, "Now in January the class is almost over and I feel *teacher's lesson [is] easy to understand*". This sentence was therefore classified under the category "Course", next under the theme "Teacher" and then under the sub-theme Kind/effective/helpful. This type of unexpected classification took place numerous times and meant that the frequency count does not total 35 for many of the themes identified although there were 35 participants.

Furthermore, some students did not submit fully completed frames. In particular the second half of a sentence, designated by the number (7) was left blank by nine of the respondents, "(6) If I could take this class again, I would like more _____ but less (7) _____ ". This point is discussed further below.

Analysis of narratives is a sub-type of story analysis identified by Polkinghorne (1995, p.5-6) who describes studies using this type of analysis as, "Studies

whose data consist of narratives or stories, but whose analysis produces paradigmatic typologies or categories". It is a wholistic analytical approach that looks at plot trajectory of individual stories from beginning to end. In the present study the analysis of the narratives followed the following steps: each case was read individually from beginning to end, twice, by the first author. Plot development, in terms of the students' experiences, emotions and intentions (narrative structure) was identified in each story. Overarching categories (or 'paradigms' to use Polkinghorne's terminology) were identified and each of the 35 narratives gathered for the study was classified according to the paradigm followed. The second author then repeated the steps and verified classification of the individual narratives into the paradigms.

4. Findings: Thematic

Two interrelated categories were identified through repeated reading of the frames: the course and the students.

Table 1. The Course (n = 35)

Themes	Details	Freq.
Enjoyed particular essay topics	Self-introduction	1
	Dream Partner	1
	Photo	1
	Summer	1
	Stamp	3
	My Friend	2
	My Goals	2
	Role Models	4
	A Great Career	4
Negative	Disliked all topics	2
	Total	(21)
Activities positively perceived	Free writing	12
	Textbook	7
	Essay prep.	18
	Speaking	3
	Vocabulary	3
	Use "Word" (computer program)	1
	Total	(35)
Suggestions for change	More speaking	4
	More writing	4
	More reading (novels)	1
	More translation	1
	More textbook	3
	Total	(13)
	Less free writing	7
	Less quiet writing	1
	Less homework	3
	Less reading	1
	Less "nothing"	14
	Total	(26)
Teacher	Kind/effective/ helpful	5
	Hard to understand	1
	Total	(6)

Note. The total of frequencies rarely adds up to 35 because any one response could contain more than one detail or omit answers.

Every assignment undertaken during the course was nominated by at least one student as their favorite assignment, although two students specified that they did not enjoy any of the topics covered. Reasons for enjoying assignments: meaningful communication. "The writing assignment I most enjoyed was *writing about my role models* because *I was able to introduce my friend*."

Unsurprisingly, for many students (18) the move from high school to university was very challenging. This seemed to be because it necessitated transitioning from high school multiple choice or complete the blank type English study to composing their own sentences, using their imagination to express their own experiences, and moreover being taught in a foreign language. Representative of this type of response is the comment "I remember once in my writing class I had a hard time trying to *write an original essay by myself*." The student goes on to explain the cause was that "[it was] *difficult to write by considering*".

One striking instance of conflicting opinions concerned the 10 minutes of free writing used as a warm-up at the start of each lesson. Many students (7) nominated free writing as the type of activity they wanted less of, and moreover, in table 2, 15 students reported that they had found it difficult to do free writing due to their low level of vocabulary or skill. However, 12 wrote about it positively, for example, "The writing assignment I most enjoyed was *free writing* because *we did it every class, so we could write long sentence[s]*. Amongst these 12 students one student said they actually wanted more free writing and one said it was very useful.

It was also revealing that the necessity of acquiring computer skills (in particular the program "Word"), and even getting access to computers, had been a struggle for some students. Eleven students referred to either to computer skills, printing or Word as difficult. One student explained, "*I don't have a PC so I was very busy for make a report*". The teacher researcher was able to grant longer time to complete homework assignments the following year.

Many respondents, fourteen in total, either specifically used the word "nothing" when asked what they wanted less of or left the space blank.

The frame did not invite students to comment upon the teacher, but perhaps it is inevitable that in this type of situation some (6) students did so. It was reassuring that the comments were generally positive.

Table 2. The Students (n = 35)

Themes	Details	Freq.
Emotions in April	Hard	15
	Word	11
	Excitement	6
	Total	(32)
Emotions in January	Regret	11
	Accomplishment	4
	Difficult/Negative	4
	Sad to end	3
	Fun	2

	Future orientation	2
	Total	(34)
Poor self-evaluations for class & homework	Due to difference between high school study and college	18
	Due to poor fundamentals (vocab., grammar, speaking etc.)	15
	Due to lack of technology skills	5
	Due to lack of topic familiarity	1
	Total	(39)
Positive self-evaluation re class & homework	Developed new skills	1
	Homework a "good fit"	2
	Total	(3)
General reflections	Gratitude	2
	Realized importance of English	6
		(8)
Future academic worries	Composition	11
	Vocabulary	9
	Speaking/Listening	9
	Study abroad	2
	Total	(31)

Note. The total of frequencies does not add up to 35 because any one response could contain more than one detail or omit answers.

Difficult class activities and newness of class activities also affected the emotions of the students. There appeared to be some shift in emotions of the students over the course of the academic year. In April many students reported the feeling the course was hard or difficult. One student wrote, "In April, at the start of this class, I felt [it was the] *most difficult in the Core Program*". By January somewhat different emotions were reported, and regret is noticeably reported by 11 students at this stage, for example, "If I could take this class again, I would like more *I raise my hand [volunteer], but less I sleep and be absent. I'm sorry:-(".*

As noted above, the frames elicited far more critical self-evaluations from students that positive one's concerning their homework and class performance. Interrelated were reports of struggles with transitioning to university style composition and challenges with writing sentences and longer pieces by themselves.

General reflections concerning their appreciation for the course had not been anticipated when the frame was first designed, however 8 students chose to complete the final sentence of the frame in this manner. One representative comment is, "Now In January, the class is almost over and I feel [I] *often learned in this lesson.*"

Many students identified writing correct grammar or poor vocabulary as future problems they would need to overcome by various means, such as study, solving problems in the textbook, free writing or, in one case "I have to act more actively in class". Unexpectedly, quite a few students (9) chose to either delete or ignore the word 'writing' from the frame and wrote about other skills they

were more concerned about. For example, one student wrote, "A future difficulty I might have with improving my ~~writing~~ skill is *communicate the words well*. However, I might overcome this by *communication with a lot of foreigners*".

Findings: Narrative
Two paradigmatic typologies were identified, first stories that had flat emotional and academic trajectories in which the emotional experiences of the students remained static throughout the course, whether they were positive or otherwise. The second paradigm was transformational in nature, in the sense that the students' emotions, attitudes and intentions changed over the course of the year as reported in their narratives.

Table 3 Narrative Paradigms (n=35)

Paradigm	Details	Freq.
Transformation	Achievement	22
	Resolve to change	7
	Total	29
Static	Endless misery	3
	Everything is always wonderful	3
	Total	6
		(35)

As indicated in table 3 the most common narrative paradigm was transformational. Either the students engaged in an upward trajectory of academic learning and/or adaptation to the communicative classroom (22) or their experiences included personal transformation due to growing self-awareness which lead to resolve to do better (7). However, an alternative, static, paradigm co-existed in the classroom. The narrative of these students included persistent attitudes, either positive or negative, and those students finished the class in much the same way as they had started (6).

For illustrative purposes we include the following frame:
Student A
(1) In April, at the start of this class, I felt [it was] *very difficult for me to listen to native speaker's English*. (2) I remember once in my writing class I had a hard time trying *to think about the contents for free writing*. (3) The main reason for this was that *I don't remember many words' spelling.* (4) The writing assignment I enjoyed most was *writing about my seal* (5) because *I love to draw pictures, so I didn't feel hard study.* (6) If I could take this class again, I would like more *to read famous story ex. Harry Potter* but less *nothing*. (7) The most useful thing I did in this class was *report writing in Word*. This is because *it is important for me to write about my idea*. (9) It was ~~easy~~/difficult to do the homework because *I must think reports contents and English grammar by myself.* (10) A future difficulty I might have with improving my writing skill is *speaking, reading and listening because I will go to study abroad*. (11) However, I ~~might~~/will

overcome this difficulty by *read an English book, speak English with a native speaker and listening to anything. For example, movie with subtitle.* (12) Now, in January, the class is almost over, and I feel *English is very difficult, but I want to speak English more!*

5. Discussion

A comparison of narrative frames with a more traditional "survey" tool for data collection leads us to the following conclusions:

Since the students' answers take the form of a story with narrative frames, it is easier to see the relationship between questions than it would be with a series of multiple-choice questions. For example, Student A in the example above wrote that at the beginning of the school year he had anxiety about his ability to understand the native-English-speaking teacher's instructions. He made efforts, however, to overcome his problem and even committed to his own method of studying to catch up with the class, writing that he would read books and watch movies with subtitles. Finally, he added, he ended the year highly motivated to study harder and wanting "to speak English more!". Thus, the narrative frame questionnaire has helped the teacher in charge of the class visualize the process of her students' growth by linking the dots. Through narrative analysis we could sense that often, although difficulties were present in the students' experiences, they were perceived as challenges, rather than being completely negative.

Another benefit observed in the narrative frame questionnaire is that it succeeded in yielding unexpected answers—ones even the student himself would not have been able to come up with if he had been presented with multiple choices. For instance, several students said the main reason for their problem was that they did not own a computer. One of the authors had not expected this to be any major source of inconvenience; it would certainly not have been one of the options if a multiple-choice question had been prepared. As shown above, the latitude given by the narrative frame format makes it possible to obtain answers closer to the learners' true sentiments.

Nevertheless, some of the blanks failed to yield any responses from some students, including (7) in the following sentence, "(6) If I could take this class again I would like more _____ (7) but less _____." Those students seemed to have difficulty either understanding what kinds of answers were expected of them or coming up with words and phrases they felt could express their feelings. If these parts had been replaced with multiple-choice questions, there is a good chance that the students would have circled some options close to their feelings. While the answers thus obtained may have been less precise, they would have provided the teacher with at least some information about the students. An alternative explanation is that those who left the space blank had nothing in particular they wanted less of, but there remains some uncertainly concerning this point. We suggest that mixed methods research, involving both narrative frames and interviews, could circumvent this problem.

Narrative frames, as a tool for collecting data, did have some disadvantages. Despite the sentence starters, the imaginative manner in which the students completed the sentences meant that sentences could not be easily categorized into simple themes based on the sentence starters.

When compared with a more traditional quantitative evaluation tool the narrative frame approach has significant drawbacks in terms of preparation time and time needed for analysis. Whereas a more traditional survey may be completed in 5 to 10 minutes at the end of a class, designing the frame, preparing the students to complete the frame, and then allowing time in class for completion meant that the process took hours rather than minutes. We accept that the decision to complete the frame in the students' second language clearly contributed to this extended process, but we believe that the process itself was of educational benefit to the students, who were after all, enrolled in a writing course. Analysis of the data also took hours, a common feature of qualitative research. Finally, the anonymous nature of the collection method in the present study meant that the teacher could not support students to complete the frame during the class.

The advantages found by other researchers using narrative frames were largely also experienced in the present study. The frames were fully controlled (designed and administered) by the teacher, it was possible to obtain detailed feedback, the feedback was limited to the researchers' requirements, and beginner students were supported to write comprehensibly in an unfamiliar genre in a second language. However, we feel the frames also had one further advantage that has not been emphasized in the literature to date. The frames easily lent themselves to a narrative analytical approach. The narrative nature of the data allowed us to perceive transformation of student attitudes, beliefs and self-perceived skills over time. We were struck by the finding that 29 respondents produced transformational narratives, indicating either academic growth, or growth in self-awareness. Furthermore, close study of the data heightened teacher researcher awareness of student emotions and struggles, and that lead to greater empathy on our part for the students. As with Hiratsuka (2018) we find that the use of narrative frames led to increased reflectivity on the part of the teacher involved.

6. Conclusion

This paper has explored the effectiveness of narrative frames as a course evaluation instrument in a first-year English writing class for English language majors at a Japanese university. Based on the data gathered through the frames the teacher researcher involved in the writing class gained confidence that the variety of writing tasks appealed to the students involved and that the students considered them helpful. Significant insights into the practical difficulties experienced by first-year students transitioning to a university learning environment, and

their complex emotional experiences were also gained. We conclude that narrative frames are an effective method of eliciting feedback from English language students in a writing course and Moreover, that narrative analysis complements thematic analysis as an analytical method in this type of qualitative study.

To conclude, we agree with Hiratsuka (2014) that narrative frames, "could be used as an alternative to students' evaluation sheets for teachers to understand their students and classes more fully" (p. 176). Therefore, other teachers may wish to utilize this method of data collection in similar courses. However, a word of caution is necessary. Anonymity should be abandoned if the frame will be completed in the second language and student level is very low. This would allow students requiring further language support to freely seek help during class time, which would lead to more comprehensible feedback.

References

Barkhuizen, G. (2014). Revisiting narrative frames: An instrument for investigating language teaching and learning. *System, 47*, 12−27.

Barkhuizen, G., & Wette, R. (2008). Narrative frames for investigating the experiences of language teachers. *System, 36*(3), 372−387.

Barnard, R. & Nguyen, G. V. (2010). Task-based language teaching (TBLT): A Vietnamese case study using narrative frames to elicit teacher's beliefs. *Language Education in Asia, 1*, 77−86.

Bodgan, R. C., & Biklen, S. K. (2007). *Qualitative research for education: An introduction to theory and methods* (5th ed.). Boston: Pearson Education.

Boice, R. (2000). Advice for New Faculty Members: Nihil Nimus. Needham Heights, MA: Allyn and Bacon.

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review, 41*, 71−88.

Creswell, J. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (5th ed.). Frenchs Forest, Sydney, NSW: Pearson Australia.

Hiratsuka, T. (2014). A Study into How High School Students Learn Using Narrative Frames. *ELT Journal, 68*(2), 169−178.

Hiratsuka, T. (2016). Employing narrative frames for needs analysis: The case of a newly-hired teacher. *OTB Forum, 7*(2), 57−63.

Hiratsuka, T. (2018). Narrative Frames as a Course Evaluation Instrument. *The Language Teacher, 42*(1), 3−7.

Kember, D., Leung, D. Y., & Kwan, K. P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education, 27*(5), 411−425.

Macalister, J. (2012). Narrative frames and needs analysis. *System, 40*, 120−128.

Nguyen, G. V. & Bygate, M. (2012). Narrative frames. In R. Barnard and A. Burns (Eds.) *Researching language teacher cognition and practice: International case studies*. Channel View Publications. Retrieved from http://ebookcentral.projquest.com 2019/5/16

Nguyen, T. (2017). *The Transition of an ESP Teacher from a GE Teacher: A Case Study. (Unpublished master's thesis).* Macquarie University, Australia.

Ryan, J., Rabbidge, M., Wang, Y., & Field, J. (2019). Satisfiers and Dissatisfiers for International Vocational Education Students: A Case Study Using Narrative Frames. *Journal of International Students, 9*(3), 795−814.

Shelley, M., Murphy, L., & White, C. (2013). Language teacher development in a narrative frame: The transition from classroom to distance and blended settings. *System, 41*(3), 560−574.

Swenson, T. & Visgatis, B. (2011). Narrative frames to assess overseas experiences. In A. Stewart (Ed.), *JALT2010 Conference Proceedings*. Tokyo: JALT.

Wette, R., & Barkhuizen, G. (2009). Teaching the book and educating the person: Challenges for university English language teachers in China. *Asia Pacific Journal of Education, 29(2)*, 195−212.

Xu, Y. (2014). Becoming researchers: A narrative study of Chinese university EFL teachers' research practice and their professional identity construction. *Language Teaching Research, 18*(2), 242−259.

Appendix A: Frame Used in the Present Study Based on Barkhuizen and Wette (2008)

Writing (Core 4) Course Feedback

Thank you for being a student in Core 4 this year. It was wonderful teaching you☺

1. Read all the sentences BEFORE beginning to write.
2. Next, write the story of your time in Core 4. Do this by filling in the blank spaces.
3. Do not write your name on the form. This is anonymous.

(1) In April, at the start of this class, I felt _____ _____ .

(2) I remember once in my writing class I had a hard time trying to _____ .

(3) The main reason for this was that _____ _____ .

(4) The writing assignment I most enjoyed was _____ _____ (5) because _____.

(6) If I could take this class again, I would like more _____ (7) but less _____ .

(8) The most useful thing I did in this class was _____ _____ .

(9) It was easy/difficult to do the homework because ___ _____ .

(10) A future difficulty I might have with improving my writing skill is _____ _____ .

(11) However, I might/will overcome this difficulty by
_____ .

(12) Now, in January, the class is almost over, and I feel
_____ .

Utilizing CLIL in a Japanese Elementary School Art Classroom: An Analysis of Student Perceptions of Teaching English and Foreign Culture Through Art

John Andras Molnar* and Kazuhiro Nomura**

*Kinjo Gakuin University
jmolnar@kinjo-u.ac.jp
**Kinjo Gakuin University
nomura-k@ kinjo-u.ac.jp

Abstract

This research explored the application of CLIL in teaching English in a Japanese elementary school. The MEXT Course of Study (MEXT, 2017) implemented in 2020 brings new challenges for elementary schools as English becomes an official subject in the curriculum. English lessons will need to foster cultural understanding and include English instruction without explicitly teaching grammar, and CLIL was identified as a potential teaching method to implement the MEXT requirements (Yamano, 2013). This study investigated the potential of using CLIL in elementary schools by teaching an art class focused on Norwegian trolls in a 4th grade, Japanese public school classroom. The research questions focused on (1) student perceptions of enjoyment and difficulty regarding having Art class in English and (2) which aspects of the class regarding content were most salient. The results of a questionnaire survey administered at the end of the lesson indicated that using English was challenging for the students, but they thoroughly enjoyed the class and the difficulty did not impede the lesson. Speaking in English was identified as the most difficult part of the lesson. Students also expressed on the survey that they learned content about cultural differences, but the specifics were inconclusive.

Keywords: Elementary School English, CLIL, Art, Young Learners

1. Introduction

Formal instruction in the English language has long been an official subject in Japanese junior high and high schools, but has yet to be formally introduced in elementary schools. This situation will change as of April 2020 when Japan's Ministry of Education, Culture, Sports, Science and Technology's (MEXT) mandated Course of Study from 2014 will take effect and change English from an elective subject (Foreign Language Activities) to a formal subject (MEXT, 2017). This change will bring many challenges to schools and teachers such as teachers having to teach a new subject with little or no pedagogic training (Tahira, 2012), the low English ability of teachers (Yahata, 2018), and additional duties on top of an already heavy load for homeroom teachers (HRTs) (Kano et. al, 2016). Additionally, elementary school English education is mandated to be taught without explicit metalinguistic grammar instruction (MEXT, 2017), thus requiring teachers to creatively teach a subject they are not familiar with. Due to this policy change, it is necessary to develop and research potential ways that can help teachers overcome the challenges associated with this new policy, yet satisfy the requirements of MEXT.

Content and Language Integrated Learning (CLIL) was identified by the researchers as a potential way to bridge the gap between teacher needs and the MEXT policy. CLIL is a method used to teach subject content usually taught in the students' L1 in a second language, and has had earlier successes in Japanese elementary school English education (Yamano, 2013). The 4C's of CLIL (Content, Communication, Cognition and Culture) match well with the MEXT objectives for elementary school English education (Ito, 2018) and thus mutually support each other.

This research uses CLIL to teach a content class in the English language medium about Norwegian trolls, and combines that with Art class (図画工作) to deliver a more visual means of reinforcement for the students (Fujii et al., 2016; Ito, 2018). While CLIL has been found as an effective means of language teaching in Europe, this research is still new to Japanese elementary school English, especially art class contexts, thus warranting further investigation.

1.1 Purpose of the Study

The purpose of this study was to test student perceptions of a CLIL class using English and Art, along with assessing what aspect of the class resonated most with Japanese elementary school students. Norwegian trolls were chosen as content for the lesson as they fit into the MEXT framework of having students think about other cultures (MEXT, 2017), and moreover offer an effective way to teach English body parts, colors, and adjectives in an integrated manner as the content is highly visual. Comparisons between Norwegian trolls and Japanese ogres (oni) are also an age-appropriate topic deemed fun for the students. Art class was selected to allow students to collaboratively express themselves, while visually

reinforcing the English learned in the content focused class. Also, few studies have been conducted regarding cross-curricular use of CLIL in Japanese elementary schools.
To this degree, the research questions were as follows:

RQ1: How does having a CLIL Art class in English affect student perceptions of enjoyment and content difficulty?

RQ2: What aspect of the class regarding content was most salient for students?

2. Background

2.1 MEXT CoS Policy for 2020

MEXT is in charge of determining public school elementary and junior high school education policy in Japan, and every several years revises their Course of Study (CoS), which are educational guidelines for different school subjects. The 2004 CoS guidelines (which took effect in 2011) mandated that elementary schools teach foreign languages in classes called "Foreign Language Activities", which were classes like Music or Art that have a limited place in the curriculum, but are not core subjects like Social Studies or Math. Foreign Language Activities were part of the 5th and 6th grade curriculum (though lower grades often also had these classes), and were required to meet at least 35 hours a year (General Union, 2017). These classes focused on getting students used to foreign sounds and cultures rather than focusing on formal grammatical instruction. While English was generally used in these classes, schools were not limited to teach English exclusively. In some cases, other cultures or languages were taught during this time. This flexibility and vagueness gave elementary school HRTs the discression to use Japanese in the classroom, or leave language instruction entirely up to Assistant Language Teachers (ALTs) (Tahira, 2012) and avoid dealing with English.

However, the 2014 CoS taking effect in 2020 represents a major shift from the 2004 CoS. In the 2014 CoS, Foreign Language Activities will be shifted down to the 3rd and 4th grade, and English will become a core subject in the 5th and 6th grade curriculum. The objective of this new policy is to have students get used to reading, writing, listening and speaking in the English language before junior high school, as well as deepening student understanding of foreign cultures. This new curriculum will focus on pronunciation and the four skills, but not through grammatical instruction. The elementary school curriculum is designed to introduce English to students generally, with junior high school classes beginning formal foundational grammar instruction as has traditionally been the practice at that level (MEXT, 2017).

With this shift in policy, there is a new burden on teachers to implement a new and unfamiliar curriculum. Where under the 2004 CoS policy HRTs could lightly teach English or rely on ALTs, the new 70-hour English curriculum for 5th and 6th grade requires weekly instruction to satisfy the mandated policy, not to mention formal assessment, thus increasing the burden on HRTs. This creates a challenge for current teachers as in many cases elementary school faculty have received little or no formal English language education since high school or university, and especially no formal training in foreign language instruction (Tahira, 2012). Thus, there is a need for creative ways to integrate English teaching into elementary school without complex pedagogic requirements. To this degree, CLIL was selected by the authors of this manuscript to explore a potential means to fill that gap.

2.2 CLIL

CLIL, a style of teaching that is popular in Europe, uses a second language to teach subject content ordinarily taught in a student's L1, with language and content instruction treated as equal (Coyle et al., 2010). Marsh (2002, p. 15) defines CLIL as "any dual-focused educational context in which an additional language, thus not usually the first foreign language of the learners involved, is used as a medium in the teaching and learning of non-language content." This technique can create a more immersive environment and contextualized use of language to discuss school subjects that go beyond foreign language textbooks.

CLIL classes operate on a 4C's framework to create this complex learning environment in the classroom. The 4Cs framework for CLIL include Content (such as subject matter, themes, cross-curricular approaches), Communication (language), Cognition (thinking) and Culture (awareness of self and 'otherness') (Coyle, 2008) to integrate both the learning of content and language. The 4C's framework overlaps with the objectives from the MEXT 2014 CoS, suggesting that CLIL is potentially an effective way of teaching foreign language in Japanese elementary schools (Ito, 2018). Below is a table outlining the 4C's of CLIL (Table 1), followed by how MEXT's objectives for elementary school English can also fit into the 4C's (Table 2).

Table 1: 4C's of CLIL (Coyle, 1999, p. 54)

Content	• coherence
	• progression
Communication	• linguistic forms and functions of the language needed by learners;
	• development of communication skills, including discourse strategies;
	• three M's, i.e. meeting new language, manipulating it, and making it 'my own'
	• coping with the unexpected;
Cognition	• development of tasks related to thinking skills;
	• construction, sequencing and evaluation of learner tasks;
Culture	• developing a sense of otherness;
	• contrasting different perceptions of events roots in different cultures;

Table 2: MEXT 2014 CoS Objectives respective of 4C's (MEXT, 2017)

| Content | • students gain knowledge of both Japanese and foreign languages such as characters (alphabet), vocabulary, expressions, sentence structure and language function. |
| Communication | • develop communication skills via reading, writing, listening and speaking |

	• develop basic skills to communicate thoughts and feelings by consciously writing
Cognition	• teaching and evaluating students' English ability **without** explicit metalinguistic grammar
Culture	• to deepen the understanding of the culture behind foreign languages, and develop an attitude to proactively communicate using foreign languages. • Comparing and contrasting different perceptions of an event from various perspectives

2.3 CLIL and Art Class in Japanese Elementary Schools

Among the current research investigating the effectiveness of using CLIL to combine Art class and English, the relatively few studies done in this area reported positive results. In one, Fujii et al. (2016) investigated CLIL in Art class using the topic of "let's use circle, triangle, and square shapes to make a picture!" with third graders. The researchers reported that at a basic level, students were both able to complete a picture drawing task, and also that the students felt they were able to successfully use English. Moreover, the researchers noted that this type of class allowed students at various levels of English and cognitive development to have hands on, visual reinforcement for the words they were learning. The researchers stated that CLIL was an effective means of teaching art and English. Similarly, Ito (2018) found CLIL a successful means to teach English and Art. In the study, 5th and 6th grade students were successfully able to learn several English phrases regarding content information about traditional fans in Nara, the topic of the CLIL lesson. Like Fujii et al. (2016), Ito also found that the CLIL class was enjoyable for students and observed that while communication often happened in the Japanese mother tongue, a lot of communication also occurred in English. Ito deemed this as natural given their level and understanding of the English language. Both results from Fujii et al. (2016) and Ito (2018) imply that there is potential for CLIL in Japanese elementary schools. The current study was created to further explore art taught via CLIL as a vector to test the efficacy of CLIL in this context.

3. Methodology

3.1 Participants

This study was conducted in a 4th grade class at a Japanese elementary school in Nagoya, Japan. There were 37 participants in the class consisting of 22 boys and 15 girls. While English education at this school officially begins at the 5th grade, the HRT in the class holds a junior high school English teacher certification, and thus occasionally conducts Foreign Language Activities classes herself.

During the troll lesson, there were three teachers in the classroom. Two were the researchers who conducted the study and taught the class, and the third was the HRT. Researcher 1 was an English native speaking, non-Japanese teacher who conducted the English parts of class. Researcher 2 was a Japanese native speaking, Japanese teacher who conducted the art portion of the class. The HRT was a Japanese native

speaking teacher who assisted with classroom management.

3.2 Materials and Research Procedure

Data in this research were collected from various vectors. The first was a questionnaire administered directly after the conclusion of the troll class to assess student reactions to the material and class. The questionnaire comprised of 10 questions, four of which were open response, and six of which were based on a 6-point Likert scale. The Likert scale questions were written to have the answer of a six being a "strong yes", and one being a "strong no" to avoid confusion. The questionnaire was conducted in Japanese with the students encouraged to answer in Japanese to allow them to express themselves freely. A limited questionnaire was created because the participants were 4th grade students in elementary school, and the researchers were concerned about survey fatigue. The questionnaire is as follows (translated from Japanese by the researchers):

Q1: What were you in charge of today? Please write in English.

Q2: Was the English class fun? (6-point Likert scale)

Q3: Was studying English difficult? (6-point Likert scale)

Q4 : What part (of studying English) was difficult?

Q5: Were you able to understand the content of the class with all the English used? (6-point Likert scale)

Q6: Did you learn anything new related to foreign countries? (6-point Likert scale)

Q7: What did you learn specifically (about foreign countries)?

Q8: Was it fun to use English in art class?   (6-point Likert scale)

Q9: As you were drawing, were you able to express yourself in English? (6-point Likert scale)

Q10: How did you feel using English outside of English class?

The questionnaire was created based on recommended practices from Borgers et al. (2000) for conducting research with children aged 8-11 years old. For example, keeping the survey short and concrete were deemed important procedures when creating the survey.   The survey data were compiled and analyzed using standard descriptive statistics (Carr, 2011).

In addition to the questionnaire, the researchers made a video recording of the class for a detailed review of the student-teacher interactions, and interviewed the HRT.

3.3 Lesson Procedure

As was previously stated, the objective of this research was to explore a potential way for elementary school teachers to

effectively implement lessons that are engaging, are manageable, and fulfill the objectives of the 2014 CoS. Based on the success of CLIL in other studies in this context, as well as the fit with the MEXT 2014 CoS aims (Table 2), this lesson about Norwegian trolls and art was created using the 4C's of CLIL as a model (Table 3).

Table 3: The 4C's in the Norwegian Troll CLIL Lesson

Content	• Learning new vocabulary about and the history of Norwegian trolls • Learn how to talk about facial features (eyes, ears, ect.) • Learn the adjective + noun English form in non-metalinguistic format (The troll has big eyes)
Communication	• To repeat target phrases and vocabulary • To tell the teachers and the class in English what their troll (created in art class) looks like • To work together in a group to collaboratively create a troll
Cognition	• To imagine and create a troll based on *hearing* a description of a • Norwegian troll (the actual image was revealed the third period)
Culture	• To compare Japanese ogres (oni) to a different culture's version of a similar mythical creature • To learn about another culture's mythology

Comparisons between Japanese ogres (oni) and Norwegian trolls were chosen as content for this research. Ogres and trolls offer a fun, age appropriate subject that students are interested in, visual content students can use in Art class, as well a base of something (ogres) all Japanese students are familiar with to compare to similar foreign content. Where Ito (2018) used traditional fans from Nara as content, the researchers considered ogres something that students already had a visual image of (like fans) and could use that image to try to imagine what a foreign version might look like. Considering the case in Ito (2018), students know what fans are and could easily understand Nara specific fans without too much cognitive load.

With the 4C's in mind, the researchers used three 45-miunte classes in one day at an elementary school to implement the complete lesson, from English to Art class, about Norwegian trolls. The first period was used to have students consider the differences between Japanese ogres (oni), and trolls (specifically from Norway due to the prepared story) in other countries. This part of the session was done in English, introducing target vocabulary (face, colors, adjectives) students would use when creating their own troll during the second period. For period one, the picture of the troll was covered to have students imagine what a troll might look like based on the teachers' English description. Next, the researchers read students a traditional story called Prince Harold and the Giant Troll adapted from Troll Forest (Torooru no mori) (Craigie, 2004). The researchers created a kamishibai panel story for the children to look at, and covered

the troll in the story, again to have student imagine what it looked like. Finally, after a short discussion between the researchers and students mixed in English and Japanese about what trolls and ogres might be like outside of Japan and what they looked like, the researchers introduced key vocabulary about facial features, adjectives, and colors before finishing the period. See Table 4 for a more general layout of the lesson with target terms.

Table 4: Lesson Plan Breakdown for Troll Class

Periods	Teacher Roles	English Vocabulary
Period 1 – Introduction of trolls and key English vocabulary (45 minutes)	R1 was T1 R 2 was T2 HRT was support	**Face**: hair, eyes, ears, mouth, nose, face shape **Adjectives**: short, long, curly, straight, big, small, oval, round, pointed, straight, curved, full, thin, square, triangle **Colors**: red, orange, yellow, blue, green, purple, white, black
Period 2 – Art class where students made the troll face (45 minutes)	R2 was T1 R1 was T2 HRT was support	Key Phrases: What is that? It's a _____ What (_____) does it have? It has red (hair).
Period 3 – Students share their art and discuss Norwegian trolls vs Japanese ogres (45 minutes)	R1 was T1 R 2 was T2 HRT was support	No new vocabulary; only students presenting their art.

The second period was the Art (図画工作) portion of the lesson. In this period, the students worked together collaboratively to make the face of what they imagined a troll to look like based on the story. The students sat in groups of five or six, and chose what part of the troll face (hair, eyes, ears, mouth, nose, face shape) they wanted to create. Students would first write down features of their troll on a worksheet, (for example, my troll has big, yellow eyes) then practice drawing their image. Each students' part of the face comprised of a noun, adjective, and a color to combine. Students then drew their creation on separate piece of paper, cut out their part, and combined the group parts to make a face. These creations were then displayed on the blackboard (see Picture 1, Appendix A for student artwork). While the students were working, both the researchers and the HRT visited each student individually and asked in English what they were working on and helped the students to answer if they had difficulty. Most of the discussion about the trolls however was done in Japanese among the groups.

Period 3 was used to have the students share their trolls with the class, and present their portrayals of Norwegian trolls. At the start of this final period, the students stood in front of the class and individually shared their different parts of the troll in English. After sharing, the researcher re-read the

kamishibai about Prince Harald, this time with the troll uncovered. With the remaining time, students were shown pictures of other trolls, and discussed differences between mystical creatures in Japan and other parts of the world. Lastly, the students completed a survey.

4. Results and Discussion
4.1 RQ1: How does having a CLIL Art class in English affect student perceptions of enjoyment and content difficulty?

In order to ascertain an answer to the first research question, data from the questionnaire and interview with the HRT were used. The results from the relevant items of the questionnaire are displayed in Table 5.

Table 5: Questionnaire Descriptive Statistics

	mean	max	min	SD
Q2: Was the English class fun? (6-point Likert scale)	5.52	6	3	0.66
Q3: Was studying English difficult? (6-point Likert scale)	3.55	6	1	1.35
Q5: Were you able to understand the content of the class with all the English used? (6-point Likert scale)	5.33	6	3	0.81
Q8: Was it fun to use English in art class? (6-point Likert scale)	5.66	6	3	0.69
Q9: As you were drawing, were you able to express yourself in English? (6-point Likert scale)	4.66	6	3	0.96

As can be seen from Q2 and Q8, the relatively high mean in response to the questions (mean 5.52 and 5.66 respectively) indicate that they students enjoyed the class. The means fall close to the maximum of 6 on the Likert scale, indicating a strong "yes". There is also a low SD (0.66 and 0.69 respectively), indicating that student responses are centered around the mean. Thus at the very least, the survey results suggest that they did enjoy the class.

While the students seemed to enjoy the class, other items on the questionnaire reveal that the CLIL class was challenging for the students. Q3 directly asked the students if the class was difficult, to which they responded with a mean that 3.55. On the Likert scale used in the survey, this response is almost right in the middle. There is a high SD (1.35) indicating the answers are spread out, and responses to the question included answers on both ends of the maximum and minimum. This suggests that there were a variety of feelings in the class toward the lesson regarding its difficulty, with some students able to participate in the class with little difficulty, and others with presumably more difficulty.

Q4, a fill in the blank question asking "What part (of studying English) was difficult?" offers more insight into what specifically was difficult about the class. Several students that responded close to the mean said that, "responding to the teacher's questions in English" was difficult, as well as "pronunciation" of the English words.

Others that answered six or one (the extremes) also identified "speaking" as the most difficult part of the class. Students at both ends of the spectrum answering the same thing could indicate that the students are not really familiar with speaking English, which is understandable as they do not have regular English classes, or regularly use English outside of class.

Though "speaking" was considered difficult, Q5 and Q9 show that the students felt they could understand the content in the class, and that they were able to express themselves when the researchers talked to them in English during the Art class portion of the class. Though the English aspect of the class seems to have been challenging as suggested from Q3, the students were able to participate in the class, and thus it can be said that the CLIL class did not considerably hinder learning in the class, and was manageable for most students.

Q10, an open response question, asked students how they felt about using English outside of a specifically English designated period (in this case, in Art class), and the results showed the students generally had a positive response. Following in Table 6 is a selection of student responses from Q10 (paired with their response from Q2 indicating enjoyment and Q3 indicating difficulty) to show the variety or responses. While the students indicate the class was challenging (Q3), they also seemed to enjoy it as is suggested in Q2 and Q10. The researchers translated the student responses from Japanese to English.

Table 6: Selections from Student Surveys for Q10

Student	Q2	Q3	Q10
1	6	2	It was really fun to use English in another class.
2	6	1	It was fun to do English and Art.
3	6	4	Using English and doing that with art was more fun than art generally is. I would like to use English again in art next time!
4	5	4	Hearing a story in English was actually more fun than I thought
5	5	3	It was fun, I would like to do it again.

The HRT reported similar observations as the students indicated in their survey. She mentioned that the students seemed to find it difficult to communicate in English, but from her observation, they did enjoy the class. She pointed out that one likely reason for students finding the class difficult, and specifically speaking, could have been their reluctance to make mistakes. She said that this same behavior occurs in other classes such as math and science, and she spends a lot of time telling them that learning takes place through mistakes, and that mistakes are alright.

Thus, looking at the results of the survey and the interview, it can be said that students generally found the class challenging (specifically speaking in English), yet enjoyed the class. While more research into how to deliver a CLIL class to students that do not regularly speak English is necessary to reduce the anxiety the students seemed to feel when speaking English, these results suggest that CLIL is a viable way to teach English in elementary school. The results

also suggest using CLIL to teach content outside of English specified times is suitable for a large class containing students with diverse language abilities as even students who found the class challenging had fun.

4.2 RQ2: What aspect of the class regarding content was most salient for students?

For the second research question, two items were taken from the questionnaire for analysis. The first was a Likert scale question, Q6: "Did you learn anything new related to foreign countries?". The results are displayed in Table 7. The mean indicates that student strongly felt they learned in the lesson about foreign countries, and the SD indicates that most students fell around the mean.

Table 7: Questionnaire Descriptive Statistics Q6

	mean	max	min	SD
Q6: Did you learn anything new related to foreign countries?	5.33	6	3	0.81

Regarding which aspects of the class were salient for students, Q7 was used (Table 8). The majority of students responded with a comment, while some responses were more specific than others. See Table 8 for some selected responses from the questionnaire.

Table 8: Student responses to Q7: What did you learn specifically (about foreign countries)?

Type	Q7 Example	Similar responses
1	There are things other than ogres in foreign countries, trolls. (specific cultural point)	11
2	Foreign culture. (vague cultural point)	11
3	I learned about a foreign story. (foreign story)	4
4	I didn't know how to say beard or ear in English, and learned that. (English)	7
5	Blank (3) or don't know (1).	4

While all the students had slightly different answers, the researchers noticed that four trends emerged among the responses. Type 1 was a more detailed response (using specific nouns, more than two words) about foreign culture, Type 2 was a short (one or two words) response about foreign culture, Type 3 specifically mentioned the troll story read in class, and Type 4 was English language related. Type 5 was blank or "don't know".

Roughly one-third of the class gave specific, detailed responses such as "differences between trolls and ogres" or "trolls can both be kind and evil", but others only gave a vague response, focused on the story, or cited English as a their main take aways from the class. The researchers were surprised that seven students responded about English content specifically in Q7, and considered this an indication that language was more salient than content for a some of the class.

While the responses to Q6 and Q7 do indicate that content related learning took place in the class, the researchers acknowledge that more research is needed on this specific point. The students seemed satisfied that they learned something as indicated by the high mean (5.33), however, what they learned specifically, or how deeply they processed the information was not uncovered here. More questions on the questionnaire examining the research question from different vectors, or a specific content centered pre/post-test quiz could have made the results in the section clearer.

5. Conclusion

With the new mandated English curriculum from MEXT (2017) to be implemented in elementary schools and the gap in pedagogic knowledge for the practitioners (Kano et. al, 2016; Tahira, 2012; Yahata, 2018), it is necessary for research to be done into how to effectively teach English in this context. In this study, CLIL was selected as a potential means to teach English in elementary schools from previous studies (Fujii et al., 2016; Ito, 2018, Yamano, 2013), and from how MEXT objectives fit with the 4C's of CLIL. Using Norwegian trolls as content, the students imagined what the trolls looked like through hearing a story, and created what they imagined a face of a troll to look like in Art class and explained it, using English vocabulary from the first period.

In general, the results seemed promising and fit in line with previous studies. Similar to Fujii et al. (2016), all the students were able to complete the task, and the student responses to Q2 and Q3 indicate that students found the activity enjoyable yet challenging. This fits the supposition in Fujii et al. that the CLIL classes can accommodate learners at various stages. Ito's (2018) research claims that students found the class enjoyable and learned several phrases were also supported by this research. That said, further research needs to be done on specifically what content the students learned, and also what English phrases they acquired. One caveat that should be mentioned is that the preparation for this class was very time intensive for the researchers, and unless elementary school teachers have access to pre-made material that could facilitate them teaching a class such as this, the HRT would have a high work burden of preparation. However, in general, this research shows that CLIL when combined with another subject matter, in this case art, does seem to be an effective means of teaching English in elementary school.

Further directions for this research would be to test CLIL classes taught by HRTs as the primary teacher in the class. While CLIL is definitely viable approach to teaching language in Japanese elementary schools, surveying how HRTs unfamiliar with the method teach would give insight into how to implement CLIL at a wider scale in classrooms.

References

Borgers, N., Leeuw, E. D., & Hox, J. (2000). Children as Respondents in Survey Research: Cognitive Development and Response Quality 1. *Bulletin of Sociological Methodology/Bulletin De Méthodologie Sociologique*, *66*(1), 60–75. doi:10.1177/075910630006600106

Borge, P. E, and Kittlesen, T. (1998) *Troll*. Oslo: SFG.

Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford: Oxford University Press.

Craigie, W. A. (2004). *Tororu No Mori No Monogatari: hokuō no minwashū* (Y. Higashiura, Trans.). Tōyō Shorin.

Coyle, D. (1999). Supporting students in content and language integrated contexts: planning for effective classrooms. In J. Masih (Ed.), *Learning through a foreign language: Models, methods and outcomes* (pp. 46–62). London: Centre for Information on Language Teaching and Research.

Coyle D. (2008). *CLIL A Pedagogical Approach from the European Perspective*. In Hornberger N. H. (Eds.), Encyclopedia of Language and Education. Springer, Boston, MA.

Coyle, D., Hood, P. & Marsh, D. (2010). *CLIL Content and Language Integrated Learning*. Cambridge: Cambridge University Press.

Fujii, et. al. (2016). Shōgakkō zuga kōsaku-ka. Gaikokugokatsudō no kyōdō gakushū no jissen-teki kenkyū ichi eigo o tsukatte-ban ni yoru hyōgen o okonau jugyō no ichikōsatsu [Elementary school arts and crafts. A Practical Study of Collaborative Learning in Foreign Language Activities A Study of Classroom Expressions Using English]. *Oita University Bulletin of the Center for Higher Education Development*, 8.

General Union. (2017). *MEXT's 2018/2020 "New course of study in foreign language education" plan*. Retrieved from http://www.generalunion.org/laws-and-rights/1696-the-2018-transition-towards-smooth-implementation-of-new-course-of-study-in-foreign-language-education

Ito, Y. (2018). CLIL in Practice in Japanese Elementary Classrooms: An Analysis of the Effectiveness of a CLIL Lesson in Japanese Traditional Crafts. *English Language Teaching, 11*(9), 59–67. doi:10.5539/elt.v11n9p59

Japanese Government Ministry of Education, Culture, Sports, Science and Technology. (2017). Shōgakkō gaikokugokatsudō gaikoku-go kenshū gaidobukku [Elementary School Foreign Language Activities and Foreign Language Training Guidebook]. http://www.mext.go.jp/a_menu/kokusai/gaikokugo/__icsFiles/afieldfile/2017/07/07/1387503_1.pdf

Kano, A., Sonoda, A., Schultz, D., Usukura, A., Suga, K., & Yasu, Y., (2016). Barriers to effective team teaching with ALTs. In P. Clements, A. Krause, & H. Brown (Eds.), *Focus on the learner*. Tokyo: JALT.

Marsh, D. (2002). *CLIL/EMILE—The European Dimension: Actions, Trends and Foresight*. Potential, Public Services Contract DG EAC, European Commission, Strasbourg.

Tahira, M. (2012). Behind MEXT's new course of study guidelines. *JALT: The Language Teacher, 36*(3), 3–8.

Yahata, M. (2018). Problems to consider and solve for the formal start of English education in Japanese elementary schools: Through the interviews with the teachers and students from Taiwan and Japan. *Bulletin of Beppu University Junior College, 37*, 93–104.

Yamano, Y. (2013). Utilizing the CLIL approach in Japanese primary school: A comparative of CLIL and EFL Lessons. *The Asian EFL Journal, 15*(4), 160–183.

Appendix A: Pictures from Class

Picture 1 displays the students' finished trolls. These are the creations from the 6 groups during the Art period. They were displayed on the board during period 3, and students explained the different aspects of their trolls (my troll has big, round eyes) to the class.

Picture 2 displays how in period 1, the researcher introduced Norwegian trolls to the students with the Japanese oni picture revealed, but the Norwegian troll hidden to force students to think about the image for themselves.

A Pilot Study on Acceptability of Machine Translation: How do Japanese Office Workers Accept the Use of Machine Translation at Their Workplace?

Kazuma Hatano*

*Soka University
khatano@soka.ac.jp

Abstract

This study aims to investigate how Japanese office workers accept the use of machine translation at their workplace. Due to the recent development of neural machine translation, there are growing expectations that machine translation can be usable in various situations in Japan, especially in business contexts. However, the present study questions the idea that translation accuracy is the only determinant of how machine translation is accepted by Japanese office workers; the study presumes that naturalness and politeness of machine-translated messages are also important factors that affect the acceptability of machine translation. To investigate it, an online questionnaire survey was administered to Japanese office workers. The results show that machine-translated messages that lack naturalness and politeness are less acceptable, which implies that these two elements can also determine acceptability. The study argues that this could be due to Japanese people's sensitivity to the hierarchical relationship. The paper also discusses implications on the value of "authentic" communication and on foreign language education.

Keywords: machine translation, business, politeness, foreign language education, Japan

1. Introduction

1.1 Development of Machine Translation in Japan

The recent development of Google's neural machine translation using deep learning and big data has attracted a great deal of attention from people all over the world due to its dramatic improvement of translation accuracy. It has also gained high expectation in Japan, where its dominant language (Japanese) has great linguistic distances from many other languages. While much of the current neural machine translation technology owes to the contribution of global private corporations such as Google, in the case of Japan, the public sector has played the central role in developing machine translation. The National Institute of Information and Communications Technology (NICT) has worked with private corporations to develop a platform that provides translations between Japanese and other languages. Due to its development, the country has now seen more possibilities than ever before to enjoy the benefit of machine translation.

The Japanese government plans to use the machine translation technology to support the lives of foreign workers. In 2018, the Japan Ministry of Justice published plans to reinforce support for foreign workers, who are expected to increase in number after the enforcement of the 2018 Amendment of Immigration Control Law. The plans include the usage of machine translation to provide multilingual services in public facilities for foreign workers, as well as other support plans such as provision of Japanese language lessons, human translators, and multilingual manuals.

1.2 Application to Business

Another expectation for machine translation is its application to more professional purposes such as reading and writing business emails, conducting meetings, and negotiating with international business partners. Machine translation may be useful for corporations because it provides multilingual translations at much cheaper cost than human translators who can translate only between two languages. To increase efficiency and productivity in their corporate activities, using machine translation may become an important choice for their international communication.

Over the years, there have been various proposals and implementation reports (ex. Shimohata, 2005; Aiken and Ghosh, 2009) on the application of machine translation to business, but some of them had to wait for more accurate machine translation technology. However, with recent developments, it is evolving to be usable in many business situations. In fact, Sumida, a NICT scientist, reports on the development of simultaneous translation technology, suggesting that it will be applicable to multilingual video conferencing in the near future (Yoshikawa, 2019). From technological viewpoint, it seems that machine translation will soon be applicable to most business situations.

1.3 Investigation for Acceptability

However, the important question here is whether using machine translation is *acceptable* in any business context. No matter how accurate a machine translation is, it may not be always acceptable for certain business communications. For example, using machine translation for dinner conversation with business partners may be taken as inappropriate not because the machine translation does not convey a message accurately but because having an indirect way of

communication using a machine is rude in this context. To understand how office workers accept the use of machine translation in business, just discussing the aspect of technological development is not sufficient; it also requires investigation for social aspects such as accepted social norms.

In fact, as machine translation has achieved its technological advancement enough to actualize practical usage in everyday life, more scholars started to discuss not only the technology itself but also its impact on society at large. For instance, Inoue (2017) argues that by 2025 there will be a national debate that will reassess the meaning of studying English in school. As another example, Takita and Nishijima (2019) published the book titled *Machine translation and future society: Will we see a world without language barriers in the future?* This is probably the first Japanese book published after Google had launched its neural machine translation in 2016, to solely discuss the relationship between machine translation and its impact on society. This compilation of discussions in the form of a book can be seen as the result of increasing awareness of the social impact of machine translation.

Discussing the social impact of machine translation is more associated with the use of machine translation for facilitating business operations than for assisting foreign tourists and foreign workers. When machine translation is more widely used at workplace, it may affect decisions on how foreign language education should be taught in school; and thus, it affects the entire Japanese population. Therefore, it is important to investigate how machine translation is accepted in the society, and in business contexts in particular.

However, there are few empirical studies that can provide data to help us understand the social impact of machine translation and to substantively judge whether it will be really accepted in the society. It is for this reason that the current research tries to collect data to understand the acceptability of machine translation.

Since there are few prior studies on this topic, it is necessary to start investigation with some rough presumptions. The present study questions the idea that translation accuracy is the only determinant of how machine translation is accepted by Japanese office workers; thus, attempts to find out other determinants would be the first step.

As one of such determinants, *politeness* is a good candidate. Hanari (2019) points out that the concept of politeness must be fully reflected in machine translation. He introduces the definition of politeness as "linguistic behavior to establish and maintain smooth human relationships" (p. 54, translated by author), referring to the theory of politeness established by Brown and Levinson (1987) and its interpretation by Usami (2003). If machine translation is used for communicating with others, ensuring smooth human relationships is an important issue. Therefore, whether messages are polite or not may affect how people accept machine translation.

The study also presumes that *naturalness* is a candidate as an important determinant of acceptability. For example, due to lack of contextual information, Google Translate can still make a weird choice of vocabularies and produce unnatural sentences we would not usually make. In fear of sending unnatural messages, the sender may be concerned about using machine translation. Therefore, it is presumed that naturalness of a machine-translated message may serve as a determinant of acceptability.

In summary, the current study identifies the following research questions:

(1) How is machine translation accepted by Japanese office workers?
(2) Do naturalness and politeness of machine-translated messages determine the acceptability of machine translation?
(3) What other aspects of machine translation determine the acceptability of machine translation?

2. Current Study

2.1 Purpose

The purpose of the current study is to provide meaningful answers to the above research questions. Since there has not been much empirical research yet on this topic, the study necessarily has to make some presumptions about how machine translation is accepted. Thus, the study positions itself as a pilot study and as the first step to more detailed investigation in the future.

2.2 Method

To understand Japanese office workers' attitude toward the use of machine translation, a questionnaire survey was administered to individuals working for listed companies on the Japanese stock market. The Institute for International Business Communication (2013) investigated the use of the English language in listed companies in Japan and provided the percentage of companies by industry that use English at workplace. Based on the data, companies were selected from industry sectors such as electric and precision machine manufacturer, international trading, chemical and pharmaceutical, and transportation equipment, which have higher percentages than other sectors. A letter of invitation to the survey was sent by postal mail to the human resources department of 1007 companies among more than 3000 listed corporations in Japan.

The survey was conducted online for two months from September 1 to November 1, 2019, and 38 responses were collected during the period (response rate: 3.8%). 6 sets of questions were prepared to elicit responses associated with naturalness and politeness of machine-translated messages. Respondents were instructed to assume that the machine translation conveys a message accurately enough not to interfere with their operations at work. The questionnaires elicit responses to 5 levels of Likert items (1. Strongly disagree; 2. Somewhat disagree; 3. Neither agree nor disagree; 4. Somewhat agree; and 5. Strongly agree), which is a widely used approach to investigate respondents' attitude and opinions. Each of the questions will be explained in more detail below, and the complete list of questions can be found in Appendix A.

2.3 Questionnaires

Question 1 asks whether respondents have favourable

opinions about using machine translation in general. The reason for setting this question is that if there are many negative responses to the subsequent questions, it will be difficult to determine whether they are responses to machine translation in general or a specific matter asked in each question. Question 1 serves as a basis on which responses to other questions are judged.

Question 2 was prepared in order to see whether respondents have negative opinions about lack of naturalness and politeness in translated texts. The question instructs respondents to assume that they use machine translation in email communications (This also applies to Question 3 and 4). The question also asks whether they have any problems in translated texts lacking "accuracy in meaning" and "accuracy in grammar." This set of questions are to be compared with responses about naturalness and politeness. The study presumes that respondents will respond less negatively to lack of naturalness and politeness, assuming translation accuracy is more important in business.

Question 3 asks the same set of questions as Question 2, but in a more specific situation; respondents are asked to assume that a translated message is supplemented with the statement "This message has been translated by machine." The statement allows respondents to assume that the fact that a message is translated by machine is explicitly explained to the other party. The study presumes that this kind of message alleviates a feeling of hesitation to use machine translation, and thus increases its acceptability.

Question 4 asks whether a machine-translated message that lacks "accuracy in grammar," "naturalness," and "politeness" is rude to the other person even if it conveys the message accurately. The second question of Question 4 asks if the degree of rudeness changes depending on the social status and work relationship of the other person. This set of questions is intended to examine whether the acceptability of machine translation is associated with rudeness, which comes forth in human relationships. Question 2 and 3 ask if respondents have "no problems" with translated messages lacking naturalness and politeness. To check if respondents perceive "no problem" as "no problem with rudeness," Question 4 asks more directly whether such a message is rude or not, and see if this perception changes depending on whom they communicate with.

Question 5 and 6 are intended to examine other possible determinants of acceptability. Question 5 elicits responses to the use of machine translation in three different business situations: dinner conversation with a guest from an overseas company for business entertainment, negotiation with an overseas company, and an in-house meeting. The first two situations are both communications with someone outside the company, but the first one is a more informal situation, and the other is a more formal business situation. The three situations are intended to examine if there is any difference in response between communications outside and inside the company, and between informal and formal business situations. In either situation, respondents are asked to assume that they are having oral communication, which requires more direct and immediate response in face-to-face

communication; the study presumes that this type of communication involves more sensitivity to courtesy than email communication.

Question 6 asks whether respondents would rather study the target language and use it in words of their own than using machine translation. It also asks if they would rather use a human translator for business even when it requires higher cost.

3.   Results

Figures 1 through 6 below show the percentage of responses per each of the 5 Likert items. Responses to 1. Strongly disagree and 2. Somewhat disagree can be interpreted as "negative" responses with varying degrees of negativity. Likewise, responses to 4. Somewhat agree and 5. Strongly agree can be taken as "positive" responses. Thus, in the data analysis below, some results are presented as negative or positive responses based on these combinations.

3.1 General attitude

Responses to Question 1 reveal that in general respondents have favorable attitude toward the use of machine translation at workplace; more than 80% of respondents rated 4 (Somewhat agree) or 5 (Strongly agree) on Likert scale (See Figure 1 below). In particular, the lowest rate on 1. Strongly disagree (2.6%) and the highest rate on 5. Strongly agree (44.7%) suggest that the use of machine translation in business is well supported by Japanese office workers.
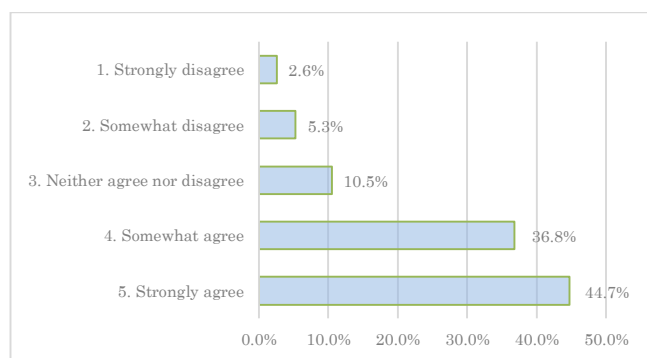

Figure 1. Responses to Question 1: General attitude toward machine translation.

If respondents respond negatively to more specific questions in the subsequent sections, it could mean that it is not because of their general attitude toward the use of machine translation, but due to the specific situation presumed in each question.

3.2 Attitude toward messages lacking naturalness and politeness

Figure 2 below reveals that 60.5% of respondents rated negatively (1. Strongly disagree or 2. Somewhat disagree) to the statement that there is no problem with messages lacking politeness and naturalness.

In addition, the statement about lack of accuracy in grammar received slightly more negative responses (65.8%) and that of accuracy in meaning received even more negative

feedback (81.6%). As can be easily predicted, it shows that accuracy in meaning is the highest priority in business email among the choices.
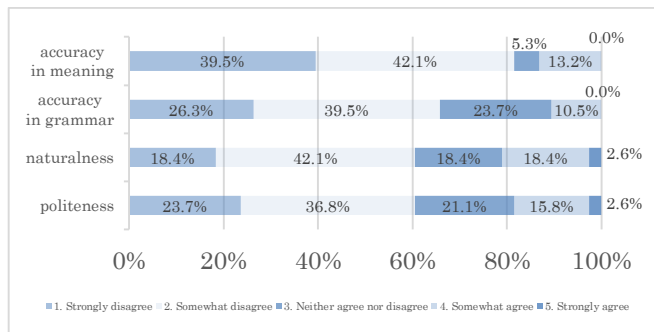

Figure 2. Result of Question 2: Attitude toward messages lacking accuracy in meaning, accuracy in grammar, naturalness, politeness.

3.3 "This message has been translated by machine"

When respondents were asked to assume that a translated message is supplemented with the statement "This message has been translated by machine," their attitude toward messages lacking naturalness and politeness seem to have changed.
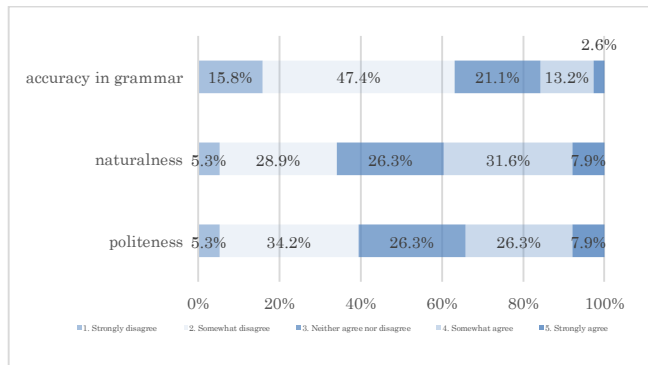

Figure 3. Result of Question 3： Attitude toward a machine-translated message when it is explicitly explained.

As you can see in Figure 3, 34.2% of respondents gave negative responses (rated 1 or 2) to messages lacking naturalness and 39.5% of them responded negatively to messages lacking politeness. Compared with the results from Question 2 (60.5%), their negative responses were weakened. Moreover, respondents' positive attitude (rated 4 or 5) increased from 21% (Question 2) to 39.5% (Question 3) for naturalness, and from 18.4% to 34.2% (Question 3) for politeness. This implies that their attitude toward messages lacking naturalness and politeness change when they know that the message explicitly explains that it has been translated by machine.

On the other hand, in the case of accuracy in grammar, the degree of change was not as much. Negative responses only slightly decreased from 65.8% (Question 2) to 63.2% (Question 3), and positive responses increased from 10.5% (Question 2) to 15.8% (Question 3). This suggests that explaining the use of machine translation is less helpful when

it comes to accuracy in grammar. Given the result from Question 2, this probably also applies to accuracy in meaning.
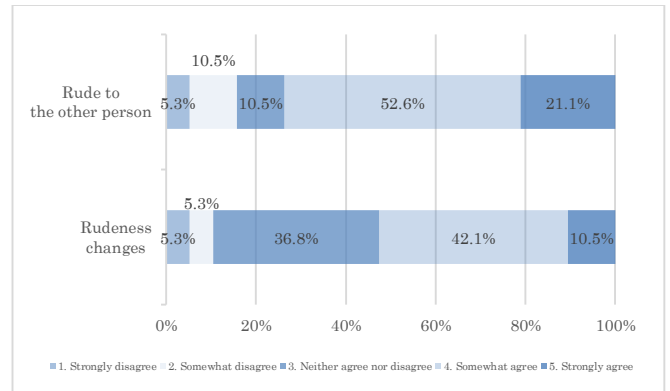
3.4 Rudeness

Figure 4. Result of Question 4: Rudeness.

As shown in Figure 4, 73.7% of respondents agreed (rated 4 or 5) and 15.8% disagreed (rated 1 or 2) with the statement that a message lacking "accuracy in grammar," "naturalness," and "politeness" is rude. Overall, a significant percentage of respondents seem to associate such a message with rudeness.

Moreover, 52.6% of respondents agreed with the idea that the degree of rudeness changes according to the social status and work relationship with the other person while only 10.6% disagreed. Although 36.8% neither agreed nor disagreed, which should be noted for closer investigation, overall more than half of respondents supported the idea.

3.5 Different situations

As indicated in Figure 5, the use of machine translation for in-house meetings received the most positive response (73.7%), negotiation the second (52.6%), and dinner conversation the least (44.7%).

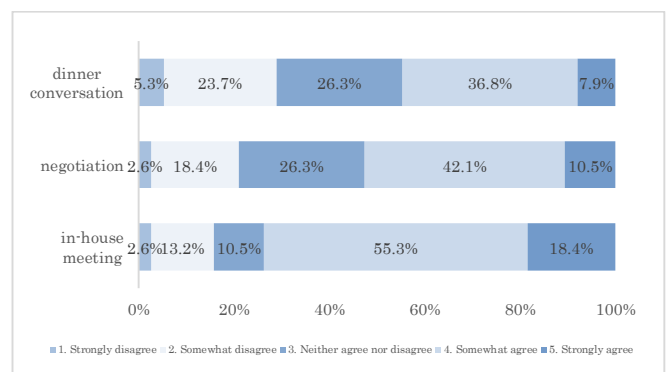Figure 5. Result of Question 5: Different business situations.

As predicted, respondents gave more favourable responses to the use of machine translation for internal communication than external. In the comparison between informal (dinner conversation) and formal (negotiation) business situations, the former received less positive response to the use of machine translation. Respondents think that using machine

translation is less acceptable at the dinner table. Dinner conversation in business is informal on one hand, but its purpose is often to facilitate business. Therefore, while speaking informally, they also have to be courteous so they can successfully establish and maintain business relationships. In the respondents' perception, negotiation is a formal business situation, but it may require only business-like conversation. Social events such as a dinner may require more attention to courtesy; and therefore, using machine translation is less acceptable.

3.6 Learning the target language / Human translator

Figure 6 shows that 76.3% of respondents think (rated 4 or 5) that learning the target language and using it for business is more desirable than using machine translation even when it conveys the message accurately.

As for the question about human translators, 44.8% of respondents responded positively, and 21.1% of them gave negative response.
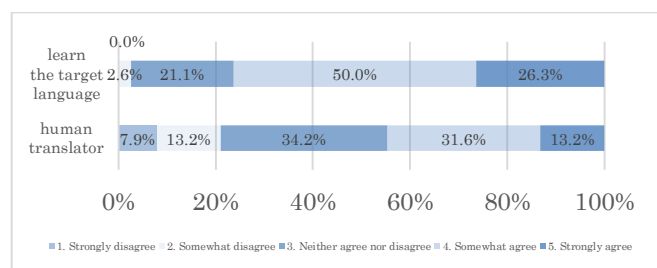

Figure 6. Result of Question 6: Machine translation vs. learning the target language / using human translators.

4. Discussion

As a pilot study, the current investigation provides meaningful data because the results suggest that, in addition to translation accuracy, naturalness and politeness can also be important variables to investigate how the use of machine translation is accepted by Japanese office workers. This implies that translation accuracy is not the only determinant of acceptability; solely discussing how successfully machine translation technology advances for translation accuracy is not enough to predict how it will be accepted by the society. The current study proposes that aspects such as naturalness and politeness should be investigated in more detail in order to understand how machine translation can be more widely accepted in the future.

In addition, there are three more important implications. One is that human relationships should be the center of discussion when we analyze naturalness and politeness; the second implication is the value of "authentic" communication; and the third is how we should address the issue of acceptability to foreign language education. The three implications will be discussed below.

4.1 Human Relationship as a Determinant of Acceptability

The responses to Questions 2 through 5 suggest that human relationships are involved in determining whether the use of machine translation is acceptable at workplace. One piece of evidence is the fact that the degree to which a message lacking naturalness and politeness is acceptable changes depending on the social status and work relationship of the other party and on the business situations where conversations take place. A decision on whether a message can be accepted could come from the sender's concerns about how the message sounds to the other party. In Japanese culture, people often change the wording according to the hierarchical relationship; it is an important social norm to consider whether the wording is appropriate for the hierarchical relationship. Therefore, human relationships can always be an important determinant of acceptability in Japan as to how naturally and politely the message is worded in order to avoid violating the social norm.

However, the result from Question 3 also shows that the violation of this social norm can be somewhat avoided by adding the statement "This message has been translated by machine." Providing this kind of statement could serve as preparation for a possible threat of violating the social norm. By doing so, the sender of a message can assume that this "heads-up" will alleviate the impact of such a threat. When both parties mutually understand the fact that their communication is done through machine translation, some unnatural and impolite language may become permissible. This suggests that one of the keys to increasing acceptability of machine translation is dealing with how we can alleviate possible threats to violate the social norm about the hierarchical relationship.

4.2 Value of "Authentic" Communication

The answers to the first question of Question 6 reveal that a significant percentage of respondents (76.3% on 4 or 5) think that it is desirable to learn the target language rather than use machine translation even if it can translate a message accurately. If translation accuracy is not an issue, then Japanese office workers place value on something other than accuracy in the act of learning a language and using it in their own words. One possible reason is that they regard using their own words in messages or conversation as an "authentic" way of communication. If this is true, using machine translation is taken as an "inauthentic" way of communication, and this perception may prevent Japanese office workers from accepting the use of machine translation.

The result from the second question of Question 6 also indicates this trend although the responses are more scattered from 1 to 5 (44.8% on 4 or 5; 34.2% on 3; 21.1% on 1 or 2). One possible explanation is that Japanese office workers may perceive that using a human translator is less "authentic" than speaking in their own words since they are getting assistance from others; but at the same time it is more "authentic" than using a machine, which is not even a person. This perception may have made responses scattered.

The above discussions imply that the perceived "authenticity" about communication should be addressed in order to predict how machine translation will be accepted in the future. In addition to naturalness and politeness, the aspect of authenticity could be another important factor that determines acceptability.

This viewpoint is also important in foreign language

education. With more advanced translation technology, producing natural and polite translations could be possible in the future. However, the perception of authenticity is not a matter of translation performance, but it is an issue of how we accept a certain mode of communication. Thus, machine translation users should understand the mode of communication using machine translation is only appropriate in certain occasions, and it should not be used in other situations. As we expect more people to use machine translation in society, it is important to teach in a foreign language classroom that in some situations the use of machine translation can be taken as inauthentic and thus not accepted as an appropriate mode of communication.

4.3 Educational Implications

In fact, some scholars have started to discuss the impact of machine translation on foreign language education. Asano (2018) discusses that in the age when machine translation (and artificial intelligence in general) is widely used in society, foreign language education should change its focus from merely promoting students' mastery of practical language skills to helping them gain wisdom necessary to lead a harmonious life with people from other countries in a global society. As another example, the JACET SIG on English Education held a conference to discuss issues on foreign language education with the awareness of the age when machine translation can be practically utilized in society (Sakai, 2019).

Some discussions on the educational implications of machine translation such as the above attempt to predict whether machine translation will be widely used in the future to the degree that foreign language education will disappear or change its focus. Such attempts at prediction are important for foreign language educators and language policy makers because if education needs to change in line with the change in society, it must be based on the judgment that the society has widely accepted the use of machine translation.

The results from the present study suggest that more empirical studies are necessary to make such judgement. Naturalness and politeness are two variables investigated in this study, but there may be more variables (other than translation accuracy) involved in the acceptability of machine translation at workplace. In addition to paying attention to the technological development that increases translation accuracy, it is important to empirically investigate other factors that affect acceptability.

It is especially important to address English education, which is the major foreign language education in the Japanese educational system from elementary to higher education. When machine translation becomes an essential part of global business communication at workplace in the future, learning English may become less valuable as qualification for office workers; and thus, we will need to reevaluate the way English education is conducted. In this sense, the most important question here is whether machine translation can really be accepted by Japanese office workers when it reaches a satisfactory level of translation performance. To answer this question, empirical studies should provide data that allow us to judge *when* and *how* English education should change.

5. Limitations and Future Research

Although the current study provides significant data as a pilot study, it also has some limitations. First, the response rate (3.8%) is too low to assume that the data represent all the Japanese office workers. Future studies should significantly improve the data collection process.

Second, the current study asked respondents to assume hypothetical situations when they answer. For example, in Question 2, respondents had to assume or imagine that they were using machine translation when they responded to the question. If they had never used machine translation, they might have responded differently. Future studies should provide actual machine-translated messages so that respondents can answer to questions in more concrete and realistic situations.

Third, the responses may change in the future, as translation accuracy improves and as more people use machine translation in the future. Therefore, with revised questions based on this pilot study, a longitudinal study should be employed to investigate the acceptability of machine translation.

The present study tested some presumptions, but each of the findings in this study must be investigated in more detail in future research. It is hoped that the current study will stimulate interest in similar research and encourage more detailed investigation.

6. Conclusion

The present study investigated how machine translation is accepted by Japanese office workers. The results from the online questionnaire survey suggested that naturalness and politeness, which are closely associated with human relationships, can be important determinants of acceptability. The study also suggested that the perceived "authenticity" of communication can be another possible determinant. The paper also discussed empirical investigation on the acceptability of machine translation such as the current study can contribute to our judgement about when and how foreign language education should constructively change.

As discussions on improving machine translation accuracy progress, it is important to question whether accuracy alone is a determinant of the acceptance of machine translation. As a pilot study, the current investigation provided meaningful data that answer that question.

References

Aiken, M. & Ghosh, K. (2009). Automatic translation in multilingual business meetings. *Industrial Management & Data Systems, 109*(7), 916–925.

Asano, K. (2018). Foreign language education in Japan in the era of artificial intelligence. *Journal of Nanzan Junior College, 39*, 95–105.

Brown, P. & Levinson, S. (1987). *Politeness: Some universals in language usage.* Cambridge: Cambridge University Press.

Hanari, T. (2019). Kkikaihonyakuto poraitonesu: Kikaihonyakuni haneisaserubeki poraitnesuto sonoshuhonikansuru ichikosatsu [Machine translation and politeness: A discussion on the

politeness that should be reflected in machine translation and its research method]. In Takita, & Nishijima (Eds.), *Machine translation and future society: Will we see a world without language barriers in the future?* (pp. 53–89). Tokyo: Shakaihyoronsha.

Institute of International Business Communication, T. (2013). *Jojokigyoniokeru eigokatsuyo jittaichosa hokokusho* [Survey report on use of English in listed companies]. Tokyo: The Institute of International Business Communication.

Inoue, T. (2017). *Jinko cho chino: seimeito kikaino aidani arumono* [Artificial super intelligence: Entity between life and machinery]. Tokyo: Shuwa System.

Ministry of Justice, T. (2018). Comprehensive measures for acceptance and coexistence of foreign nationals. December 25, 2018. Tokyo: The Ministry of Justice.

Sakai, S. (2019). Jidaino henkani tachiaite [Witnessing the changing times]. *Proceedings of Language Education Expo 2019* (gengo kyoiku expo 2019 yokoshu), 1.

Shimohata, S. (2005). Machine translation tools for daily tasks. *The Journal of Information Science and Technology Association, 55*(8), 331–334.

Takita, Y. & Nishijima, Y. (Eds.). *Machine translation and future society: Will we see a world without language barriers in the future?* Tokyo: Shakaihyoronsha.

Usami, M. (2001). Poraitonesuriron karamita keiihyogen: dokoga konpontekini kotonaruka [Respectful expressions from the perspective of politeness theory: What is fundamentally different]. *Language (Gengo), 30*(12), 18–25.

Yoshikawa, K. (2019). Tagengono chikujitsuyaku kansei tsugiwa koseidono dojitsuyaku [Completion of multilingual consecutive interpretation and next phase of development for high-precision simultaneous interpretation]. *Nikkei-Science, October Issue*, 9–11.

Appendix A: Survey Questions
Respondents are asked to rate their opinions based on the following Likert items: 1. Strongly disagree; 2. Somewhat agree; 3. Neither agree nor disagree; 4. Somewhat agree; 5. Strongly agree.

Question 1:
I would like to use machine translation in business if it translates what I want to say correctly.

Question 2:
(1) I have no problem with lack of "accuracy in meaning" in machine translation because in business it is enough to convey general messages.
(2) I have no problem with lack of "accuracy in grammar" in machine translation because in business it is enough to convey accurate messages.
(3) I have no problem with lack of "naturalness" in machine translation because in business it is enough to convey accurate messages.
(4) I have no problem with lack of "politeness" in machine translation because in business it is enough to convey accurate messages.

Question 3:
(1) I have no problem with lack of "accuracy in grammar" if the translation is supplanted with the statement "This message has been translated by machine."
(2) I have no problem with lack of "naturalness" if the message is supplanted with the statement "This message has been translated by machine."

(3) I have no problem with lack of "politeness" if the message is supplanted with the statement "This message has been translated by machine."

Question 4:
(1) Any machine-translated message that lacks "accuracy in grammar," "naturalness," and "politeness" is inappropriate (rude) even when the message is correctly conveyed.
(2) Whether the machine-translated message that lacks "accuracy in grammar," "naturalness," and "politeness" is inappropriate (rude) or not depends on the social status and work relationship of the person to whom you send the message.

Question 5:
(1) It would be acceptable to use machine translation for oral communication with a guest from overseas companies at the dinner table, when the machine translation is accurate enough not to interfere with any of your operations.
(2) It would be acceptable to use machine translation for negotiation with other companies, when the machine translation is accurate enough not to interfere with any of your operations.
(3) It would be acceptable to use machine translation for in-house meetings, when the machine translation is accurate enough not to interfere with any of your operations.

Question 6:
(1) Even when the machine translation is accurate enough not to interfered with any of your operations, it is desirable that you (or your colleagues) study the target language and communicate in words of your own.
(2) Even when the machine translation is accurate enough not to interfered with any of your operations, it is desirable that you use human translators despite higher cost.

ライティング意識の質的分析に基づく教材案

仲川　浩世*・幸重　美津子**

*関西外国語大学短期大学部
hgaeru@yahoo.co.jp
**学習院大学
yuki@balloon.ne.jp

要旨
本研究の目的は、学習意欲の低い大学生を対象に、ライティングの苦手意識を軽減するための実践研究を行い、その結果を踏まえて今後の教材案を示唆することである。リメディアル・レベルや上級者を対象としたニーズアナリシスや指導法に関する研究は、すでに行われている。しかしながら、学習意欲の低い学習者のための、ライティング意識に関する研究はまだ少ない。そこで、本研究ではフィードバックを伴ったライティング指導モデルを構築し、対象学生の振り返りとインタビューを質的に分析して、ライティング意識の変容を分析した。その結果、当該学習者は、ライティングの苦手意識を軽減させたことが明らかとなった。また、自律してライティング活動に取り組む姿勢もみられるようになった。本研究では、学習者のさらなる自律性育成の一助となる教材の可能性を提案する。

キーワード：ライティング、学習意欲、教材案、自律性、機械翻訳

1. はじめに

　日本人大学生にとって、入学前のライティング学習不足による苦手意識が問題視されている（保田・大井・板津, 2014; Okada, 2018）。高校では、文法・リーディングが重視されているが、大学では「自分の考えを論理的に書く」という能力が求められる。そのため、ライティングに対して、苦手意識を持つ学習者の増加が顕著となった（ベネッセ教育総合研究所, 2016）。このライティングの苦手意識が英語嫌いへと結びつき、学習意欲の低下の要因となったとも考えられる。

　一方、リメディアル教育（TOEIC　202.78 点程度）（牧野, 2017）や ESP・EAP 学習者に対する研究（石川・伊東, 2017）はすでになされている。しかしながら、学習意欲の低い学習者のライティング意識に関する調査はまだ少ない。そこで本稿では、当該学習者のための指導モデルを構築し、ライティング意識の変容を探るために、学習意欲の低い大学生を対象とした実践研究を実施し、その意識の変容を中心にした教材の可能性について論じ、今後の教材案を示唆することを目的とする。

　本研究では、指導モデルにフィードバックを導入した。これは、第二言語ライティング指導におけるフィードバックの有効性が実証されているからである（Bitchener & Storch, 2016）。本研究で示すフィードバックとは、ライティングに対する言語的な誤り訂正、及び内容や構成に関するコメント（Storch, 2010）と、口頭の教員による学習者への発話の言い直し（Lyster, Saito & Sato, 2013）と定義づける。

　また、ライティング後に振り返り（Reflection）をさせ、その日の学習活動を内省させた。振り返りとは、Farrell（2017）を参考とし、「学習者が自分のライティングを省みて、今後の活動へと発展させること」と定義づける。このようにして、調査協力者のライティング意識の変容を検証する。

2. 研究課題

　先行研究から得た理論的考察を踏まえて、以下に研究課題を 2 つ設定した。
　1）　指導モデルは調査協力者のライティング意識にどのような影響を与えるか。
　2）　ライティング意識の変容から、どのような教材案が示唆されるか。

3. 研究方法
3.1.　調査協力者

　調査協力者は、関西圏の私立外国語大学の短期大学生 13 名（男 6 名、女 7 名　TOEFL ITP テスト　380～440 点）である。2017 年度 2 月末から 3 月までのライティングの集中講座登録者であった。本講座は、1 年次の春学期の単位未履修者のため、英語の習熟度は中級未満にもかかわらず、学習意欲の低下という問題がみられた。

3.2.　実践概要

　講座は 11 日間、90 分授業を 30 回、ライティング指導モデルを用いた実践を 6 回繰り返した。図 1 は指導モデルのプロセスである。指導モデルは 5 段階に分け

られる。第 1 に、社会問題を扱ったリーディング・テキストを読ませた。第 2 に、英語で Rephrasing（自分の言葉で言い直し）をさせた。その際に、口頭でフィードバックを与えた。第 3 に、個人によるプロセス・ライティングに取り組ませた。まず、命題を与え、クラス全体でbrain-storming（Raimes, 1983）を行い、アイデアを共有させた。そして、完成させたパラグラフを提出させ、言語的な誤り訂正を提供した。第 4 に、その日の学習活動について、日本語で振り返りを書かせた。そして、これに対し、教員は質問の回答や励ましのコメントを与えた。すなわち、振り返り（提出物）を通して、コミュニケーションの促進を試みたのである。最後に、教員は提出物返却後、学習者を促して、自分のライティングを見直させた。このようにして、自律した書き手育成を目指したのである。

```
1) Reading
2) Rephrasing ←Oral CF
3) Process writing←WCF
4) Reflection←L1 によるコメント
5) 提出物の見直し（教員のコメントによる活動の内省）
```
図 1. 指導モデルのプロセス

また、図 2 は集中講座のスケジュールである。初日にオリエンテーションとプリテスト（30 分/社会問題をテーマにした、パラグラフ・ライティング）を実施した。尚、本研究の全てのライティングテストでは、辞書の使用を許可してはいない。

そして、2 日目から 5 日目までの間に、毎日指導モデルを繰り返し、6 日目に中間テスト、7 日目と 8 日目にも指導モデル、9 日目にポストテスト（プリテストと同様のもの）、10 日目に期末テスト、11 日目には、プレゼンテーションと統一テスト（マークシート）を行った。

```
〈実践スケジュール〉
Day 1      オリエンテーション・プリテスト
Day 2-5    指導モデル
Day 6      中間テスト
Day 7-8    指導モデル
Day 9      ポストテスト
Day 10     期末テスト
Day 11     プレゼンテーション・統一テスト
```
図 2. 実践スケジュール

4. 結果
4.1. ライティング能力の結果
ライティング能力の伸びは、①総語数と②Test of Written English（以下 TWE とする、満点 5 点）によって、判断した。TWE は、5 点「首尾一貫したパラグラフ」、4 点「内容の理解を妨げる誤りを含んでいる」、3 点「文構造における誤りが顕著にみられる」、2 点「文レベルの誤りが顕著にみられる」、1 点「書き手が命題を理解してはいない」という基準によって成り立って

いる。

その上、プリテストとポストテストの TWE と総語数の平均値の差を有意水準 1%で、両側検定の t 検定により検討した（三浦他, 2004）。その結果、どちらも有意に向上した（表 1）。また、総語数の結果は、13 名中 2 名のみが、プリテストと比較すると、ポストテストでは下がってしまった（図 3）。

表 1. 調査協力者のライティングの伸び（N = 13）

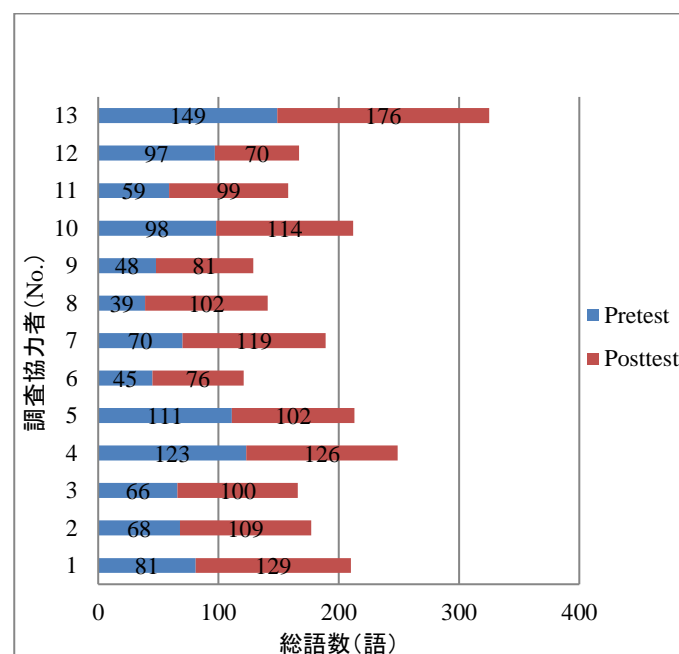|	Pretest		Posttest			
	M	SD	M	SD	T	P
TWE	2.13	0.77	2.49	0.64	4.73	.00**
総語数	81.08	32.89	107.92	27.33	3.85	.00**

注：*p < .05,　**p < .01

図 3. プリテストとポストテストの総語数の変化

4.2. 意識の変容の結果
次に、調査協力者のライティングの意識の変容について言及する。テキストマイニング（樋口, 2014）によって、学習者の振り返りを共起ネットワークで示す。また、指導後に上位者と下位者の学習者に協力を求め、インタビューも行った。本研究では、1 日目のプリテストと、中間テストの前日の 5 日目と、10 日目のポストテスト後のみを示す。
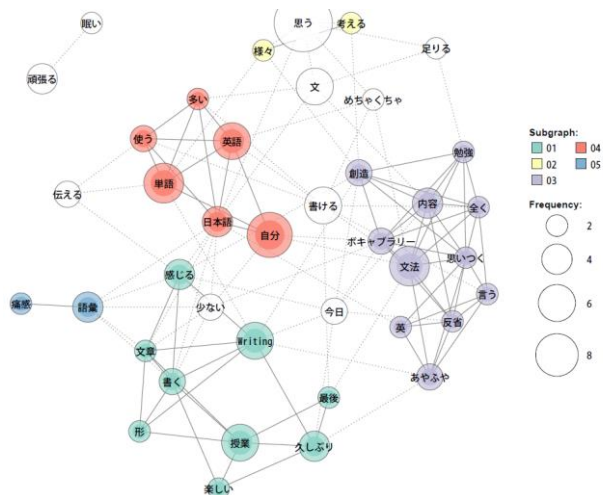
図 4. Day 1 の調査協力者の共起ネットワーク（*N* = 13）
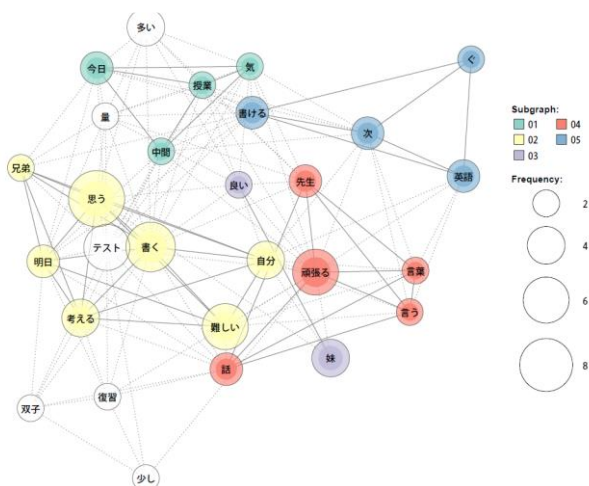
図 5. Day 5 の調査協力者の共起ネットワーク（*N* = 13）
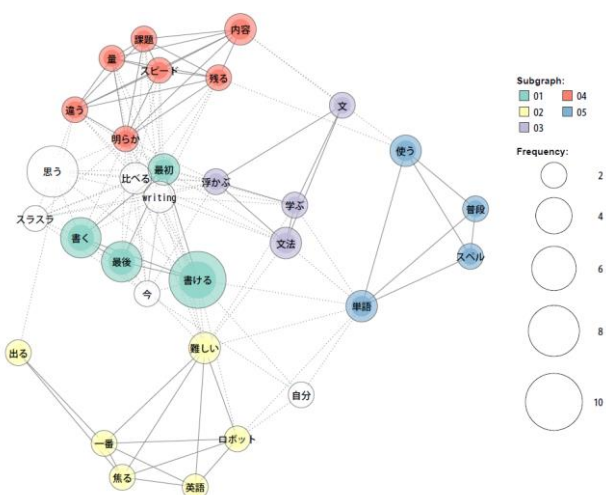
図 6. Day 10 の調査協力者の共起ネットワーク（*N* = 13）

　図 4、図 5、図 6 を比較してみると、以下のことが明らかとなった。図 4 の 1 日目は「単語」「あやふや」「久しぶり」と、ライティングに対する自信のなさが見受けられた。また、習熟度別に分析してみると、上

位者は「アイデアが不足している」「ライティングの書き方を理解していない」と述べ、下位者は「文法自体があやふや」「勉強が足りない」と英語力のなさを示している。
　次に、図 5 を参照し、5 日目について論じる。この日は中間テストの前日ということもあり、「書く」「テスト」「頑張る」というキーワードが顕著にみられた。例えば、上位者は「次の授業は中間テストなので気を引き締めて頑張る」と述べ、下位者は「明日の中間では 10 行は超えたい」と、両者ともテストに向けて意欲を向上させていたようであった。したがって、テストに向けて取り組むことから、ライティングに対する苦手意識が、少しずつ軽減されたことがわかる。
　そして、図 6 の 10 日目は、「書ける」「浮かぶ」「スピード」とライティングの上達を暗示する肯定的な語句が目立つようになった。上位者は「普段からもっとスペルを意識して取り組みたい」「最初と比べて書けるようになっている」と述べ、下位者は「どうしたら文を増やせるのか」「短文ばかり書いている」と自分の問題点に気づき始めた。
　さらに、上位者と下位者に協力を求め、インタビューからライティング意識の変容を探ったところ、以下のことが明らかとなった。

〈ライティングについて〉
　上位者→自分のペースだし、日々やってて、回数を重
　　　　　ねるごとに上手く書けるようになった。
　　　　　長く書けるようになり、紙が 2 枚目とかに
　　　　　なってできるようになった。

　下位者→自分のようにライティングにつまずいてい
　　　　　る場合、残りを家でもう少し練って、改めて
　　　　　次の時間に出したい。宿題にしてほしい。

〈振り返りについて〉
　上位者→質問を書いたら、文法や構文の使い方など、
　　　　　先生が答えてくれたので、よかった。

　下位者→書くことは大体同じだが、続けることで自
　　　　　分の意識づけになってよかった。

5. 考察
　4 の実践の結果を踏まえて、研究課題を考察する。まず、研究課題 1)「指導モデルは調査協力者のライティング意識にどのような影響を与えるか」においては、調査協力者は、概ねライティングへの意欲を向上させたということがわかった。さらに、少人数のためより授業がインタラクティブとなり、協同学習の効果もみられ、動機づけ向上にも寄与した(図 7 参照)。
　本研究で述べる、協同学習の利点を Chen（2018）は、「学習者が少人数のグループで協力することから、教育者は、授業内の学びを改善させ、学習者同士の知

識の共有を支援する」と指摘している。

図 7. 指導モデルの効果

　また、習熟度別に検討してみると、上位者の学習者の方がライティング活動を講座開始当時から、楽しみながら取り組んでいた。対照的に、下位者の学習者は苦手意識が高く、徐々に自分の問題点に気づき始めた。そして、インタビューでは、「授業外課題」を希望し、明確な目標設定をして、自律性を高めていったことが明らかとなった。

　次に研究課題 2)「ライティング意識の変容から、どのような教材案が示唆されるか」においては、学習意欲が低い学習者に特化した教材開発が必要であると言えよう。とりわけ、授業開始前のニーズ分析によって、学習者の習熟度や補足すべきスキルも検討しなければならない。

　また、下位者のコメントから、「授業外課題」の導入も考えられる。上位者に比べて、ライティングの苦手な下位者は、楽しみながらゲーム感覚で学ぶ教材を与えることで、苦手意識が軽減され、前向きに取り組む可能性もあると考えられる。
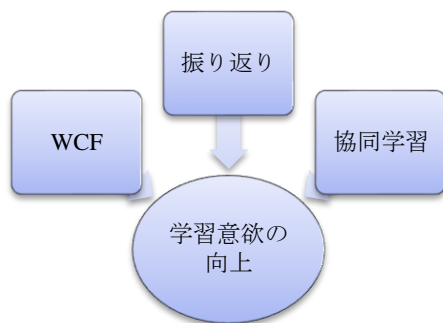
　先行研究では、学習者は、繰り返し活動を行っているうちに、自分の問題点を解決し、コミュニケーション能力を発達させると伝えられている。言い換えれば、教員の介入が少ないほど、自律した書き手になると報告されている（Chen, 2018）。したがって、学習者が最終的には、自ら取り組むことのできる教材を考案することが求められよう。

6. 今後の課題
　今後の課題として、以下の 3 点が考えられる。
　1) 個々の学習者に見合った教材の開発
　2) 授業外課題の工夫
　3) 自律した学習者育成のための指導実践

　まず、同じ習熟度のクラスにおいても、学習意欲によって、ライティングの伸びは変容する。そのために、より綿密な質的分析が必要となる。よって、学習者とのコミュニケーションを欠かさず、常に必要な副教材を準備しておくことが望ましい。

　また、下位者の調査から、授業外課題の必要性が明らかとなった。現在のところ、授業外課題については、統一カリキュラムなどとの兼ね合いという評価面に関する問題点がある。それゆえ、早急な導入は困難である。そのため、担当教員の裁量で補足できる程度に、課題を提供することを考慮すべきであろう。その結果、学習習慣の定着も想定できる。

　最後に、自律した学習者の育成のための指導については、最も調査が必要である。ある限られた指導を特定の学習者集団に提供しても、必ずしもうまくいくとは限らない。したがって、パイロット的に実践研究を繰り返し、教材開発と共に新しい指導法を探求する必要がある。

7. 教育的示唆：教材案
　本研究から得られた教育的示唆として、授業外課題の導入が具体案として考えられる。現在でも、Call 教室やタブレットを用いた教材が、英語教育現場に普及しつつある。しかしながら、施設やコストの問題で、まだ完全に定着しているとは言い難い。今後は、学習者が授業外に、自分のペースで、オンラインやスマホのアプリを通して、ライティング学習ができることを目指したい。

　学習のモチベーションと自律性を高めるために行ったパイロット授業では、授業外学習のサポートとなる MT（Machine Translation）アプリの導入が有効であった。ライティング授業のアサインメントに対して、MT を一定の指示通りに使用させたものであるが、コース終了時の質問紙調査（無記名）では、その「利便性」よりも「学習効果」に刺激を受けた学生が過半数を占め、さらに「自分で書く自信」につながったことが分かった。下位学習者にとっては、自分にもできるという期待が誘因（内発的動機）となり、評価・提出物（外発的動機）に辿り着くための facilitator としての役割を果たしたと考えられる。上位学習者にとっては「単語や語句の正しい使い方を知れて表現の幅が広がった」という報告が目立った。いずれの場合も、「自分で発信する自信」がついたという、積極的な発信に向けてのモチベーションの向上が見られたことが大きな変化であった。

　さらに自律学習を促進するためには、英語学習に資する OCW（Open Course Ware）を紹介したり、MOOCs（Massive Open Online Courses）を授業外の課題や反転授業として利用することもできる。そのほか、授業に参加できない学習者のための課題としてもその可能性は限りない。いずれの場合も、ただ知識として紹介するだけでなく実際に、実践させるといったイニシアチブをとる必要もあるだろう。SNSs などのサイバー環境に対する抵抗感が限りなく少ない昨今の学習者たちにとっては、AI（Artificial Intelligence）を活用した MT や、オンライン学習こそ学習意欲・自律性向上の一助になると期待される。今後、これらを活用した教

材開発に取り組む側と連携した実証的な研究が必要となるだろう。

8. まとめ

本研究は、学習意欲の低い学習者のライティング意識の変容をテキストマイニングによって、質的に分析し、それを基にライティング教材案を提案した。調査の結果から、自律した学習者育成のための授業外の課題や動機づけ向上となるライティング教材が望ましいということが明らかとなった。その一例として、オンライン学習や機械翻訳の導入の可能性を示唆した。

しかしながら、本研究の調査協力者数は 13 人と限られている。今後、より多くの学習者を対象とした実証的なデータを継続的に収集し、特定の学習者集団のニーズを調査することによって、より具体的な教材開発を実施する必要があるだろう。

参考文献

ベネッセ教育総合研究所. (2016). 『中高の英語指導に関する実態調査 2015』 Retrieved from
https://berd.benesse.jp/global/research/detail1.php?id=4776

Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development*. Bristol: Multilingual Matters.

Chen, Y. (2018). Perceptions of EFL college students toward collaborative learning. *English Language Teaching*, *11*, 1-4. doi: 10.5539/elt.v11n2p1

江利川春雄. (2012). 『協同学習を取り入れた英語授業のすすめ』大修館書店.

Farrell, S.C. (2017). *Research on reflective practice in TESOL*. New York, NY: Routledge.

樋口耕一. (2014). 『社会調査のための計量テキスト分析』ナカニシヤ出版.

石川有香・伊東田恵. (2017). 「ESP 教育対象としての「工学系英語学習者」－工学英語語彙学習の観点から―」『中部地区英語教育学会紀要』46, 253-260.

Kobayakawa, M. (2011). Analyzing writing tasks in Japanese high school English textbooks: English I, II, and Writing. *JALT Journal*, *33*, 27-48.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46,* 1-40.

牧野眞貴. (2017). 「リメディアル学生のニーズに基づく英語授業の検討とその実践」『LET 関西支部研究集録』16, 73-85.

三浦省五. (監修). 前田啓朗・山森光陽(編著). 磯田貴道・廣森友人(著). (2004). 『英語教師のための教育データ分析入門―授業が変わるテスト・評価・研究』東京：大修館書店.

Nassaji, H. (2017). Negotiated oral feedback in response to written errors. In H. Nassaji & E. Kartchava (Eds.), *Corrective feedback in second language teaching and learning* (pp. 114-128). New York: Routledge.

Okada, R. (2018). Challenges in teaching Japanese EFL students to express themselves logically. *Asian Education Studies*, *3*, 73-82. doi:10.20849/aes.v3i1.335

Raimes, A. (1983). *Techniques in teaching writing*. Oxford University Press.

Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, *64*, 103-131.

Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies*, *10*, 29-46.

Storch, N., & Wigglesworth, G. (2010). Learners processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition*, *32*, 303-334.

田地野彰・寺内一・金丸敏幸・マスワナ紗矢子・山田浩. (2008). 「英語学術論文執筆のための教材開発に向けて―論文コーパスの構築と応用―」『京都大学高等教育研究』14, 111-121.

保田幸子・大井恭子・板津木綿子. (2014). 「日本の高等教育における英語ライティング指導の実態調査」*JABAET Journal, 18,* 55-78.

日本の小学校外国語科における文及び文構造の分析
―English Grammar Profile との比較から―

山田　浩*

*高千穂大学
h.yamada@takachiho.ac.jp

要旨
2017 年の学習指導要領改訂により、小学校では外国語活動が中学年に移行され、高学年では外国語科が新設された。文部科学省が公表した外国語科用の教材『We Can!』に関しては、現在のところ文法的観点からの研究は少ないが、学習指導要領には文及び文構造が明記されているため、文法面の分析も重要性が高まっている。そこで本研究では、学習者が CEFR の各レベルで身に付けている文法項目に関する情報を提供する English Grammar Profile と、学習指導要領及び『We Can!』を比較することにより、外国語科で扱う文法項目について検討した。分析の結果、A1 レベルの学習者が身に付けている文法項目と、学習指導要領及び『We Can!』に出現する文法項目には違いが見られた。この中には、小学生が身に付けることができ、かつコミュニケーションにおいて重要な役割を果たす文法項目も含まれていた。この結果から、外国語科における文法項目の取り扱いや指導上の留意点を考察した。

キーワード：小学校外国語教育、学習指導要領、We Can!、English Grammar Profile

1. はじめに

　2017 年 8 月、小学校の学習指導要領が改訂された[1]。これまでの日本の英語教育については、文法・語彙等の知識がどれだけ身に付いたかという点が重視され、英語によるコミュニケーション能力の育成を意識した指導、特に「話すこと」及び「書くこと」に関わる言語活動が十分に行われていないことや、習得した知識や経験を生かし、コミュニケーションを行う目的・場面・状況等に応じて適切に表現することなどに課題が指摘されてきた（文部科学省, 2017a）。新しい学習指導要領ではこれらの課題を踏まえ、他教科と同様に外国語教育においても、「知識・技能」、「思考力・判断力・表現力等」、「学びに向かう力・人間性等」の三つの資質・能力を有機的に関連づけながら、バランスよく育成することが求められている。

　例えば、英語によるコミュニケーションを支える土台となる語彙や文法の知識についても、言語活動と効果的に結び付けることによって、実際のコミュニケーションの中で児童生徒が思考・判断・表現する際に活用できるように指導することが必要である。そして、このような学習を繰り返すことにより、児童生徒に自信が生まれ、主体的に学習に取り組む態度が涵養されることが期待されている。

　このような教育を実現するために、教師はさまざまな手立てを講じる必要があるが、これまでのところ日本における小学校の外国語教育については、語彙的な側面に着目した研究が主流であり、文法的な側面に焦点を当てた研究が少ないという状況がある（本多・志村, 2017）。そこで本稿では、今回の改訂により新設された小学校外国語科において児童が学習すべき文法項目を検討することにより、小学校における文法指導の内容と方法について考察する。

2. 先行研究
2.1 小学校外国語教育の概要

　社会や経済の急速なグローバル化の進展に伴い、学校教育における外国語教育の充実が求められてきたことを受けて、文部科学省は 2008 年改訂の学習指導要領で、小学校において外国語活動を新設した。これ以前にも、総合的な学習の時間の中で外国語教育は行われていたが、目標や内容は各学校で定められていたため、教育の機会均等や中学校との接続において課題が見られた。そこで文部科学省は、国として各学校において共通に指導する内容を示し、一定のまとまりをもって活動を行うため、高学年に外国語活動を設置し、一定の授業時数（年間 35 単位時間、週 1 コマ相当）を確保することとした。この外国語活動の目標は次のとおりである（文部科学省, 2008）。

　　外国語を通じて、言語や文化について体験的に理解を深め、積極的にコミュニケーションを図ろうとする態度の育成を図り、外国語の音声や基本的な表現に慣れ親しませながら、コミュニケーション能力の素地を養う。

なお、この目標に基づいて行われる外国語活動は、教科のような数値による評価にはなじまないものとし、教科とは位置付けないことが適当とされた。こうして日本全国で行われるようになった外国語活動について、文部科学省は成果と課題を次のようにまとめている（文部科学省, 2017a）

> 小学校では、平成 23 年度から高学年において外国語活動が導入され、その充実により、児童の高い学習意欲、中学生の外国語教育に対する積極性の向上といった成果が認められている。一方で、①音声中心で学んだことが、中学校の段階で音声から文字への学習に円滑に接続されていない、②日本語と英語の音声の違いや英語の発音と綴りの関係、文構造の学習において課題がある、③高学年は、児童の抽象的な思考力が高まる段階であり、より体系的な学習が求められることなどが課題として指摘されている。
> また、小学校から各学校段階における指導改善による成果が認められるものの、学年が上がるにつれて児童生徒の学習意欲に課題が生じるといった状況や、学校種間の接続が十分とは言えず、進級や進学をした後に、それまでの学習内容や指導方法等を発展的に生かすことができないといった状況も見られている。

ここで挙げられた外国語活動の成果を維持しつつ、課題を改善するために、2017 年の改訂では小学校中学年から外国語活動を導入し、高学年では外国語科が新設された。中学年では「聞くこと」、「話すこと」を中心とした活動を通じて外国語に慣れ親しみ、外国語学習への動機付けを高めた上で、高学年から段階的に文字を「読むこと」、「書くこと」を加えて総合的・系統的に扱う教科学習を行い、中学校への円滑な接続が図られることとなった。

また、外国語科が教科として位置付けられたことにより、文字の認識や日本語と英語の音声の違いへの気付き、語順の違い等の文構造への気付きなど、言語能力向上の観点から言葉の仕組みの理解などを促す指導が求められるようになった。外国語活動は中学年の 2 年間で計 70 単位時間、外国語科は教科として高学年の 2 年間で計 140 単位時間が配当された。この教育課程は 2018 年度から 2 年間の移行期間を経た後、2020 年度から全面実施される。

2.2 小学校外国語教育の教材

2008 年の導入以降、外国語活動用の教材は文部科学省により作成され、全国の小学校に配布されてきた。最初に登場したのは『英語ノート』であり、2009 年度から 2011 年度まで活用された。2012 年度から 2017 年度までは『Hi, Friends!』が使われ、2018 年度からは外国語活動用に『Let's Try!』、外国語科用に『We Can!』が作成

された。なお、表 1 のとおり、2020 年度からは各社が小学校外国語科用の教材を出版する。

表 1. 東京都で採択された 2020 年度用小学校外国語科の教材一覧（東京都教育委員会, 2019 をもとに作成）

教材名	出版社名	採択数
NEW HORIZON Elementary English Course	東京書籍	19
Here We Go!	光村図書出版	11
ONE WORLD Smiles	教育出版	8
CROWN Jr.	三省堂	8
Junior Sunshine	開隆堂出版	3
JUNIOR TOTAL ENGLISH	学校図書	3
Blue Sky elementary	啓林館	2

現行の『We Can!』までの教材を対象にした研究はたくさん見られるが、語彙的観点から分析を行うものが多く、文法的観点から分析を行うものは少ないとの指摘がある (本多・志村, 2017)。語彙に関する研究では、小学校外国語教育で扱うべき語彙を、使用されている教材から明らかにしたり、客観的な語彙リストとの比較を通して検討したりするものが多い (神谷・長谷川・町田・長谷部, 2009; 中條・西垣, 2010; 西垣・中條・西岡, 2007; 本田・星加・田所, 2018 など)。他方、文法に関する研究では、小学校外国語教育用の教材に含まれる表現の中で使用される文法を明らかにするものは多いものの、客観的な文法項目リストと比較して検討を加えているものは見当たらない (仁科・藤原・松岡, 2009; 藤原・仁科・松岡, 2010; 本多・志村, 2017; 本田・星加・小川, 2008)。

このように、先行研究が語彙に偏っている理由の一つとしては、外国語活動の学習指導要領には語彙に関する記述がある一方で、文法に関する記述はないことが考えられる。ところが、外国語科の学習指導要領には、新たに文及び文構造という節が設けられており、学習すべき文法項目が規定されている。中学校以降の学習指導要領では、文及び文構造に文法事項が加えられ、学習すべき文法項目についてより細かく整理されている。この違いは、次の記述が表すとおり、小学校の外国語科においては示されている文及び文構造について、基本的な表現として指導することが意図されているためである（文部科学省, 2017a）。

> 文及び文構造については、（中略）言語活動の中で、文法の用語や用法の指導を行うのではなく日本語と英語の語順の違い等の気付きを促すようにしたり、基本的な表現として繰り返し聞いたり話したりするなどして活用したりすることが求められる。繰り返し触れることによって英語の語順に気付かせ、その規則性を内在化させたり、自ら話したり書いたりする中でどのように語と語を組み合わせれば自分の伝えたいことが表現できるのかということに意識を向けさせたりするようにする。

すなわち、小学校の外国語科においては、得意なものを紹介し合う活動の中で、児童が I am good at playing tennis. という表現に触れて、その意味を把握したり、自ら活用したりできるようになる指導が求められている。一方で、動名詞を文から取り出して用語の説明をしたり、変形の練習をしたり、用法を理解して異なる表現の中で活用したりすることは、中学校以降で指導することとされている。

　このように指導方法には違いがあるものの、小学校外国語科でも学習指導要領に文法項目に関する記述が登場している。小学校と中学校の円滑な接続という外国語科新設の目的を踏まえると、文法的観点から小学校の外国語教育を検討する研究の重要性が高まってくると考えられる。したがって本研究では、客観的な文法項目リストとの比較を通して、小学校外国語科で扱うべき文法項目を検討することを目的とする[2]。

3.　研究方法
3.1　分析対象
　小学校外国語科の学習指導要領における文及び文構造に記載のある文法項目、外国語科用教材の『We Can!』に出現する文法項目、English Grammar Profile（EGP: Cambridge University Press, 2015）に記載のある文法項目を分析対象とした。『We Can!』本体は絵や日本語での記述が多いため、本研究では『We Can!』本体に加え、教師用指導書（文部科学省, 2018a; 2018b）に記載のある音声スクリプトと指導者の表現例も分析対象とした。

　EGP とは、世界各国の英語学習者がヨーロッパ言語共通参照枠（CEFR）の各レベルで身に付けている文法項目をリスト化したものである。CEFR に基づいて文法項目を整理しているリストは、Core Inventory for General English（North, Ortega, & Sheehan, 2015）や CEFR-J Grammar Profile（投野, 2016）など、他にも存在している。ところが、これらのリストの元となるコーパスには、世界中の英語学習教材に記載されている英文が含まれており、その文法項目を実際に学習者が身に付けているかどうかは不確かである。一方で、EGP は世界中の英語学習者が産出した英語をコーパスとして利用しているため、CEFR の各レベルの学習者が実際に使用できる文法項目に関する情報を提供している。したがって本研究では、EGP を分析対象とした。なお、小学校の外国語教育の到達目標は CEFR の A1 レベルに設定されているため（文部科学省, 2016）、EGP の A1 レベルに記載のある文法項目を分析対象とした。

3.2　分析方法
　文法項目の整理には、さまざまな分類方法があるが、本研究は日本における小学校と中学校の円滑な接続を念頭に置いているため、小学校から高等学校までの学習指導要領に基づいて分類した（文部科学省, 2017a; 2017b, 2018c）。例えば、EGP では「There＋is＋〜」と「There＋are＋〜」を異なる文法項目として分類してい

るが、本研究では学習指導要領に基づいて「There＋be動詞＋〜」という同一の文法項目として整理した。

4.　結果と考察
4.1　全体結果
　学習指導要領、『We Can!』、EGP のどれか一つにでも出現している文法項目を抽出すると、全部で 31 の項目が抽出された（付録参照）。出現の有無について整理すると、表 2 のとおり 31 の項目は 4 パターン（I〜IV）に集約された。次節では、パターンごとに考察を行う。

表2. 文法項目の分類結果の概要

パターン	学習指導要領	We Can!	EGP	項目数
I	有	有	有	12
II	有	有	無	1
III	無	有	有	5
IV	無	有	無	13

4.2　各パターンの考察
　パターンIに該当する項目数は 12 であった。これらの項目は、学習指導要領と『We Can!』の両方に記載があり、EGP の A1 レベルにも含まれていることから、児童が学習する文法項目として適当であると考えられる。ただし、付加疑問文の取り扱いには注意が必要である。付加疑問文はパターン I に含まれる肯定文、否定文、疑問文の知識を必要とするが、小学校から高等学校までの学習指導要領には独立した文法項目としての記載はない。したがって、どの校種で指導するかは定められていないが、一般的には中学 2〜3 年生で学習することが多い。ところが、『We Can!』には出現しており、EGP では A2 レベルに記載があるため、児童に対して付加疑問文を指導する際には、後述するパターンIVと同様の配慮が必要となる。

　パターンIIに該当する項目は命令文のみであった。命令文は学習指導要領と『We Can!』の両方に記載されているが、EGP では A2 レベルに含まれている。教室内で命令文は教師からの指示の中で頻繁に活用されるため、児童には大量のインプットが与えられる。ところが、EGP では A2 レベルに記載されていることから、児童の中には命令文を理解することはできても、産出することには困難を感じる者が出てくる可能性がある。学習指導要領でも、文法指導の留意点として次のような記載がある（文部科学省, 2017a）。

> 　言語材料については、平易なものから難しいものへと段階的に指導すること。また、児童の発達の段階に応じて、聞いたり読んだりすることを通して意味を理解できるように指導すべき事項と、話したり書いたりして表現できるように指導すべき事項とがあることに留意すること。

　したがって、児童を評価する場面では、命令文を理解

できているかどうかを問うことにとどめ、命令文を産出できるかどうかは中学校以降での学習に回すことも考える必要があるだろう。

パターンⅢに該当する項目数は 5 であった。これらの項目は学習指導要領に記載はないが、『We Can!』には出現しており、EGP でも A1 レベルに記載されている。したがって、児童はパターンⅢに含まれる項目を身に付けることができる可能性がある。中学校以降の教師が、学習指導要領の記述のみから小学校の学習内容を把握する場合、児童はこれらの項目に触れたことがなく、中学校で初めて学習するという認識を持つかもしれない。ところが、児童は授業の中でこれらの項目を含む表現に触れており、中にはこれらの文法項目を身に付けている児童もいる可能性がある。そこで、同じ内容を同じ方法で繰り返し学ぶことにより生じる学習意欲の低下を避けるため、指導の際には配慮が必要である。また、このような接続の不具合を避けるために、パターンⅢの項目については学習指導要領に記載してもよいのではないかと思われる。

パターンⅣに該当する項目数は 13 であった。この中には、EGP の B1 レベルに記載されている項目や、分詞構文のように学習指導要領では高等学校で学習することになっている項目も含まれている。これらの項目は児童にとって理解することも難しいことが考えられるため、教師はジェスチャーや言い換えを使って児童の理解を促す工夫が必要である。それでもなお理解に苦労する児童が多い場合には、たとえ教科書に含まれていても、授業内では扱わないというような配慮も必要であろう。または、このような混乱を避けるために、あらかじめこれらの項目を教科書から省いてもよいかもしれない。

5. おわりに

2008 年に導入された小学校外国語活動の成果と課題を踏まえて、2017 年には外国語科が新設された。小学校と中学校の円滑な接続に向けて、外国語科の学習指導要領には文及び文構造という節が設けられていることから、本研究では小学校外国語科で学習すべき文法項目について検討した。本研究から得られた教育的示唆をまとめると、以下の 3 点である。

第一に、各校種の教師は、学習指導要領の記述に基づいて互いの学習内容を把握するだけではなく、実際に使用される教材にどのような内容が含まれているかを知ることも重要である。本研究では、『We Can!』を分析対象としたが、上述のとおり 2020 年度からは各社が教材を出版する。どの教材を採択するかは地域や学校によって異なるため、児童によって学習内容に差が生じる可能性がある。今後は複数の教材を比較して分析することにより、中学校以降の指導で学習内容の差を補償するような教育が必要になるだろう。

第二に、教師はジェスチャーや絵などの非言語的手段を活用したり、言い換えをしたりして、児童の理解を促す工夫が必要となる。従来は中学校や高等学校で学習していた内容や、EGP では A2 レベル以上に含まれる文法項目は、理解や産出に困難を生じる児童が出てくる可能性がある。どの項目をどこまで身に付けることを目標とするのかを明確にし、発達段階に合わせて適切な指導をすることが必要である。また場合によっては、教材には含まれていても授業内では取り扱わないなどの配慮が求められる。

弟三に、学習指導要領には記載されていないが、EGP で A1 レベルに記載があり、『We Can!』にも含まれている文法項目は、次回の改訂で学習指導要領に記載してもよいかもしれない。上述のとおり、2020 年度からは複数の教材が出版される。そこで、少しでも教育内容の差を小さくするためには、児童が身に付けることができると思われる文法項目は学習指導要領に記載することも一つの方法である。

今後の研究としては、各社が出版する複数の教材を比較分析することが必要である。その際には、特定の文法項目の出現箇所や出現頻度などの情報も有益になるだろう。また、教材に含まれる文法項目を、児童がどの程度身に付けることができたかの調査を行う必要がある。EGP のレベル分けに加えて、実際に日本の小学生が学校教育の中でどのような文法項目を身に付けることができるのかを明らかにすることが、学習指導要領に基づいて行われる小学校から高等学校までの外国語教育を円滑に接続する上で、重要となってくるだろう。

参考文献

Cambridge University Press. (2015). *English Grammar Profile.* Cambridge: Cambridge University Press. http://englishprofile.org/english-grammar-profile/egp-online

中條清美・西垣知佳子．（2010）．「小学校『英語ノート』の語彙分析」．『英語コーパス研究』17，115-126.

藤原康弘・仁科恭徳・松岡結．（2010）．「小学校外国語活動における品詞・文法へのコーパス言語学的アプローチ」．

　　『JACET Chubu Journal』8，15-31.
本多尚子・志村昭暢．（2017）．「小学校外国語活動教材と中学校英語教科書における文法項目の比較」．『北海道教育大学紀要教育科学編』68(1)，119-135.
本田勝久・星加真実・小川一美．（2008）．「小中連携を視野に入れた英語活動カリキュラムの開発に向けて」．『教科教育学論集』7，19-28.
本田勝久・星加真実・田所貴大．（2018）．「小学校英語デジタル新教材 We Can!の語彙分析」．『日本児童英語教育学会研究紀要』37，169-185.
神谷昇・長谷川信子・町田なほみ・長谷部郁子．（2009）．「『英語ノート（試作版）』の語彙の特徴：品詞と意味の観点から」．『Scientific approaches to language』8，119-145.
文部科学省．（2008）．『小学校学習指導要領解説　外国語活動編』
　　http://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2009/06/16/1234931_012.pdf
文部科学省．（2016）．『次期学習指導要領等に向けたこれまでの審議のまとめ（第 2 部）（図画工作、美術、芸術（美術、工芸）、芸術（書道）、家庭、技術・家庭、体育、保健体育、外国語）』
　　https://www.mext.go.jp/content/1377021_1_5.pdf
文部科学省．（2017a）．『小学校学習指導要領解説　外国語活動・外国語編』
　　http://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2019/03/18/1387017_011.pdf
文部科学省．（2017b）．『中学校学習指導要領解説　外国語編』
　　http://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2019/03/18/1387018_010.pdf
文部科学省．（2018a）．『We Can!①指導編』東京：東京書籍.
文部科学省．（2018b）．『We Can!②指導編』東京：東京書籍.
文部科学省．（2018c）．『高等学校学習指導要領解説　外国語編・英語編』
　　http://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2018/07/13/1407073_09.pdf
西垣知佳子・中條清美・西岡菜穂子．（2007）．「小学校英語テキスト出現語彙の意味領域による分析」．『日本児童英語教育学会研究紀要』27，15-25.
仁科恭徳・藤原康弘・松岡結．（2009）．「小学校英語教育のための重要フレーズおよび定型表現の抽出：前置詞・WH 疑問詞表現の場合」．『JACET Kansai Journal』11，14-32.
North, B., Ortega, A., & Sheehan, S. (2015). *Core Inventory for General English* (2nd Ed.). British Council/EAQUALS.
　　https://englishagenda.britishcouncil.org/sites/default/files/attachments/pub-british-council-eaquals-core-inventoryv2.pdf
東京都教育委員会．（2019）．『令和 2～5 年度使用教科書採択地区別の採択結果（公立小学校）』
　　http://www.kyoiku.metro.tokyo.jp/school/textbook/adoption_policy_other/adoption_result/results_2020_public.html
投野由紀夫．（2016）．「学習者コーパスによる英語 CEFR レベル基準特性の特定と活用に関する総合的研究」．『平成 24 年度～平成 27 年度科学研究費補助金（基盤研究（A））研究課題番号 24242017 研究成果報告書』
　　http://cefr-j.org/PDF/TonoKaken2012-2015FinalReport.pdf

付録　学習指導要領、『**We Can!**』、**EGP** の比較結果

文法項目	学習指導要領	We Can!	EGP	例文
＜文＞				
単文	○	○	○	I want a new ball. I'm happy.
重文	×（中）	○	○	*She won the World Cup and she also got a silver medal in the Olympic Games as a member of Nadeshiko Japan.* *I can't jump, but I can swim.*
複文	×（中）	○	×（A2）	*She lost her right leg when she was nineteen.* *What do you always do before you go to bed?* *If you think she can't do it, put a triangle on the picture.*
肯定、否定の平叙文	○	○	○	I play baseball. He is a good soccer player. I don't like soccer very much. She isn't a teacher.
肯定、否定の命令文	○	○	×（A2）	Go straight for three blocks. Please be quiet, David. Don't run here.
疑問文のうち、be 動詞で始まるもの	○	○	○	Are you from Canada?
疑問文のうち、助動詞（can、do など）で始まるもの	○	○	○	Do you like blue? Can you dance well?
疑問文のうち、疑問詞（who、what、when、where、why、how）で始まるもの	○	○	○	When is your birthday? What time do you get up?
疑問文のうち、疑問詞（which、whose）で始まるもの	×（中）	○	×（B1）	*Which picture matches the diary?*
代名詞のうち、I、you、he、she などの基本的なものを含むもの	○	○	○	I want a new ball. This is my hero.
動名詞や過去形のうち、活用頻度の高い基本的なものを含むもの	○	○	○	I am good at swimming. I enjoyed fishing. I saw the blue sea. It was beautiful.
＜文構造＞				
主語＋動詞	○	○	○	I sometimes get up at 6:00. I went to Okinawa.
主語＋動詞＋補語のうち、主語＋be 動詞＋｛名詞・代名詞・形容詞｝	○	○	○	David is a good tennis player. This is me. It was fun.
主語＋動詞＋補語のうち、主語＋be 動詞以外の動詞＋｛名詞・形容詞｝	×（中）	○	×（A2）	*It seems interesting.* *That sounds interesting.*
主語＋動詞＋目的語のうち、主語＋動詞＋｛名詞・代名詞｝	○	○	○	I like baseball. I play it on Saturdays.
主語＋動詞＋目的語のうち、主語＋動詞＋｛動名詞・to 不定詞・how（など）to 不定詞｝	×（中）	○	×（A2）	*I want to know how to use computers, so I have a computer class, too.*

文法事項				例文
主語＋動詞＋目的語のうち、主語＋動詞＋〔that で始まる節・what などで始まる節〕	×（中）	○	×（A2）	*Can you guess what he/she can do?* *I hear it's very exciting.*
主語＋動詞＋間接目的語＋直接目的語のうち、主語＋動詞＋間接目的語＋〔名詞・代名詞〕	×（中）	○	×（無）	*I can teach Maria Japanese.*
There＋be 動詞＋〜	×（中）	○	○	*Are there any differences between club activities at elementary schools and at junior high schools?*

＜文法事項＞

助動詞（can〔能力〕）	○	○	○	She can play baseball well. I can't play the piano. Can you dance well?
助動詞（can〔依頼〕）	×（中）	○	○	*Can you teach me Japanese?*
助動詞（can〔許可〕）	×（中）	○	×（A2）	*You can look at the picture in the textbook.*
助動詞（may〔許可〕）	×（中）	○	×（B1）	*May I ask you some questions?*
現在進行形	×（中）	○	○	*Now she is playing for Seattle Reign FC in America.* *What are the students doing?* *What's missing in your town?*
現在完了形	×（中）	○	×（A2）	*Have you ever had Indian curry?*
助動詞などを用いた未来表現	×（中）	○	○	*You will make a mini poster like this with your partner.* *You will be a junior high school student soon.*
to 不定詞＜名詞としての用法＞	○	○	○	I want to go to Italy.
to 不定詞＜副詞としての用法＞	×（中）	○	×（B1）	*Many people go to Shinto shrines or temples to make a new year wish.*
現在分詞や過去分詞の形容詞としての用法	×（中）	○	×（無）	*In France they have a very exiting bicycle race in July.* *Make a pair with a person sitting in front of you.*
受け身	×（中）	○	×（A2）	*It was built in 2012 and it's 634 meters high.* *It's called "borscht" or red soup.*
分詞構文	×（高）	○	×（無）	*They parade along the main street dancing in colorful costumes.*

凡例1：例文中の正字体は学習指導要領から、斜字体は『We Can!』から抜粋。
凡例2：（中）中学校の学習指導要領に記載あり（高）高等学校の学習指導要領に記載あり（A2）EGP の A2 レベルに記載あり（B1）EGP の B1 レベルに記載あり（無）EGP のどのレベルにも記載なし

英語教育改革に対する中高教員の理解と実践
—クリティカルな視点から—

熊澤　雅子*

*桜美林大学
kumazawa@obirin.ac.jp

要旨
　グローバル化が加速する現在、世界共通のコミュニケーションツールとして英語の重要性が高まっている。英語力が経済的有利への近道だという言説が広がる中で進むトップダウンの教育改革は、政治的・経済的な動機を背景に教育上の議論が十分にされないまま実施されるケースが広く見られ、日本でも、コミュニケーション言語としての英語が強調される近年の学習指導要領改訂では、施策と実践との埋められないギャップが指摘されている。
　本研究では、この施策と現場のギャップの実情と要因への理解を深めることを目的に、日本の中高に勤務する4名の英語教員から収集された質的データを「クリティカル」な視点で分析した。結果、教育改革と教員心理には明らかな距離があり、教員は「他者」から課された教育改革よりも自分に近いステークホルダーのニーズを優先すること、また改革が広がらない要因として学校単位の組織的な取り組みの不在があることなどが明らかになった。

キーワード：英語教育改革、言語政策、教員心理

1.　はじめに

　グローバル化が加速する現在、経済的・文化的交流における世界共通のコミュニケーションツールとしての英語の重要性の高まりは、言語教育のあり方にも大きな変化をもたらしている　（Block & Cameron, 2002）。新自由主義が席巻する中、英語が世界語となった時代に生き残るためには、英語力を身につけることが経済的有利への近道だという言説を背景に、国家による教育政策が策定され、英語コミュニケーション能力育成に向けた動きが加速している。

　日本においても、「英語コミュニケーション力」育成への期待は、経済界からの要請を受けて政府の英語教育政策の目玉となって久しく、近年においては、小学校における英語教育の拡大や大学における英語を媒介とした授業の増加など、様々なレベルでカリキュラム改革が行われている。本研究のテーマである中等英語教育においては、コミュニケーション重視が始まった1989年公布の学習指導要領以降、「英語で積極的にコミュニケーションを図ろうとする態度」（文部科学省, 1989）の育成に始まり、コミュニケーション活動の導入を経て、現在では英語による英語の授業を原則として、「読む・聞く・話す・聞く」の4技能の言語活動を通してコミュニケーションを図る素地を身につけることが目標とされている。

　このように英語コミュニケーション能力の育成に向けて大きく舵をとってきた日本の英語教育改革だが、トップダウンの政策が教育現場に浸透する難しさが指摘され、その成果の評価は思わしくない（e.g., Butler &

Iino, 2004; Humphries & Burns, 2015; Kumazawa, 2013; Nishino, 2009, 2011, 2012; Nishino & Watanabe, 2008; O'Donnell, 2005; Saito, 2017; Sakui, 2004; Taguchi, 2005; Tahira, 2012）。これらの文献では、教育改革を阻んでいる要因として、知識偏重・受け身の日本の学校教育文化や知識偏重の入試制度（Nishino, 2009, 2012; Taguchi, 2005; Saito, 2017; Sakui, 2004）、教員の知識・英語力不足（Suzuki & Rodgers, 2014; Taguchi, 2005）や教育観からくる抵抗（Kumazawa, 2013; Nishino, 2012; Saito, 2017; Sakui, 2004）、生徒のモチベーション・英語力不足（Nishino & Watanabe, 2008; Taguchi, 2005）、また学習指導要領そのものの曖昧な記述（Tahira, 2012）などが指摘されている。

2.　研究の目的

　上述したように1980年代から続く英語コミュニケーション力の育成に重きをおいた英語カリキュラム改革をテーマにした研究はほぼ一貫して政策と実践との乖離を問題点として指摘しており、その要因もある程度解明が進んでいる。本研究は現職の英語教員から収集された質的データを、critical language policy（以下、CLP）（Tollefson, 2006）の視点で分析することにより、国家主導の教育改革に伴う複雑な社会的・文化的プロセスが教育の現場にどのように及んでいるかを再検証することを目指した。

　Tollefson（2006）によると、教育政策研究におけるcritical theory には主に3つの特性がある。第一が政治的に中立的な既存の研究を批判し、教育政策が社会的

強者の利益を守っているとの立場で政策を論じる視点、第二が社会変革を目指す方向性、最後に教育施策が社会的不平等を生み出し維持している仕組みを明らかにするという意識、である。既存の社会構造の影響をこのように「クリティカル」に捉える視点は、近年の英語をめぐる言説でも徐々に存在感を増している（久保田, 2015）。英語が経済的有利をもたらすという新自由主義的言説への疑問が上がる中（Pennycook, 2007）、日本の英語教育政策をクリティカルな視点で検証することは、既存研究に新たな貢献をもたらすことが期待できる。

　このような背景をもとに本研究では、中高英語教員の「声」を聞くことを通して、英語コミュニケーション力育成に向けた改革が政策策定者の意図通りに進まない要因を、質的手法を用いてクリティカルな視点で調査した。設定した研究設問は以下の通りである。
1)　中高英語教員は、コミュニケーション力を重視する英語教育政策を、どのように実践と関連づけているのか。
2)　国の推進するカリキュラム改革が、教員の実践に浸透しない傾向にあるのはなぜか。

3.　研究手法
3.1 研究参加者
　研究参加者は現職の中学あるいは高校の英語教員 4 名である。抽出方法はオープン・サンプリング（Strauss & Corbin, 1998）で、筆者のネットワークから、教員経験、学校のタイプ、性別などをもとに抽出した候補者に協力を依頼し、了承を得た先生方を対象者とした。2 名は学校の管理職に就いているベテラン教員で、あとの 2 名は 20 代の若手教員である。勤務校は、3 名が進学校という位置付けの私立高校、もう 1 人は公立中学である。結果として、了承を得られた先生方の組み合わせは偏りもあるが、ベテラン・若手が 2 名ずつ、また公立と私立、中学と高校、女性と男性を対象者に含めることができた。研究対象者の基本情報は下記の表 1 に示された通りである。また以下にそれぞれの研究参加者について、授業実践の様子や英語教育観、学習指導要領への基本的な意見など、データを解釈する際に必要となる事項を中心に紹介する。

名前	教員経験	性別	役職	勤務校（タイプ）
羽田先生	38 年	男	英語科主任	A 校（私立高校）
須賀先生	29 年	女	教頭	B 校（公立中学）
加藤先生	3 年	男	特になし	C 校（私立高校）
岩井先生	2 年	男	特になし	D 校（私立高校）

表 1. 研究対象者 [1]

　羽田先生は政令指定都市にある私立高校で英語科主任を長く務め、読解や文法、大学受験対策の授業を担当している。基本的に日本語を使う従来型の教授スタイルだが、最近はタブレット機器を使用したインタラクティブなスタイルを試みたり、ネイティブ教員と協同でアクティブラーニングを取り入れたり、ベテランでありながらも新しい授業に挑戦する向上心を持ち続けている。学外にもネットワークを持ち、ここ数年はそのネットワークを利用して様々な研修会に参加して自己研鑽を行ったり、英語教育改革を検討する文部科学省の会議に中高教員の代表として参加するなど、意欲的に英語教育に取り組んでいる。学習指導要領の改訂について意見を尋ねると、現状の英語教育に問題点があることは認め「我々の方も考えなきゃいけないことはある」としつつも、文法の基本を教えることの重要性を「皮膚感覚で」知っており、基礎が固まらないところに 4 技能を積み上げようとしても成果が上がらないのではないか、と感じているとのことであった。

　須賀先生は、大都市近郊にある公立中学で教頭を務めている。現在は職務上、定期的に授業は持っていないが、学内の英語科活動には必要に応じて関わっている他、地域の公立中学英語教員の研究会に属し活動を続けている。また勤務校のある市で地元を題材にした英語教材の開発に関わるなど、英語教育に対して高い関心と情熱を維持している。自分の英語教育の最終目標は、生徒たちが「英語で世界とつながりたい」という思いを抱いて卒業していくことであると語り、その思いは授業でコミュニケーションが成立した時の楽しさを通して育まれる、と信じている。学習指導要領の改訂について意見を求めると、自分は初めから基本的なコミュニケーションを目指して授業をしてきたので「基本ライン」は変わらず、ただその量と質を試行錯誤の中で改善するのみ、との答えであった。高校で教鞭をとる他の三人の研究参加者は口を揃えて大学入試改革の影響を懸念していたのと対照的に、須賀先生はその影響はほとんど感じていないと語り、影響を受けるのは「旧態依然」のやり方をしている高校ではないか、とのことであった。

　加藤先生は、東北地方の中都市にある大学付属の私立高校に勤務して 3 年目の若手教員である。最終的には CLIL を取り入れた授業をすることが理想だと語り、教員初年度から授業では英語を多く使い、4 技能を使ったタスクを取り入れながらアウトプットにつなげる授業を試みている。この点では、近年の学習指導要領の意図するところと教育理念は近いといえよう。ただ現在は CLIL の実現は難しいと感じているとのことで、その理由として、自分の勉強不足以外に、担当する進学コースの生徒のほとんどが付属の大学への推薦合格を目指しているため、その試験に合格することが自分にとってより喫緊の課題になっていることを挙げている。文科省主導の英語教育改革については、アウトプット重視で多少の文法的な間違いは気にせずどんどん使って流暢さを上げる方が大事であるという自分の信念とある程度方向性が同じである、と考えている。

岩井先生はこの研究の参加者中最年少で、教員歴 2 年目、日本中部の地方都市の私立高校で教えている。勤務校は進学する生徒が過半数を占め、岩井先生は成績がトップクラスの国公立大学や難関私立大学を目指す生徒向けの進学コースを担当している。岩井先生は高い英語力を生かして授業でも英語を多用し、テクノロジーを積極的に活用した意欲的な授業を行っている。学校にもその力を認められ、2 年目にしてオープンキャンパスでの模擬授業や出張講義を任されている。学習指導要領の改訂にともなうコミュニケーションや 4 技能の重視については、高校 3 年間での実現は難しいのではないか、と感じており、岩井先生自身は 2 年目に入り進学コースの生徒に対しては運用力より文法の型をたくさん覚えさせることに重点を置いた、いわゆる「トラディショナルな」アプローチの有効性を感じていると述べている。

3.2 データ収集と分析

　本研究は質的研究手法をとっており、データは筆者が全て直接収集した。主なデータソースは研究参加者とのインタビューや、筆者が学校訪問や授業観察を行った際のフィールドノートである。インタビューは、学校訪問時あるいは学校訪問後に一度、1 人 1 時間から 2 時間かけて行った。形式は、半構造化インタビューを基本とし、事前に用意したインタビュー項目をもとに、回答から研究テーマにつながる項目が生じた際は質問を追加してさらに掘り下げて行く形をとった。質問項目は主に、1）研究参加者の教員歴、2）英語教育観、3）勤務校の特徴、4）学校としての英語教育への取り組み・システム、5）英語教育改革（学習指導要領改訂）に伴う個人的な意見と取り組み、などである。インタビュー後、録音した音声データを文字に書き起こした。

　学校訪問は各校一回行い、学校近辺や校内を観察するとともに、教頭職のため通常授業を担当していない須賀先生以外の研究参加者については授業観察も行った。A 校はちょうど訪問日に受験生の保護者向けの学校説明会があったので参加させてもらった。また C 校では英語科主任の先生に面会し学校の沿革や概要、英語科の全体方針などについて聞き取りを行うことができた。上記全てにおいてフィールドノートを取り、手書きのものは訪問後すぐデジタル化した。それ以外にも、英語科のカリキュラム、保護者説明会資料、学校パンフレット、学校の年間スケジュールなど入手可能なものを参考資料として持ち帰った。

　データ分析はクリティカルな視点を保ちつつ、研究設問の答えを研究参加者の視点に照らして探ることを目指した。収集したインタビューデータは、筆者がメモを取りながら複数回聞くとともに、文字起こし後マージンにメモを取りながら熟読し、浮かび上がってきたテーマを筆者の視点で解釈した。その解釈にあたりフィールドノートや持ち帰ったデータを随時参考にし

た。

4. 結果

4.1 カリキュラム改革と実践の関連：心理的距離

　まず研究設問（1）の「中高英語教員は、コミュニケーション力を重視する英語教育政策を、どのように実践と関連づけているのか」について、本研究参加者の声から浮かんだテーマが「学習指導要領と教員との心理的距離」である。以下、引用を交えて詳細を説明する。

　まず最初の引用は、筆者が指導要領改訂に伴う変更についての意見を尋ねた際の、公立中学に勤務する須賀先生とのやりとりである。

須賀：まあどっちでも。どっちでもいいというより、楽しく本当に外国語に関わってくれるぶんにはいいと思うんです。まあ正直どっちでもいいんです。でもやれっていうからやるし。
熊澤：（略）結局降りてきたものをそのままやらなくてはいけないということに対しては、もう仕方ないって割り切られているわけですよね。
須賀：割り切ってるか割り切っていないかは別にして、やらなきゃならないですからね。

ここで須賀先生は、指導要領改訂に伴う変化は、賛成・反対にかかわらず従わなければならないものであるという前提で、「正直どっちでもいい」と述べている。この発言の裏から読み取れるのは、教育改革は「制度」としては受け入れるが、その中身となる自分の信念や実践には大きな変化はない、というニュアンスであろう。実際、須賀先生は別のところで、学習指導要領が変遷を重ねても自分の実践の柱となる部分はなんら変わっていない、とも明言し、むしろ学校で責任のある立場として、「制度」そのものの善し悪しより、現場の教員が「楽しんで外国語の授業をやってくれればいい」が変化に伴う「負担増が一番心配」との懸念を述べてもいる。

　では若手の教員の視点からは、教育改革はどう受け止められているのであろうか。ともに 20 代の加藤先生と岩井先生のインタビューから引用してみよう。両者とも、最近の自分の授業実践に影響を与えている要因を問われての答えである。

加藤：外的な要因っていうので考えるなら、大学入試が一番大きいかな、と思います。やはり、生徒のモチベーションが向かうのは自分自身のことなので、大学入試でどう結果を出すかっていうことが一番大きいかなっていうのは、この 3 年間過ごして来て思ったので。

岩井：まあ 3 年目くらいになってくるとやっぱり意識したのは社会的なところ。結局一番レベルの高い科を持たせてもらっているので、結果出さなきゃいけないな、っていうのは 3 年目、今年度すごく意識していま

すね。

この引用から読みとれるのは、Nishino （2012）が指摘した通り、教員の実践に影響をより大きく持つのは、遥か彼方から降りてくる教育政策ではなく、生徒や学校という目に見えるステークホルダーのニーズである、という点である。前述したように、両者とも、指導要領の変遷とその狙いを理解した上でコミュニケーション能力を高める英語教育を志して教員となり、初年度から様々な試みで実践的な英語力養成を目指してきた、進取の精神に富む先生方であり、過去の文献が指摘したように、教員の力や知識不足（Suzuki & Rodgers, 2014; Taguchi, 2005）、や教育観からくる抵抗 （Kumazawa, 2013; Nishino, 2012; Saito, 2017; Sakui, 2004）によって英語教育改革を実践できないわけではない。実際 2 人とも教師になりたての頃は、タスクを多用しアウトプットに重きをおいた授業を試みたと述べている。ところが、2 年目、3 年目となり周りの状況が一通り見えるようになってくると、加藤先生においては、勤務先の大学付属高校で生徒にとって最大の目標は系列大学への入学試験に合格することであると理解し、授業では学習指導要領に記載された目的を達成することより進学入試に向けた力をつけることを優先し始める。岩井先生も、私学に勤務している以上、保護者や志願者に対してわかりやすい成果を出すことを期待されていると感じ、その期待に英検合格に向けた指導をすることで応え、勤務校の教員としてのミッションを果たそうと努めるようになる。この 2 人の実践の変化から、現場で教壇に立つ教員にとって、指導要領の遂行よりも大きな心理的な場所を占めるのは、自身のおかれた教育環境でのミッションを果たすことである、ということが見てとれるであろう。

　さらに、教育政策が教員の心理に内在化されない理由について一歩踏み込んで言及したのは、進学校で教鞭をとる羽田先生である。羽田先生は、長年受験指導を行い、さらに大学入学や就職後の卒業生の英語との関わりを見てきた経験から、直近の学習指導要領に対して、4 技能の発達は大事だとは思うが「新しいカリキュラムは先走りすぎ」という感想を漏らしている。中等英語教育で注力すべきはしっかりしたベースを作ることであり、その基礎があればこそ生徒は進路に応じて意味のある英語の使用ができるようになるのであって、その基礎固めを軽視して「使う」ことを目指すのは本末転倒ではないか、と指導要領の方向性に疑問を投げかけている。このような考えを持つ羽田先生は、過去に文科省の会議で中高教員の代表の 1 人として発言する機会があった際の経験を振り返り、自分の思うところを話す時間は与えられたが、その発言が会議の意思決定に何の影響も持たなかった、むしろ同席していた企業側の代表者の意見が重視され、「一応聞いた形はとってるけども、決まったらどこかでもう決まって

いるんだな」と感じた、と述懐している。この感想が示す通り、教育策定にあたって長年の教員経験に基づく提言を行う場を与えられた教員ですら、それが教育政策には反映されていないと感じているのであるなら、教育政策が教員の心理に内在化しないのは、むしろ当然のことであろうと思われる。

4.2 改革が浸透しない理由：組織レベルの取組みの不在
　上記に紹介したように、研究設問（1）の、中高英語教員が英語教育政策をどのように実践と関連づけているのかの答えとして、本研究の参加者の 4 名の教員にとって教育政策は、見知らぬ他者が作ったトップダウンの制度上の枠組みであり、制度としては受け入れるが、教員の心理には内在化されない心理的な距離のあるものであるということが見てとれた。そしてまさにこの点が、研究設問（2）の「国の推進するカリキュラム改革が、教員の実践に浸透しない傾向にあるのはなぜか」の答えの一つとなると思われる。つまり、少なくともこの研究に参加した現職の教員たちにとって、学習指導要領は「制度」であり自分たちの実践とは離れたところに存在するものであるため、改革に伴う変化は、自身の理念や使命に「合う」なら教室運営に取り入れるが、「合わない」場合は、自分の経験に基づく信念や、自分により近いステークホルダーのニーズを尊重する。この心理的距離が、教育改革と実践との乖離を生む一つの要因となっていると言えよう。

　さらに、教育改革が進まないもう一つの要因としてデータから抽出されたテーマが「組織レベルでの取り組みの不在」である。この点についても、管理職側のベテラン教員 2 名と、若手教員 2 名の異なる二つの視点から、具体的な引用を交えて詳述する。

　まず管理職である須賀先生と羽田先生であるが、2 人とも声を揃えて学校を挙げて英語教育改革を行う難しさを語っている。須賀先生の公立中学では、須賀先生以外の 3 名の英語教員が各学年を担当している。3 年間を通しての教育目標や英語使用のガイドラインなどがあるか尋ねたところ、「それぞれの先生の意識」と裁量によって決まっているとの答えだった。補足として須賀先生は、「部会もそう開けるものではない」ので「それぞれの先生にお任せ」な部分が大きいと述べている。

　羽田先生も同様に、英語科として一致した取り組みの不在を強調し、まず「頭になる人の力量」不足だと多少自虐的に説明している。その後、英語科に広がらない取り組みの一例として ICT の導入を挙げ、高校生の授業でタブレット端末を全面導入した際、羽田先生自身は早速活用して生徒一人一人にインタラクティブにフィードバックできるなどの利点を感じたが、「使わない先生は使わない」現状であると指摘している。この背景として羽田先生は教員の多忙を挙げ、以下のように述べている。

羽田：やっぱり専任の先生が忙しいんですね。（略）（非

常勤の先生からやる気のある）先生が中心になってやっていただく、っていう形でもいいのかなって。専任の方は負担をなかなかこう、新しいことに発展するまで割けないっていう。

　須賀先生、羽田先生ともに、自身の理想や理念はあっても、多忙な教員に遠慮があり、学校としての細かい達成目標やガイドライン、新しい取り組みを一律に課すことができないでいる実情が浮かび上がる。

　では逆に、若手の先生方からの視点では、どう映っているのであろうか。こちらも両者とも、学校全体での取り組みの不在を訴えている。これについて加藤先生は、勤務校の規模の大きさと保守的な文化、教員の多忙を指摘し、例として、Can-Do を作る話が出た際も、進学コースや国際コースなど多岐にわたるカリキュラム編成のため、全体のコンセンサスが得られず頓挫したことを述懐し、以下のように述べている。

加藤：（Can-Do を作る話になっていたが）結局できていなくて、結局コースとか学年によって大きく差が出てしまうので全体として作るってなるとなかなか進まなかったり。（略）やっぱりこれまでやって来たことが根付いているというか、良くも悪くも保守的なので、そこが難しいなって感じですね。

それに加え、自分の新しい試みを他の先生と共有して学校全体のカリキュラムをよりよくしたいと思っても、英語以外の校務分掌が忙しい中、全体の改革にまで声をあげる余裕がなく、その状態を「良くも悪くも個人戦」と評している。

　一方、若手であっても岩井先生の学校に対する視点はより批判的である。組織レベルでの英語教育改革への取り組みの不在の原因として岩井先生がまず挙げているのは、古くからいる英語教員の意識の低さや英語力不足で、20 人近くいる勤務校の英語教員の中で積極的に授業で英語を使うのは 3 人程度で、それ以外の英語の授業はおそらくオールジャパニーズだが、「結局たぶんできないと思うんで、何も触れられない」状態だと訴えている。この点は、過去の文献（Suzuki & Rodgers, 2014; Taguchi, 2005）で報告されている現実と通じるものがある。また岩井先生は、英語科内での取り組みの以外に、より全体的な学校カリキュラムの問題点に触れ、「コミュニケーション力」の育成は英語科のみの課題ではないと指摘している。

岩井：正直学習指導要領が変わっていく中で、まあ他の教科もそうなのかもしれないですけど、英語だけすごいスポットライトを浴びている感じ。でもどっちかっていうと僕は国語を改革させるべきではないかなっていう風に最近はすごく感じていて。やっぱりその、英語ってコミュニケーションのツールだと思うので、自分の考えが深くならないと英語を使って発信するこ

とができないんじゃないかなっていうふうに。

　岩井先生は、研修で他校を見学した際に、学校全体でアクティブラーニングに取り組んでいる学校の日本語ディベートの実践に感銘を受けたが、そのような取り組みを自校に持ち帰るのは「僕 1 人ではどうにもならない」という無力感を訴えていた。

　上記のように、本研究の参加者の教員の話からは、英語教育改革が教室内の実践に浸透しない要因の一つとして、組織レベルでの共通認識や取り組みの不在があることが示された。本研究の教員たちは、自己の授業実践では賛同できる部分に関しては積極的に取り入れ授業の質を上げようという思いを持って教壇に立っている。しかし、マネジメント側にいる教員は部下への遠慮から、若手側の教員は膠着した学校の文化や年配の教員の意欲不足を感じ、個々の取り組みを学校全体の取り組みへと広げることができずにいる、という実情が浮かび上がった。

5.　考察

　上記の結果をクリティカルな視点で考察すると、日本における英語教育改革の問題点がいくつか見えてくる。以下、マクロとミクロの二つのレベルで問題点を論じる。

　まずマクロレベルの問題として、政策決定のプロセスが挙げられる。「2. 研究の目的」で述べた通り、クリティカルな視点で見た言語政策の制度上の性質として、政策は社会的強者の利益のために作られるという点がある。これを日本の英語教育政策の文脈に置き換えると、政策決定のプロセスのステークホルダーは、政策決定を行う側（＝文部科学省）と、英語が使える人材を必要とする経済界と考えることができる。ところが、本研究のデータで見られた教員の意識上では、政策の 2 大ステークホルダーである政策の作り手とその成果を享受する日本企業は、「他者」である可能性が示唆された。そのように「他者」によって作られた教育改革は、実際にそれを実践する教員にとっては単なる制度であり、必要最低限受け入れねばならないものであるが、政策で決められた目標は教員たちの意識に内在化されることはない。過去の研究でも指摘された通り（Nishino, 2012）、教員は制度という箱の「中身」は自分たちにより近いステークホルダーの利益を重視して決めており、また政治的またイデオロギー的な意図が多くの場合、教育上の目標と一致しない（Yamagami & Tollefson, 2011）という現実が、本研究でも実証された。これをクリティカルな視点で解釈すると、トップダウンの政策決定のあり方そのものが、利益を享受すべき「社会的強者」というステークホルダーの望む結果につながらない要因になっている、と捉えることができる。つまり、単に「制度」を押し付けるだけでは教室の中にまで改革を浸透させることはできず、その改革の成果を期待するのであれば、「政策」とその「実践者」

との距離を縮める必要があると言えよう。さらに言えば、政策の「成果」が誰にとってのものであるべきかという根本的な命題があらゆるステークホルダーの目線から議論される必要がある、ということが示唆として得られるのではないだろうか。

　次にミクロレベルの問題として、学校という組織における明確な"power"の不在を挙げたい。Tollefson（2006）は、CLPにおける"power"を"the ability to control events in order to achieve one's aims"（p. 46）と定義づけ、Pederson（2002）を引用してCLPの枠組みの研究において power は、社会構造と個人の agency（個人が目的を達成しようとする意志）とのダイナミックな関係に焦点をあてて考察されるべきだ、と述べている。本研究においては、前述したように大きな意味での「社会構造」（＝政策）は個人の agency まで power は持たない可能性が示されたが、より小さなレベルでの社会構造と考えらえる「学校」あるいは「英語科」という組織と個人の関係はどうであろうか。本研究では、研究参加者全てが、学校や英語科を挙げての改革に向けた取り組みの難しさを語っており、本人たちには改革に向けた意思があっても、管理職側からのトップダウンも、若手側からのボトムアップも、どちらも全く機能していなかった。この結果から、少なくとも日本における英語教育改革に関しては、ミクロレベルの「組織」において明確な power holder、つまりカリキュラム改革に向けた取り組みを遂行する力を持つリーダー的な存在、がいないという問題点が見えてくる。言語政策研究の分野では、政策を実施する教員の power を重視する研究（Chimbutane, 2011; Cincotta-Segi, 2011）がある一方で、政策の安定した実施にはある程度の組織だった取り組みが必要であると指摘する文献（Johnson, D. C. & Johnson, E., 2015）もある。日本における研究（Humphries & Burns, 2015; Saito, 2017）でも、カリキュラム改革の成否には学校内での共通認識と同僚間のサポートが重要な鍵となることがいわれている。「個人戦では限界がある」という若手の先生の言葉は、英語教育改革の推進には、改革に向けて power holder を持った「組織」の構築が必要であることを表していると言えよう。

　最後に、ミクロレベルの問題としてもう一つ付け加えたいのは、教員の多忙である。近年の社会問題としてメディアでも取り上げられることの多い教員の働き方改革であるが、教員の負担軽減無しに教育改革無し、という点はこの研究でも4名の教員の声からはっきり示されていることを強調しておきたい。

6. まとめに

　最後に、本研究の限界と、今回の研究から示される英語教育改革に向けた今後への示唆を述べる。まず限界であるが、一番大きなものは、本研究のサンプルの問題である。質的研究である本研究は、全国の教師に一般化できる知見を目指したものではなく、少数の教員の声に深く耳を傾けることを目的としてはいるが、オープン・サンプリングで抽出された4名の研究参加者の特性に何らかの偏りがあることは否めない。特に多忙な中高の教員が英語教育に強い関心と使命感なしにはこのような研究への協力を承諾してくれることはないと考えると、本研究の協力者4名がいわゆる「普通の」教員の声を代弁しているかどうかには、疑問を拭い去ることは難しい。この種の研究の遂行上、常につきまとう問題であるが、積極的に声を上げないがおそらく多数派を占めるであろうと思われる「サイレント・マジョリティ」の意識と実践をどのように調査するかは、今後の課題といえる。

　次の限界として、質的研究では避けるのが難しい研究者の主観の問題が挙げられる。本研究のデータ収集と分析は筆者一人で行ったため、半構造化を基本としたインタビューで収集したインタビューデータについては、フォローアップの質問ややり取りで筆者自身の経験からくる影響や主観を取り除くことは不可能であり、またその分析においても同様に主観が入ってくることは避けられない。この点において本研究の客観性や再現性にはある種の問題点があることは否定できない。しかし他の質的研究がそうであるように、本研究が、データの意味を研究者（＝筆者）と研究参加者が共同で構築する作業であると考えれば、本研究の知見は「ある」視点からの現実の提示として、意味を持つと考えることは十分可能である。

　次に、本研究の結果から得られる今後への示唆を述べる。まず問題点として示された学習指導要領と英語教員の心理的距離を埋めるためには、政策決定プロセスの民主化と、改革の実践者である教員、その実践者にとって重要なステークホルダーである生徒が利益授与者となるような政策の策定が必要である点をあげたい。政策が「他者」から課されたままでは、それが教員の実践に浸透することは難しい。公教育全体の利益を平等に網羅する政策を作ることは容易ではないが、少なくともその策定のプロセスにおいて、当事者の利益が当事者の目線で十分議論されるということがまず必要なのではないだろうか。具体的なステップとしては、政策の作り手は、政治的またイデオロギー的な意図と教育上の目標の乖離から目をそらして前者に重きをおくのではなく、その乖離を認識し、双方にとって利益となる政策のあり方を政策の実践者を交えた話し合いの中で追究することを提案したい。

　またもう一つの示唆として、日本の中学・高校という組織における power の構造を、今後の研究によってより深く理解することの必要性が挙げられる。例えば、日本にも数は少ないが英語教育改革の成功例は存在する。このような学校がイノベーションを実現する際、power がどこにあり、それが個人の教員の agency にどのように働きかけていたのかを調査することも一つの可能性であろうし、あるいはより広い視点で日本の学

校という組織の特異性への理解を深め、個人ではなく集団としての「教員認知」を学校内で形成する方法を考えて行くことも、今後の研究領域として重要になってくるのではないと思われる。

注

謝辞

参考文献

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research. Techniques and procedrues for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage.

Block, D., & Cameron, D. (2002). Introduction. In D. Block, & D. Cameron (Eds.), *Globalization and language teaching* (pp. 1-10). New York, NY: Routledge.

Butler, Y. G., & Iino, M. (2005). Current Japanese reforms in English language education: The 2003 "Action Plan". *Language Policy, 4*, 25-45.

Chimbutane, F. (2011). Rethinking bilingual education in postcolonial contexts. Clevedon, UK: Multilingual Matters.

Cincotta-Segi, A. (2011). Taking in, talking around and talking about the L2: Three literacy teaching responses to L2 medium of instruction in the Lao PDR. *Compare: A Journal of Comparative and International Education, 41*(2), 195-209.

Humphries, S., & Burns, A. (2015). "In reality it's almost impossible": CLT-oriented curriculum change. *ELT Journal, 69*(3*)*, 239-248.

Johnson, D. C., & Johnson, E. (2015). Power and agency in language policy appropriation. *Language Policy, 14,* 221-243.

久保田竜子. (2015). 奥田朋世(監訳). グローバル化社会と言語教育：クリティカルな視点から. くろしお出版.

Kumazawa, M. (2013). Gaps too large: Four novice EFL teachers' self-concept and motivation. *Teaching and teacher education, 33.* 45-55.

文部科学省. (1989). 高等学校学習指導要領（平成元年度改訂）. https://www.mext.go.jp/a_menu/shotou/old-cs/1322571.htm

Nishino, T. (2009). *Communicative language teaching in Japanese high schools: Teachers' beliefs and classroom practices* (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.

Nishino, T. (2011). Japanese secondary school teachers' beliefs and practices regarding communicative language teaching. *JALT Journal, 33*, 131-155.

Nishino, T. (2012). Modeling teacher beliefs and practices in context: A multi-method approach. *Modern Language Journal 96*(3), 380-399.

Nishino, T., & Watanabe, M. (2008). Communication-oriented policies versus classroom realities in Japan. *TESOL Quarterly, 42*(1), 133-138.

O'Donnell, K. (2005). Japanese secondary English teachers: Negotiation of educational roles in the face of curriculum reform. *Language, Culture, and Curriculum, 18*(3), 300-315.

Pennycook, A. (2007). The myth of English as an international language. In S. Makoni & A. Pennycook (Eds.), Disinventing and reconstituing languages (pp. 90-115). Clevedon, UK. Multilingual Matters.

Saito, Y. (2017). High school teachers' cognition of the policy of "English classes in English," and their classroom practice. *The Language Teacher, 41*(1), 3-8.

Sakui, K. (2004). Wearing two pairs of shoes: Language teaching in Japan. *ELT Journal, 58*(2), 155-163.

Taguchi, N. (2005). The communicative approach in Japanese secondary schools: Teachers' perceptions and practice. *The Language Teacher, 29*(3), 3-12.

Tahira, M. (2012). Behind MEXT's new Course of Study Guidelines. *The Language Teacher, 36*(3), 3-8.

Tofflefson, J. W. (2013). Critical theory in language policy. In T. Ricento (Ed.), *Language policy: Theory and method* (pp. 42-59). Malden, MA: Blackwell.

Yamagami. M. & Tollefson, J. W. (2011). Elite discourses of globalization in Japan: The role of English. In P. Seargent (Ed.) *English in Japan in the era of globalization* (pp. 15-37). London: Palgrave Macmillan.

全国都道府県における英語教育研究の実態調査
—全国市レベルの取り組み—

藤尾 美佐*・山口 高領**・青田 庄真***・新井 巧磨****・飯田 敦史*****
奥切 恵******・金子 淳*******・鈴木 健太郎********・多田 豪*********
辻 るりこ**********・中竹 真依子***********・濱田 彰************
横川 博一*************・木村 松雄**************

*東洋大学
misa_f@toyo.jp
**秀明大学
takane46@gmail.com
***筑波大学
aota.shoma.fu@u.tsukuba.ac.jp
****早稲田大学
tack@aoni.waseda.jp
*****群馬大学
a.iida@gunma-u.ac.jp
******聖心女子大学
okugiri@u-sacred-heart.ac.jp
*******山形大学
jun_kaneko@e.yamagata-u.ac.jp
********北海道教育大学
suzuki.kentaro@k.hokkyodai.ac.jp
*********東邦大学
iekoshinbeegprx@gmail.com
**********神田外語大学
tsuji-r@kanda.kuis.ac.jp
***********青山学院大学
nakatake@aoyamagakuin.jp
************明海大学
hamada.akira@meikai.ac.jp
*************神戸大学
yokokawa@kobe-u.ac.jp
**************青山学院大学
bluemountain333@rc4.so-net.ne.jp

要旨

産官学の連携は、現在の英語教育に不可欠な課題である。本研究は、各自治体が教育委員会などを通じて取り組んでいる英語教育のテーマを調査、共有し、今後の英語教育に多角的に貢献する目的で発足した。本年度は、昨年度の調査結果を踏まえ、都道府県および市の教育委員会を対象に、さらに大規模なアンケート調査を行った。本報告では、昨年度の調査結果より得られた7つの英語教育のテーマに焦点を当て、自治体の各テーマの重視度を6件法で、具体的な取り組みを自由記述の形で質問した。前者は記述統計及びクラスタ分析を、後者はコーディング及びテキストマイニングを用いて分析した。その結果、7つのテーマに関して、都道府県の回答の方が市レベルよりも高い重視度（平均値）を示す傾向があり、また自由記述では、都道府県が政策の大枠に関する様々な取り組みを示す一方で、市レベルでは、より実践的な共通（類似）の取り組みを示唆する回答が多かった。

キーワード：自治体、アンケート調査、（都道府）県教委、市教委、英語教育研究テーマ

1. はじめに

大学英語教育学会関東支部では「全国都道府県英語教育研究のテーマ」を調査研究する特別委員会を 2018 年度に発足させた。その目的は、各自治体が英語教育についてどのような研究テーマを掲げ、どのような取り組みを行っているのかについての実態調査を行い、全国の自治体と研究成果を広く共有することで、英語教育の発展に多角的に貢献することである。

2018 年度は、47 都道府県と 20 の政令指定都市に属する教育委員会に対し、各自治体が進めている、もしくは重視している英語教育の研究テーマを調査した。質問紙調査の回答データを量的・質的に分析した結果、各自治体が進める英語教育研究のテーマは「1. コミュニケーション能力養成」「2. 四技能統合」「3. 児童・生徒の主体性」「4. 連携・協働」「5. 国際理解」「6. グローバル化」「7. ふるさと」というキーワードに分類でき

ることが明らかになった。各キーワードの具体的な内容を表 1 に示す。詳細は奥切他（2019）、飯田他（2019）、JACET 関東支部特別研究プロジェクト（2019）を参照されたい。

表 1. 2018 年度調査で抽出された英語教育研究テーマと目標

1. コミュニケーション能力養成
　　実際の場面で「使える」英語、特に目的や状況に応じて対応できるコミュニケーション能力の育成が目標

2. 統合
　　4 技能統合型の言語活動を指導できる教員研修を行うことが目標

3. 児童・生徒の主体性
　　一人ひとりの学習者に配慮し、探究型学習など、学習者参加型の授業を開発することが目標

4. 連携・協働
　　小学校、中学校、高等学校の教育・研究の連携や、系統性を持った英語教育プログラムの構築が目標

5. 国際理解
生きる力を養うために国際理解教育を推進し、国際感覚を養っていくことが目標
6. グローバル化
グローバル社会で活躍できる児童・生徒の育成が目標
7. ふるさと
外国語（英語）教育と自国・地域理解を関連づけ、アイデンティティを育むことが目標

　一連の分析結果から都道府県単位での英語教育研究のテーマは把握できたものの、市区町村レベルではどのような研究テーマが進められているのかについては未解明であった。そこで、2019 年度の調査では、全国の都道府県に加えて、各市の教育委員会における英語教育研究の現状と課題を把握することを目的とした[注1]。

2. 研究方法
2.1 調査対象
　47 都道府県教育委員会（以下、県教委または都道府県教委）と 815 の市教育委員会（市教委）を調査対象とした。

2.2 質問紙の作成と調査の方法
　表 1 に挙げた 7 つの英語教育研究テーマについて、(a) それぞれをどの程度重要視しているか、(b) 具体的な取り組みの内容は何かについて回答を得る質問紙を作成した。各テーマの重要度については 6 件法で質問した（1. まったく重視しない、2. 重視しない、3. あまり重視しない、4. 少し重視している、5. 重視している、6. とても重視している）。取り組みの具体的な内容については、自由記述による回答を求めた。項目 6c のみ「はい・いいえ」の 2 択による回答を求めた。表 2 に質問内容と解答方法を示す。

表 2. 質問紙の項目内容と回答方法

回答	項目内容
記述	0a. 英語教育について、参考にしている自治体があれば、自治体名とその理由を回答してください。
記述	0b. 英語教育政策に関して、貴自治体の独自の取り組みはありますか
6 件	1a. 貴教育委員会は、研究テーマ「コミュニケーション能力養成」をどの程度重視していますか。
記述	1b. どのような取り組みをしていますか
6 件	2a. 貴教育委員会は、研究テーマ「4 技能統合」をどの程度重視していますか。
記述	2b. どのような取り組みをしていますか
6 件	3a. 貴教育委員会は、研究テーマ「児童・生徒の主体性」をどの程度重視していますか。
記述	3b. どのような取り組みをしていますか
6 件	4a. 貴教育委員会は、研究テーマ「連携・協働」をどの程度重視していますか。
記述	4b. どのような取り組みをしていますか
6 件	5a. 貴教育委員会は、研究テーマ「国際理解」をどの程度重視していますか。
記述	5b. どのような取り組みをしていますか
記述	5c. 国際理解に関して、具体的にどのような国や地域を念頭に置いていますか。その理由も教えて下さい
6 件	6a. 貴教育委員会は、研究テーマ「グローバル化」をどの程度重視していますか。
記述	6b. どのような取り組みをしていますか
2 件	6c. 国際化とグローバル化を明確に区別していますか。
記述	6d. 上の質問（「国際化とグローバル化を明確に区別していますか。」）の回答が「はい」の場合、どのような区別か教えてください。「いいえ」の場合、どのように考えて「いいえ」と判断したか教えてください（自由記述）。
6 件	7a. 貴教育委員会は、研究テーマ「ふるさと」をどの程度重視していますか。
記述	7b. どのような取り組みをしていますか

　2019 年 7 月に、各自治体に質問紙を郵送した。各教育委員会内で最も英語教育に詳しい人物が回答者となるよう指示し、2019 年 9 月末日までに、Web もしくは郵送で回答を送るよう依頼した。なお、回答した自治体が特定されるような形でデータを公表しないなど、各自治体に対して想定される不利益が起こらないよう配慮した。

2.3 分析方法
　質問紙の回収率は全体で 31%（266 件）、都道府県単位では 40%（19 件）、市単位では 30%（245 件）であった。データ入力者と確認者をそれぞれ分けてデータの集計を行った。

　各自治体の傾向を量的に分析するにあたって、平均値や中央値などの記述統計、および回答の割合を算出した。また、各質問項目に対する回答が似た傾向を示す自治体を層別化するためにクラスタ分析（ユークリッド平方距離による Ward 法）を行った。

　各自治体の自由記述に対する回答は、コーディングおよびテキストマイニング用いて分析した。

　コーディングにあたっては、まず抽出した自由記述回答文に暫定的にコーディングをした後、暫定版のコーディング・システムを用いて再分析し、コーディングに修正を施した。その後、コーディング・システムを確定・構築し、各コードの定義付けを行うことによって英語教育に係る主題を抽出した。

　また、分析に際し、コーディングといった手法のみでは分析者の恣意的・主観的要素が混入し、客観的とは言い難い解釈になってしまうことがある。客観性を担保する方法として、テキストマイニングの手法も用い、特に都道府県教委と市教委の相違について分析することとした。分析には KH Coder Version: 3.Alpha.17g) を用いた。

3. 量的分析の結果と考察

表 3 に、英語教育研究 7 テーマの重要度に関する、全自治体の記述統計を示す。平均値は 4.71 から 5.33 の範囲にあった。中央値はすべての項目において 5.00 となった。これらの結果は、各自治体が 7 つの英語教育研究テーマに対して、どれも「重視している」と回答する傾向にあったことを示唆している。

表 3. 全自治体の記述統計

英語教育研究テーマ	n	M	Mdn	SD	Min	Max
1. コミュニケーション能力養成	186	5.33	5.00	0.70	2	6
2. 四技能統合	181	4.81	5.00	0.95	1	6
3. 児童・生徒の主体性	178	5.09	5.00	0.73	3	6
4. 連携・協働	180	5.02	5.00	0.75	3	6
5. 国際理解	182	4.76	5.00	0.79	2	6
6. グローバル化	174	4.71	5.00	0.85	3	6
7. ふるさと	179	4.76	5.00	1.01	1	6

表 4 に都道府県単位での記述統計を示す。平均値は 5.06 から 5.76 の範囲にあり、中央値は 5.00 から 6.00 の範囲にあった。また最頻値はすべての項目において 6 という結果であった。すなわち、全体の結果と比較して、都道府県単位では 7 テーマの重要度をより高く評価する傾向にあったと言える。

表 4. 都道府県単位の記述統計

英語教育研究テーマ	n	M	Mdn	SD	Min	Max
1. コミュニケーション能力養成	17	5.76	6.00	0.44	5	6
2. 四技能統合	17	5.59	6.00	0.51	5	6
3. 児童・生徒の主体性	16	5.44	5.50	0.63	4	6
4. 連携・協働	17	5.47	6.00	0.62	4	6
5. 国際理解	17	5.18	5.00	0.81	4	6
6. グローバル化	16	5.50	6.00	0.63	4	6
7. ふるさと	16	5.06	5.00	0.85	4	6

表 5 は市単位の記述統計を示している。中央値は 5.00 であったものの、平均値は 4.63 から 5.29 の範囲にあり、都道府県単位と比較して各テーマの評定値が下がる傾向にあった。回答 6（とても重視している）を選択した割合を都道府県と市で z 検定により比較した。その結果「ふるさと」を除く 6 項目で、都道府県の方が「とても重視している」と回答する傾向にあることがわかった。具体的な数値を表 6 に示す。

表 5. 市単位の記述統計

英語教育研究テーマ	n	M	Mdn	SD	Min	Max
1. コミュニケーション能力養成	169	5.29	5.00	0.70	2	6
2. 四技能統合	164	4.73	5.00	0.95	1	6
3. 児童・生徒の主体性	162	5.06	5.00	0.73	3	6
4. 連携・協働	163	4.98	5.00	0.75	3	6
5. 国際理解	165	4.72	5.00	0.78	2	6
6. グローバル化	158	4.63	5.00	0.82	3	6
7. ふるさと	163	4.73	5.00	1.02	1	6

表 6. 回答 6 の割合に対する都道府県と市の比較

英語教育研究テーマ	都道府県	市
1. コミュニケーション能力養成	77%	40%
2. 四技能統合	59%	19%
3. 児童・生徒の主体性	50%	27%
4. 連携・協働	53%	23%
5. 国際理解	41%	14%
6. グローバル化	56%	15%
7. ふるさと	38%	24%

さらに、市単位の回答データに対し、類似の回答パタンを示す市とその特徴を検証するためにクラスタ分析を行った結果、表 7 に示す 3 つのクラスタが抽出された。クラスタ 1 は 7 つのテーマのうち 6 つを「重視する」と回答し、グローバル化のみ「少し重視する」と回答するグループであった。クラスタ 2 は、7 つのテーマすべてに対し「とても重視する」と回答していた。最後にクラスタ 3 は、コミュニケーション能力育成を「重視する」のに対し、その他のテーマは「少し重視する」、また「ふるさと」は「あまり重視しない」と回答していたことが分かった。

表 7. クラスタ分析による市単位の下位分類

回答	クラスタ 1 ($n = 93$、 61%)	クラスタ 2 ($n = 37$、 24%)	クラスタ 3 ($n = 23$、 15%)
とても 重視する		コミュニケーション 四技能統合 主体性 連携・協働 国際理解 グローバル化 ふるさと	
重視する	コミュニケーション 四技能統合 主体性 連携・協働 国際理解 ふるさと		コミュニケーション
少し 重視する	グローバル化		4 技能統合 主体性 連携・協働 国際理解 グローバル化
あまり 重視しない			ふるさと

このように、昨年度の調査より得られた 7 つの英語教育研究テーマについて否定的な回答を示す傾向が見られなかったことから、奥切他 （2019） で抽出した「英語教育 7 テーマ」の重要度は高いということが、本調査からも裏付けられた。しかしながら、市単位では質問項目に対する無回答も多く （31%から 36%）、こうしたテーマで英語教育の研究を行っていない自治体もあることが示唆された（例えば、市単位では「コミュニケーション能力育成」を除く 6 テーマに対し、「と

ても重視している」と回答する割合が都道府県と比較
して低くなっていた）。次節では、都道府県単位と市単
位で英語教育研究のテーマに対する視点が異なる可能
性について、自由記述回答の分析結果を報告する。

4. 質的分析の結果と考察

4.1 英語教育7テーマにおける都道府県・市教委の特徴
　コーディングによって抽出された都道府県・市それ
ぞれの英語教育の取り組みに関する特徴について、表8
にその結果を示す。

表8. 英語教育7テーマ、11項目から成る自治体の回答とそ
れぞれの特徴

	都道府県教委	市教委
1.英語教育の取り組みについて参考にしている自治体名	・福井県 ・東京都 ・秋田県 ・群馬県 ・岩手県 ・静岡県 ・徳島県 ・岐阜県	・さいたま市 ・高崎市 ・相模原市 ・川西町 ・成田市 ・岐阜市 ・岡崎市　など
2.独自の取り組み	・英語教育研究事業（プロジェクト）	・授業外での取り組み ・外部試験 ・ALT配置
3.コミュニケーション能力養成の取り組み	・教員研修 ・授業外での取り組み ・授業実践	・ALT活用 ・授業外での取り組み ・授業実践 ・教員研修実施
4.4技能統合の取り組み	・教員研修 ・教材開発 ・授業実践	・教員研修及び講演会の実施 ・異校種間での授業観察 ・指導力・授業改善の取り組み
5.児童・生徒の主体性の取り組み	・小・中・高における全体的な取り組み ・高校での取り組み	・教員研修 ・ALT活用 ・授業作り及び授業実践
6.連携・協働の取り組み	・小中高連携 ・小中連携 ・中高連携 ・小中高–県教委–市教委連携	・小中連携 ・小中高連携 ・保幼小中連携 ・幼小中連携
7.国際理解の取り組み	・高校生の海外派遣 ・海外留学支援 ・英語力強化事業 ・国際バカロレア導入	・ALT活用 ・中学生の海外派遣 ・オリンピック・パラリンピック関連事業との連携
8.国際理解で意識する国や地域	・英語圏 ・アジア ・ヨーロッパ	・英語圏 ・アジア ・ヨーロッパ ・アフリカ
9.グローバル化の取り組み	・グローバル人材事業 ・留学・語学研修支援	・外部試験補助 ・短期留学・海外派遣事業 ・教員のスキルアップ
10.グローバル化と国際化の違い	・国際化＝国家間 ・グローバル化＝世界規模	・国際化＝国家間 ・グローバル化＝世界規模
11.ふるさとの取り組み	・教材作成 ・プロジェクト ・ボランティアガイド養成	・各教科との連携 ・教材作成 ・ボランティア活動

　まず、「1. 英語教育の取り組みについて参考にしている自治体名」として、都道府県教委では福井県、東京都、秋田県、群馬県、市教委では、さいたま市（埼玉県）、高崎市（群馬県）、相模原市（神奈川県）、川西町（山形県）などが挙げられた。参考にしている理由は様々であったが、主に、子供の学力が数値的に近い、パフォーマンステストの実施と授業改善の取り組みが優れている、CEFR-A2相当以上の英語力を有する学生の割合が多い、教育課程特例校制度による独自の英語教育を展開している、あるいは近隣県・近隣市であるという回答が得られた。

　「2. 英語教育政策に関する自治体独自の取り組み」においては、都道府県では、主に英語教育研究事業・プロジェクトに積極的に取り組んでいるとの回答が得られた。また、市教委においても、授業外での取り組み（例　イングリッシュ・キャンプ、イングリッシュ・デイ、スピーチコンテスト、留学生との交流）、外部検定試験実施及び受験助成（例　英検、GTEC）、ALTを増員し、積極的に登用している自治体が多く見られた。

　「3. コミュニケーション能力養成の取り組み」に関しては、都道府県では、教員研修の実施に加えて、授業外の取り組みとして高校生によるディベート大会、イングリッシュ・キャンプ、高校生の海外語学研修を実施していた。また、授業実践においても、CAN-DOリストを活用した指導と評価の実践、タブレットを活用した「聞く」「話す」能力向上を図る取り組みも見られた。一方で、市教委では、ALTに関する記述が多く見られ、ALTの積極的登用、ALTによる研修会開催、ALTとのティーム・ティーチングを実施している自治体が多く見られた。また、授業外の取り組みとしては、イングリッシュ・キャンプや中学生を対象とした海外相互派遣を実施している自治体も見られた。

　「4. 四技能統合の取り組み」では、都道府県では、教員研修を実施する一方で、教材開発として、英四技能テスト結果を活用した指導・評価モデルを作成したり、四技能評価テスト集を発行したりする取り組みを行っている自治体が見られた。加えて、授業実践においては、指定校での統合型の授業実践及び公開、英語による発信力を育成する指導と評価の実践研究に従事

する自治体も見られた。市教委においても、都道府県同様、教員研修や講演会が頻繁に行われている一方で、中学校での四技能統合型の授業を小学校教諭が参観するといった異校種間での授業観察を導入している自治体もあった。

「5. 児童・生徒の主体性に関する取り組み」では、小・中・高における系統的な学習到達目標を CAN-DO リストの形式で作成する試みや高校におけるアクティブ・ラーニング型授業開発支援プログラムを実施する都道府県があった。一方、市教委では、地元大学との連携による研修会の開催、ALT とのやり取りを多く取り入れた言語活動の促進、あるいは、問題解決型・探究型学習に基づく授業作りや振り返りによる学習者自身の学習経過を可視化する取り組みが見られた。

「6. 連携・協働における取り組み」では、都道府県は、小・中・高の連携にとどまらず、県教委・市教委との連携を強化している自治体も見られた。こうした連携を通して、異校種間の授業参観を実施したり、合同研修会を開催したり、CAN-DO リストの作成に取り組んでいることがわかった。一方、市教委では、小・中の連携のみならず、（保育園・）幼稚園から小学校・中学校をつなぐ連携を試みる自治体も見られた。これは、市内の大多数の子供が公立幼稚園から就学するという実情を考慮し、幼・小・中が連携することでカリキュラムを作成し、ALT がそれぞれの学校で授業を担当する取り組みであった。

「7. 国際理解の取り組み」に関しては、都道府県では、高校生の海外語学研修・海外派遣に積極的に取り組んでいた。また、海外留学支援も行っていた。海外留学以外の取り組みとしては、留学生との触れ合い交流、国際バカロレア導入準備、イングリッシュ・キャンプやスピーチコンテストを実施している自治体も見られた。一方、市教委では、ALT を積極的に活用することに加えて、中学生の海外派遣にも取り組んでいた。また、姉妹都市との交流やオリンピック・パラリンピック関連事業と絡め、ホストタウンとしての国際交流に取り組む自治体もあった。

「8. 国際理解に関して念頭においている国や地域」は、都道府県・市ともに多様であった。都道府県では、英語圏に加えて、教育交流協定・姉妹間交流としてドイツ（ニーダーザクセン州）、台湾（新竹市）、中国、韓国、カンボジアや渡航の利便性（例　季節定期便）を考慮し、香港など、在籍児童・生徒数を増やしている学校との関係で、フィリピンやアラブ中東諸国という回答が得られた。一方、市教委では、ALT の出身国であるアメリカ、カナダ、イギリス、パキスタン、ガーナ、フィリピン、メキシコ、キューバなどの国々が、姉妹都市・友好都市としては、アメリカ（ナコーテス市）、フィリピン（バコール市）、ブルガリア（シリストラ市）、イタリアが挙げられた。これらの国・地域に加えて、オリンピック・パラリンピックのホストタウ

ンとして、ドイツ、ハンガリー、ベナン共和国、ラグビーW 杯開催市として、フィジー、ウルグアイ、カナダ、ナミビアいった多様な国・地域も回答として得られた。

「9. グローバル化の取り組み」に関しては、都道府県においては、グローバル人材事業を実施したり、留学・語学研修支援を行なったりしていた。また、留学生との交流事業にも積極的に取り組む自治体も見られた。一方、市教委では、外部検定試験補助、短期留学・海外派遣事業、自国文化を外国人観光客に紹介する活動を実践している自治体もあった。さらに、ALT を積極的に活用することで、グローバル社会に対応できる人材育成に努めたり、教員のスキルアップ・派遣事業を実施することで、グローバル化に対応したりしているという回答も得られた。

「10. 国際化とグローバル化の区別」に関しては、都道府県・市教委ともに共通の見解が得られた。両者を特に区別することなく英語教育政策に取り組んでいる自治体が見られる一方で、国際化は、「国家間あるいは 2 国間」の関係を指し、グローバル化は、「物事を世界規模あるいは地球規模」で物事を捉えるという認識で区別している教育委員会もあった。

最後に、「11. ふるさとに関する取り組み」では、都道府県では、ふるさとを紹介する教材を作成し、生徒が英語でふるさとを紹介できる力を育成したり、ふるさとの魅力についてプレゼンテーションを実施したりするなどの取り組みが見られた。また、高校においては、外国人にふるさとの良さを紹介する活動を支援することで、ボランティアガイドを養成する自治体もあった。市教委においても、都道府県同様、地域の教材作成や観光地における英語ボランティア活動に取り組む自治体がある一方で、他教科との連携を意識してふるさと事業に取り組む自治体も見られた。具体的には、社会化の授業や総合的な学習の時間で学んだ郷土に関する内容を活用し、英語の時間ではその魅力を伝えるという言語活動を実施するものであった。

4.2 テキストマイニングによる分析

次に、自由記述をテキストマイニングした結果について記述する。ただし、コーディングによる分類は済んでいること、また紙幅の関係から、以下の 2 点に焦点をあてることとする。

1 つ目は、分析に取り上げる項目は、①「2. 英語教育政策に関する自治体独自の取り組み」と、②「8. 国際理解に関して念頭においている国や地域」の 2 項目に絞ることである。理由は、この 2 項目が、もっとも記述量が多かったことに加え、本調査の本質に最も強く関わる内容であると考えられるからである。

2 つ目は、「（都道府）県教委」と「市教委」とで、記述内容に相違があるか、という点に絞って分析を行う点である。なぜなら、今回の調査の目的の一つは、

両者の間の取り組みに相違があるかどうかを明らかにすることが目的の一つだからである。

まず、①「2. 英語教育政策に関する自治体独自の取り組み」であるが、前処理をへて、弾き出された「抽出語の出現頻度数」が表9であり、それを棒グラフで可視化したのが 図1である。

表9. 抽出語の出現頻度数

	抽出語	出現回数		抽出語	出現回数
1	英語	203	21	平成	33
2	小学校	148	22	補助	32
3	外国	113	23	年	30
4	実施	109	24	開催	29
5	授業	100	25	中学生	28
6	教育	89	26	学習	27
7	指導	77	27	研究	27
8	ALT	74	28	英	26
9	研修	72	29	作成	26
10	活動	69	30	支援	26
11	事業	52	31	活用	25
12	行う	46	32	検	25
13	派遣	42	33	検定	25
14	市	41	34	イングリッシュ・キャンプ	23
15	対象	40	35	向上	20
16	教員	39	36	受験	19
17	配置	39	37	交流	18
18	中学校	37	38	中	17
19	市内	34	39	訪問	17
20	推進	33	40	連携	17

	抽出語	品詞	出現回数
1	英語	名詞	203
2	小学校	名詞	148
3	外国	名詞	113
4	実施	サ変名詞	109
5	授業	サ変名詞	100
6	教育	サ変名詞	89
7	指導	サ変名詞	77
8	ALT	未知語	74
9	研修	サ変名詞	72
10	活動	サ変名詞	69
11	事業	名詞	52
12	行う	動詞	46
13	派遣	サ変名詞	42
14	市	名詞C	41
15	対象	名詞	40
16	教員	名詞	39
17	配置	サ変名詞	39
18	中学校	名詞	37
19	市内	名詞	34
20	推進	サ変名詞	33

図1. 抽出語の頻度数

この頻度数は、高い順からそれぞれ「英語」「小学校」「外国」「実施」「授業」となっている。これら抽出語をJaccard係数に基づいて共起ネットワークを描画させる際、特に「（都道府）県教委」と「市教委」（外部変数）に結びつきが強い語を可視化させた（図2）。

回答した「（都道府）県教委」数は19、「市教委」数は245であり、回答数は圧倒的に「市教委」の方が「（都道府）県教委」より多い。しかし、図2では「（都道府）県教委」（右の四角）と結びついている語は22であるが、「市教委」（左の四角）と結びついている語はわずか12である。回答数が少ない「（都道府）県教委」に結びついている語の数が「市教委」の2倍近い数となっていて、逆転現象が生じている。しかも「（都道府）県教委」につながる語の頻度数はすべて少なく多様である一方、「市教委」とつながっている語には、頻度の高い語もある。以上のことから、「（都道府）県教委」と結びついているのは「教材」「海外」「学校」など英語
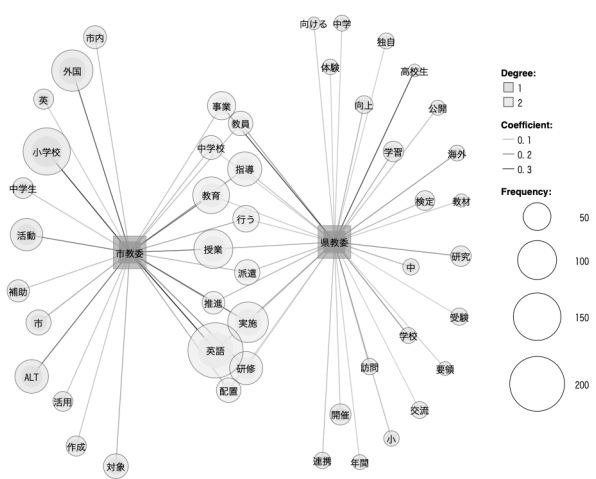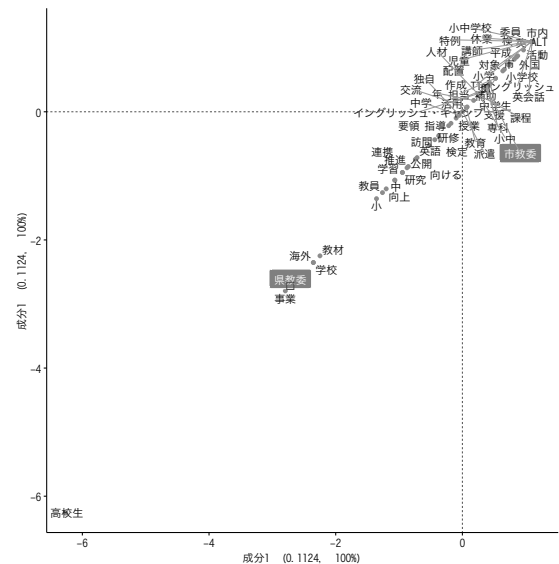

図2. 共起ネットワーク図
「（都道府）県教委」と「市教委」

教育政策的に大枠を示す事柄が、（頻度は低いものの）多いことがわかる。一方、「市教委」と結びついているのは、「市内」「小学校」「外国」「ALT」など、日々の業務として、小学校・中学校を直接、指導・管理することに関連する言葉が多いように思われる。しかもそれらは、多くの市教委において共通しているがゆえ、まとまったものとして扱われ、頻度数が高く認識されたのではないかと考えられる。

同様の傾向は、分析の数理モデルが異なるが、図3の多重対応分析図でも確認できる。

図3. 多重対応分析 「（都道府県）県教委」と「市教委」

「（都道府）県教委」（中央の四角）の周囲には、「教材」「海外」「学校」など英語教育政策的に大枠を示す言葉が集まり、「市教委」（右上の四角）の周りには、「市内」「小学校」「外国」「ALT」「イングリッシュ・キャンプ」など、具体的な取り組み例を示す言葉が集まっている。しかも、極度に密集しており、緊密さゆえ、ひとかたまりとして認識されたものに思われる。いずれにせよ、「（都道府）県教委」の自由記述は、英語教育政策的な大枠の言葉で記述されている傾向があり、「市教委」は具体的な取り組み例を記述している傾向があると理解することができる。

表10. 抽出語の出現頻度数

	抽出語	出現回数		抽出語	出現回数
1	国	29	21	オーストラリア	5
2	アメリカ	19	22	オリンピック	5
3	都市	19	23	タウン	5
4	ALT	16	24	ホスト	5
5	交流	16	25	学校	5
6	地域	16	26	在籍	5
7	姉妹	15	27	世界	5
8	特に	15	28	台湾	5
9	英語	11	29	友好	5
10	出身	10	30	ドイツ	4
11	アジア	8	31	ブラジル	4
12	中国	8	32	外国	4
13	文化	7	33	韓国	4
14	カナダ	6	34	国々	4
15	フィリピン	6	35	国際	4
16	教育	6	36	市	4
17	児童	6	37	多い	4
18	生徒	6	38	置く	4
19	理解	6	39	特定	4
20	イギリス	5	40	念頭	4

	抽出語	品詞	出現回数
1	国	名詞C	29
2	アメリカ	地名	19
3	都市	名詞	19
4	ALT	未知語	16
5	交流	サ変名詞	16
6	地域	名詞	16
7	姉妹	名詞	15
8	特に	副詞	15
9	英語	名詞	11
10	出身	名詞	10
11	アジア	地名	8
12	中国	地名	8
13	文化	名詞	7
14	カナダ	地名	6
15	フィリピン	地名	6
16	教育	サ変名詞	6
17	児童	名詞	6
18	生徒	名詞	6
19	理解	サ変名詞	6
20	イギリス	地名	5

図4. 抽出語の頻度数

次に、②「8. 国際理解に関して念頭においている国や地域」について検討する。抽出語の頻度数を表 10、それを可視化したものを図 4 に示す。

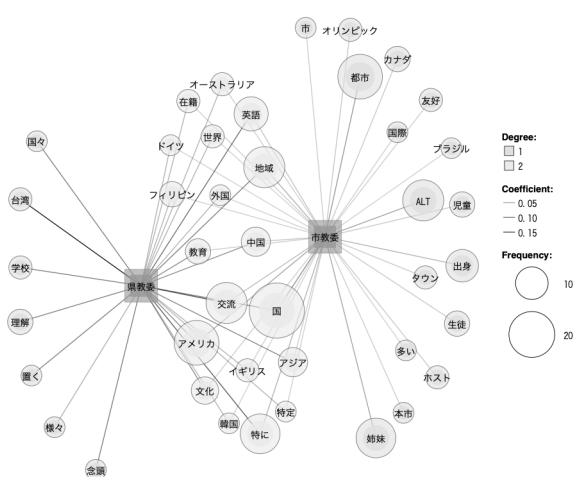ここでは、頻度の高い順に「国」「アメリカ」「都市」「ALT」、「交流」となっている。これを①と同様、描画したものを図 5 に示す。

図 5. 共起ネットワーク図 「（都道府）県教委」と「市教委」

図 5 では、「（都道府）県教委」（左の四角）と主に結びついている語は 7、「市教委」（右の四角）と結びついている語は 16 である。回答した「（都道府）県教委」の数（19）と、「市教委」の数（245）を比較すると 2 倍強となっている。ここでも「（都道府）県教委」と結びついている語は「国々」「理解」「念頭」「様々」など抽象的な傾向が強く、「市教委」と結びついている語は「ALT」「姉妹」「都市」「オリンピック」など、具体的な事柄の傾向が強い。それは、図 6 の多重対応分析図においても、同様の傾向を確認することができる。
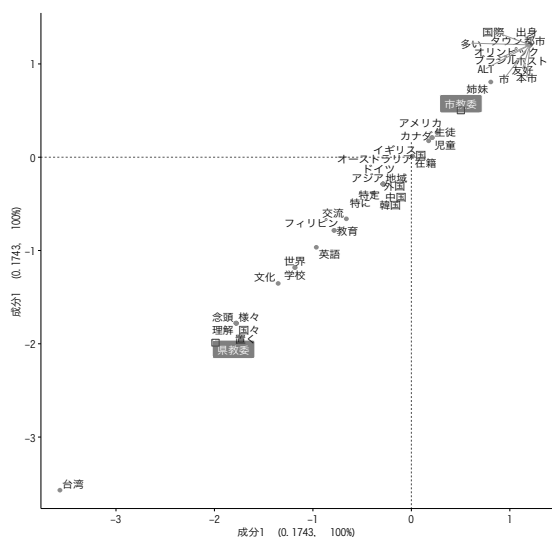

図 6 多重対応分析 「（都道府）県教委」と「市教委」

①と②の分析結果から、「都道府県教委」は英語教育政策的な全体的・抽象的な言葉で多く記述され、「市教委」は個別的・具体的な取り組み例などを記述している傾向があると言える。

4.3 自由記述回答から分かったこと

自由記述分析結果から、都道府県教委では英語教育政策に関して様々な取り組みを行なっている一方で、市教委では、概して共通の取り組みを行なっていることが分かった。また、市教委では、都道府県教委と比べてより実践的・具体的な取り組みに関する回答が多く得られた。英検や GTEC などの外部試験受験助成、小・中・（高）連携による CAN-DO リストやパフォーマンステスト開発、研修会・講演会実施、ALT の積極的雇用・活用、オリンピック・パラリンピックでのホストタウン事業との連携、ラグビーW 杯とのタイアップ、海外姉妹都市との積極的交流など、様々な活動や取り組みが見られた。こうした取り組みから、市教委では与えられた環境を最大限に活かし、英語教育政策に取り組んでいることが明らかになった。

5. 結論

本研究の分析結果より、昨年度、都道府県および政令指定都市への調査結果で得られた 7 つのテーマは、都道府県教委のみならず市教委でも重要視されていることが明らかになった。同時に、市教委では、都道府県に比べ、テーマの重要度についての平均値が低めであったこと、また具体的な取り組みについては、テキストマイニングより、都道府県が「教材」「海外」「学校」など英語教育の政策に関する大枠を表す語が多いのに対し、市教委では「小学校」「ALT」「イングリッシュ・キャンプ」など、より具体的な取り組みを表す共通（類似）した語が多いことが明らかになった。

この政策の大枠と取り組みへの具体性という特徴の違いは、都道府県と市という行政レベルの違いの反映とも考えられる一方、市教委の取り組みに大きな共通点（類似点）があったことは、各自治体の独自の取り組みとともに、今後の興味深い研究課題でもある。今後は、特徴のある自治体への個別インタビューなども含め、さらに自治体の取り組みの詳細に迫る調査を行ない、英語教育に対する考え方や政策について明らかにしていく計画である。

注
1 本研究は JACET と大修館書店との協議の上に立って計画・実施する産学連携事業である。

参考文献
飯田敦史・山口高嶺・奥切恵・青田庄真・新井巧磨・鈴木健太郎・多田豪・辻るりこ・中竹真依子・濱田彰・藤尾美佐・米山明日香・木村松雄.(2019).「教員養成課程コアカリキュラムの実態調査：大学教職担当者の見解から」*JACET KANTO Journal, 6*, 23–41.

JACET 関東支部特別研究プロジェクト. (2019). 「移行期間に各自治体が取り組む英語教育 :『自治体における英語教育研究テーマに関する調査』から」『英語教育 (2019 年 3 月号)』67 巻 (13 号), 34–39.

奥切恵・濱田彰・中竹真依子・辻るりこ・米山明日香・藤尾美佐・木村松雄. (2019). 「JACET 関東支部特別研究プロジェクトＡ : 都道府県・政令市における英語教育研究テーマに関する実態調査」*JAAL in JACET Proceedings*, *1*, 38–45.

樋口耕一.「KH Coder」（http://khcoder.net）

樋口耕一. (2014).『社会調査のための計量テキスト分析－内容分析の継承と発展を目指して－』京都 : ナカニシヤ出版.

英文読解方略に対して間接ストラテジーが与える影響についての質問紙調査
—共分散構造分析を用いて—

森 明智*

*名古屋外国語大学
amori@nufs.ac.jp

要旨

EFL/ESL 環境下の英文読解方略の研究は、読み手と英文という 2 者間の関係によって読解が進むことを前提として分析を行ってきたが、英文読解方略の習得を扱った先行研究は、（1）読解のモデルを示す教員、（2）仲間（教員含む）との英文の疑問点の話し合い、（3）読解の目的設定と自己省察を中心とする反復学習、という、間接ストラテジー（Oxford, 1990）と定義される内容を含めた分析結果を報告してきた。本研究は、これらの間接ストラテジーが英文読解方略に与える影響を分析するため、英文読解方略から成る 40 の下位尺度の質問紙に間接ストラテジーによる 12 の下位尺度を加えて探索的因子分析を行い、因子間の因果モデルを作成し、先行研究では扱われてこなかった共分散構造分析を行った結果、間接ストラテジーは 5 因子中の 2 因子として表れ、英文読解方略の 3 因子に有意な影響を与えていた。最後にこの結果に基づく教育的示唆を示す。

キーワード：英文読解方略、間接ストラテジー、探索的因子分析、共分散構造分析

1. はじめに

本研究は、英語リーディング授業において、学習者が自律した読み手となる指導を念頭に、英文読解において読み手が使用する英文読解方略の習得過程に注目している。特に先行研究の分析結果から、読み手と英文以外の要素として、読解のモデルとしての教員、疑問点を話し合う仲間、読解の目標設定と自己省察を伴う反復学習、などのの間接ストラテジーと定義される要素が英文読解方略の習得の前提とされている点に焦点をあてて、質問紙を用いた分析を行った。英文読解方略と間接ストラテジーの双方を含めた下位尺度による質問紙を作成し、探索的因子分析を行い、因子の尺度値を用いた因果モデルの作成と共分散構造分析を行った結果、間接ストラテジーの下位尺度に十分な負荷があり(0.4以上)、間接ストラテジーの因子が英文読解方略の因子に有意な影響を与えていることが示された。よって、英語リーディング授業において、学習者に英文読解方略を習得させ、読解力を高めたい場合、単に読み手と英文の 2 者だけを念頭に置くのではなく、読解のモデルを示す教員の役割、疑問点をめぐる仲間や教員との話し合い、自分の読解の様子を振り返る自己省察とその反復、という間接ストラテジーの要素を含めることに要因がある可能性が示された。

2. 先行研究とリサーチクエスチョン

外国語あるいは第 2 言語によるリーディングの読解モデルは、Block(1986) の研究に見られるシンクアラウド法に基づきつつ、読み手が意識している内容を直接聞き取る手法から始まり、語彙および文法に依拠するボトムアップによる読解方略と、背景知識に依拠するトップダウンの読解方略の 2 種が明らかにされているが、80 年代より、どちらにも偏重せず、相互に補助し合いながら読解が進む "Interactive Model" が妥当なモデルとして提唱されている(Carell, Devine & Eskey, 1993)。

また、これらの読解時に読み手が意識する英文読解方略は、読み手が英文への理解に効果的に取り組み、意味をとらえるための知的活動、として定義され(Barnett, 1988)、これらをめぐる先行研究の分析結果からは、リーディングにおける理解(comprehension)は、(1) ボトムアップ・トップダウンの双方の読解方略の相互作用(interaction)に基づき、（2）読み手と英文の間の 2 者間で行われ、（3）優れた読み手は多様な英文読解方略を使いこなす、などの点が共通した結論とされている（Grabe, 2012）。

この（3）の観点に基づきつつ、80 年代より EFL/ESL 環境下における英語リーディング力向上を目的として、読み手に対して英文読解方略の習得をめぐる研究が行われている（Carell, Liberto & Pharis, 1989; Hiromori, 2005; Ikeda, 2013; Ikeda & Takeuchi; 2003; Ikeda & Takeuchi, 2006; Janzen, 2002; Kai, 2009; Kern, 1989; Kimura, 1999;

Kimura, et al., 1993; Klingner & Vaughn, 1999; Taki. 2013）。

　Grabe（2012）は、これらの英文読解方略の習得を扱った研究に対して、（i）学習者自身が読解のプロセスを分析しつつ英文読解方略に対する気づきを促す指導、（ii）教員自身が読解のモデルを解説して明示し、直接的に学習者に英文読解方略の使用を促す指導、（iii）学習者間の少人数のグループを作り、読解を進める中での話し合いを通して英文読解方略の使用を促す指導、などの分類を行い、さらに、これらの研究の結果、成果が見られた結果に共通する内容として（iv）長期間にわたる英文読解方略の使用の反復、（v）自分の読解について学習者自身がリーディングの目的を持った上で自己省察を行う、などの点があることを示している。

　具体的には、Carell, Liberto, & Pharis,（1989）、Kimura（1999）らの研究は（i）に含まれ、Kern（1989）、Kimura, et al.（1993）、Hiromori（2005）、Ikeda & Takeuchi（2003）、Taki　（2013）らの研究は、いずれも（ii）に分類される研究であり、Klingner & Vaughn（1999）らの Collaborative Strategic Reading の研究は（iii）に分類され、Ikeda & Takeuchi（2006）の研究は、リーディングの目標設定を行いつつポートフォリオによる自己省察を何度も行うことによる英文読解方略の習得の研究であり、（iv）および（v）に含まれる。また Kai（2009）の分析は、読解後の retelling を通した英文読解方略習得の研究であるため、復習を行う観点から見れば（iv）に含まれる。

　上記の（i）～（v）の5点をまとめるならば、読み手に対して英文読解方略の習得を促す指導という観点に立つ際に、読み手と英文という2者だけでは解釈できない学習の要素が入っていると言える。特に Ikeda（2013）は、少人数を対象とした英語リーディング学習に対する長期間のポートフォリオの分析を行う中で、学習者が目標設定を行った上で、それに対する長期にわたる自己省察とリーディング学習の自己調整を行うことは、英文読解方略とは異なるストラテジーであると明確に定義づけた上で、これらのストラテジーが英文読解方略を習得していく際の要因となる可能性を示している。

　外国語学習における方略研究を行った Oxford（1990）の定義によれば、上記の（i）～（v）における、リーディング学習の目標設定、リーディングのモデルとしての教員あるいは仲間、教員や仲間との疑問点の話し合い、自己省察を伴う長期の反復学習などの学習の要素は、目標言語に直接に関連しないが、言語学習を支え、管理する方略、と定義される"Indirect Strategies"（『間接ストラテジー』（宍戸・伴訳, 1994, pp.13-25））"に含まれる内容であり、また後に Oxford（2017）自身が外国語学習のためのストラテジーを自己調整学習（self regulative learning）の観点を用いて再構築、再定義した点にも見られるように、Zimmerman（1989）らが主張する自己調整学習ストラテジーとして定義づけることが可能な内容と言える。

　逆に、ボトムアップやトップダウンのような英文読解方略は、先述の Oxford（1990）における外国語学習での学習方略の分類では、目標言語に直接関連する方略、として定義される"Direct Staretgies"（『直接ストラテジー』（宍戸・伴訳, 1994, pp.13-25））に含まれる。例えば、典型的なボトムアップに分類される読解方略である、単語の意味を調べる、文法構造に注目する、であれば、Direct Staregy に含まれる Analyzing Expressions の方略であり、あるいは典型的なトップダウンの読解方略として分類される、自分の知識や経験を内容を理解するのに役立てる、であれば、同じく Direct Strategy に含まれる Reasoning Deductively の方略であると言える。

　よって本調査では、先行研究にて扱われてきた（i）～（v）の内容を、英文読解方略とは異なる間接ストラテジーとして定義づけた上で分析を行う。

　そのうえで、学習者に対する英文読解方略の習得を目的とした指導を対象とした研究において、研究の余地があると思われる点について述べる。

　まず、間接ストラテジーが英文読解方略の使用に与える影響の有無について、同一の間接ストラテジーを用いた研究の間においても影響に対する評価は一定ではない。(i) の指導では、Carell, Liberto, & Pharis（1989）らは、影響が見られた事を報告している一方で、Kimura（1999）では影響が見られなかったという報告がなされている。また、(ii) の指導では、Kern（1989）、Hiromori（2005）、Ikeda & Takeuchi（2003）らの研究では、影響があるとする報告が成されている一方で、Kimura, et al.（1993）や、Taki（2013）らの研究では、影響が無いとする報告がなされており、この点について検証の余地があると言える。

　加えて、研究手法においても研究の余地があると言える。学習者に対して、間接ストラテジーを取り入れる事による英文読解方略使用の影響を分析した研究は、学習者を実験群と統制群に分け、さらに具体的な条件を加えた複数の群を設定した上で、指導の効果を質問紙の結果などに基づく量的データに対する分散分析による有意差の有無によって論じる手法を用いている。あるいは同一群に対する事前事後のデータに対する t 検定や分散分析などの統計を用いた比較による分析手法もある。

　これらの手法は有効である一方で、間接ストラテジーが英文読解方略に与える影響の分析という観点に立つならば、間接ストラテジーを原因とし、英文読解方略を結果とする因果関係があるとする仮説をたてることができるため、このような因果モデルを分析できる回帰分析の手法も、この調査を行う際に有効な分析方法である可能性があることは否定できないと思われる。また、既に述べたように、英文読解とは単一の読解方略によってのみ理解が進むことは現実的でなく、複数の読解方略が関連し合いながら理解が進むとされており、また間接ストラテジーにおいても同様に単一のストラ

テジーのみが有効とされる結論には議論の余地があると言える。よって、複数の英文読解方略に対する複数の間接ストラテジーが与える影響を分析することを目的とすれば、共分散構造分析が明らかにできる側面があることは認められる可能性があると考えられる。

なお、本来は（ⅰ）～（ⅴ）の指導方法を実際の学習者に行った結果に対する分析を行い、結果を報告するべきであるが、本調査はその予備的調査として、特定の指導を行う前の段階として、学習者の心理において間接ストラテジーが英文読解方略に与える影響の有無を検証する事を調査の目的とする。

よって、本研究は以下のリサーチクエスチョンを設定する。

RQ: 英文読解方略に対して、読解のモデルとしての教員、仲間（教員含む）との話し合い、リーディングの目的明確化と自己省察を含む反復学習、という間接ストラテジーが与える影響について共分散構造分析を用いた場合、有意差を伴う影響がみられるかどうか。

3. 研究方法
3.1 研究協力者
愛知県内の私立大学の 3 学科 81 名 （1 クラス 20 名前後（TOEIC 平均 485）を対象に、2019 年 4 月 11 日にて、研究への協力の了解の元に調査を実施した。週 1 回の Intensive Reading の授業にて説明文を対象に英語リーディング学習を行う授業であった。

3.2 分析手法
質問紙法を用いるにあたり、英文読解方略用の質問紙（5件法）を選択する際に、本調査が日本人英語学習者を対象としているため、Carell（1989）の質問紙を元にしつつ、同じく日本人英語学習者を対象に改良がくわえられたIsaji（2003）による質問紙（40項目）を選んだ。

英文読解方略の下位尺度の内容は、英文を効果的に読むため（24 項目）、英文を読むとき（5 項目）、知らない単語があるとき（1 項目）、英語を読んで理解することを難しくしている要因（5項目）、英文を上手に読むことができる要因（5 項目）の5 つに分かれ、すべて5 件法の選択肢である（1:全く同意しない～5:強く同意する）。

上記の下位尺度に対して新しく加える下位尺度を考慮する際に、あくまでも先行研究にて扱われた（ⅰ）から（ⅴ）までの内容に合致する間接ストラテジーに限定した内容を含めた12項目を、英文を読めるようになるために工夫していること、という内容で追加した（計52項目）。よって、本来の間接ストラテジーには、様々な方略が含まれているが、（ⅰ）から（ⅴ）までの内容に合致しないと判断されるものは下位尺度には含めなかった。（例として、間接ストラテジーには学習者の感情に関連するAffective strategyも含まれて

いるが、英語リーディング学習における読解方略の習得に関連する先行研究には含まれない内容であるため、本調査の質問紙からは除外した。）

上記の内容をふまえた上で、本調査で新しく追加された下位尺度の設定にはOxford（1990）のStrategy Inventory of Language Learning （SILL）内のIndirect Strategyに含まれる下位尺度を参考としつつ、本調査が英語リーディングを対象とした分析であるため、英語リーディングの内容を示す形として質問の文言を表記した。具体的な下位尺度の内容を以下に示す。なお、各項目の後に、Oxford（1990, pp.151-173）による間接ストラテジーの内容における定義と、それに加えてO＋数字という表記を用いて、SILLからの下位尺度の番号をも合わせて付記した。なお、52番のみはSILLには該当する下位尺度がなかったが、日本での英語リーディング学習を行う際の現状を鑑みて、下位尺度の一つとして加えた。

41 英語リーディングのより良い読み手になるための自分の学習方法を持っている。あるいは探している。（Arranging and planning your learning （Finding out about language learning）: O33）
42 英単語が使われる状況を思い浮かべるなどして、知っている英語の単語や表現を増やそうとしている。（Arranging and planning your learning （Planning for a language task）: O4）
43 英語リーディングで間違いに気づいた場合、復習して次に生かそうとしている（Evaluating your learning & Arranging and planning your learning）: O31）
44 英語リーディングを勉強できる十分な時間を確保できるようにスケジュールを作る。（Arranging and planning your learning （Organizing）: O34）
45 出来るだけ多くの英語を読める機会が増えるように工夫している。（Arranging and planning your learning （Organizing）: O36）
46 英語のリーディングの明確な目標（TOEICやTOEFL）を持っている。（Arranging and planning your learning （Setting goals and objectives）) : O37）
47 自分の英語のリーディングの進歩について普段から考える。（Arranging and planning your learning （Finding about language learning）) : O38）
48 一緒に英語リーディングの学習ができる人を探すようにしている。(Cooperating with others （cooperating with peers）) : O47）
49 英文を読んで分からない時は先生や友人に聞くようにしている。（Cooperating with others （cooperating with proficient users of the new language）) : O48）
50 英文中の知らない単語については日を置いて英文を読み返しつつ覚える。（Arranging and planning your learning & Evaluating your learning: O9）
51 英単語帳を作ったり、英単語帳を使用し、知らない

単語を何度も読んで覚える。（Arranging and planning your learning （organizing））：O6）

52 英語の文法を集中的に学習した事がある。(Arranging and planning your learning （Identifying the purpose of a language task）)

　なお、統計的分析を行うにあたり、探索的因子分析に対しては SPSS ver.25 を使用し、共分散構造分析には R i386 3.5.3 を使用した。

4. 分析と結果

　探索的因子分析と共分散構造分析の結果を示す。質問紙の結果に対する Bartlett の球面性検定の結果は p <.000 であり、52 項目全体の α 係数は 0.87 であった。スクリープロットの結果、7 因子構造として分析を開始した。最尤法を選択し、英語リーディング学習を対象としている質問紙であるため、因子間に相関があることを想定し、プロマックス回転による分析を行い、負荷量 0.4 以下となった下位尺度は除いて分析を行った。

　分析後、23 の下位尺度からなる 5 因子構造となり全体の分散の 56.2%を説明する結果となった。以下、探索的因子分析の結果を表 1 に、因子間相関を表 2 に示す。

表 1. 探索的因子分析の結果

	I	II	III	IV	V
第 1 因子(α = 0.79)：トップダウン					
7	**0.68**	0.01	0.00	-0.18	-0.05
14	**0.63**	0.05	-0.18	0.04	-0.06
18	**0.61**	0.31	-0.15	0.12	0.09
22	**0.60**	-0.03	0.09	-0.05	0.01
8	**0.59**	-0.16	0.05	0.13	0.06
30	**0.50**	-0.01	0.08	0.04	0.23
25	**0.47**	0.02	0.14	0.09	-0.12
第 2 因子(α = 0.72)：英語リーディングの困難さ					
16	0.09	**0.79**	-0.03	0.08	-0.06
31	0.06	**0.76**	-0.05	-0.01	-0.05
34	-0.09	**0.51**	0.32	-0.18	0.02
33	-0.33	**0.45**	-0.01	0.18	0.06
36	0.12	**0.42**	0.12	0.01	-0.05
第 3 因子(α = 0.73)：復習・反復を含む計画的な学習					
4	0.00	-0.18	**0.70**	0.19	-0.18
45	-0.19	0.16	**0.66**	0.14	0.03
42	0.15	0.15	**0.60**	-0.24	0.06
50	-0.02	-0.01	**0.51**	-0.15	0.33
41	0.05	0.06	**0.45**	0.26	-0.07
第 4 因子(α = 0.70)：文法を含むボトムアップ					
12	0.08	-0.04	0.02	**0.68**	0.05
11	-0.02	0.00	0.09	**0.64**	0.17
39	-0.13	0.14	-0.11	**0.64**	0.00
27	0.11	-0.07	0.13	**0.48**	-0.11
第 5 因子(α = 0.74)：他者との協力による目標達成の学習					
46	-0.09	-0.05	-0.06	0.05	**1.03**
49	0.17	-0.06	0.11	0.05	**0.52**

表 2. 因子間相関

因子	2	3	4	5
1	-0.11	0.30	0.38	0.32
2		0.06	0.02	0.12
3			0.31	0.28
4				0.22

　以上の探索的因子分析の結果、第 1、第 2、第 4 因子が質問紙では 40 番以内の英文読解方略に含まれる下位尺度から構成され、第 3、第 5 因子が 41 番以降の間接ストラテジーに含まれる下位尺度から構成されている。

　これら 5 つの因子の尺度値に基づく共分散構造分析の結果を以下に示す。因果モデルの作成にあたり、(1) 本調査にて新しく加えた間接ストラテジーが従来の英文読解方略に影響を与えており、(2) 間接ストラテジーと英文読解方略のそれぞれの因子の間には相関があるという 2 点による仮説に基づいて因果モデルを作成した。

　その結果、間接ストラテジーである第 5 因子が英文読解方略である第 2 因子に与える影響が低く、適合値が許容範囲とならなかったため、第 5 因子が第 2 因子に与える影響を外して再度分析を行ったところ、適合値は、カイ 2 乗値 = 0.12、自由度 = 1、Pr（＞Chisq）= 0.73 となり、帰無仮説は棄却されなかった。また、GFI = 0.99、AGFI = 0.99、Bentler-Bonett NFI ＝ 0.99、RMSEA = 0 90% CI:（NA, 0.2092199）、CFI = 1、SRMR = 0.0099 となり、適合値は許容範囲となった。
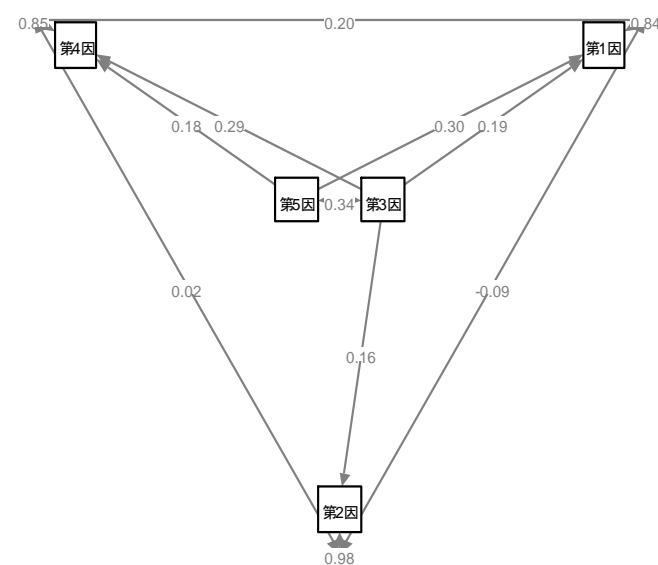
　上記の尺度値に基づく共分散構造分析のパス図を図 1 に示す。

図 1. 共分散構造分析の結果（パス図）

共分散構造分析による因子間の相関について、有意差基準を 5%とした場合の分析結果について述べる。まず、英文読解方略は第 1 因子（トップダウン）、第 2 因子（英語リーディングの困難さ）、第 4 因子（文法を含むボトムアップ）の 3 つの因子として検出されたが、第 1 因子と第 4 因子との間には有意な相関が見られたが（p＝0.038）、第 2 因子と、第 1 因子および第 4 因子の間には有意な相関が見られず（第 1 因子と第 2 因子間の相関: p＝0.372 ／ 第 2 因子と第 4 因子間の相関: p＝0.825）、第 2 因子は他の英文読解方略とは関連が低い独立した因子である可能性が示された。

次に、間接ストラテジーは、第 3 因子と第 5 因子の 2 因子として検出され、有意な相関が見られた（p＝0.003）。

次に間接ストラテジーの因子が英文読解方略に与える影響について、有意差基準を 5%としたうえで述べる。

第 3 因子（復習・反復を含む計画的な学習）が第 4 因子（文法を含むボトムアップ）に与える影響には有意差が見られた（p＝0.009）が、第 1 因子や第 2 因子に与える影響には有意差が見られなかった（第 3 因子が第 1 因子に与える影響: p＝0.087 ／ 第 3 因子が第 2 因子に与える影響: p＝0.15）。

また、第 5 因子（他者との協力による目標達成の学習）が第 1 因子（トップダウン）に与える影響には有意差が見られたが（p＝0.006）、第 5 因子が第 4 因子に与える影響には有意差が見られなかった（p＝0.11）。

4. 考察
本章では、リサーチクエスチョンに対する分析結果に基づく考察と本調査の限界について述べる。

RQ: 英文読解方略に対して、読解のモデルとしての教員、仲間（教員含む）との話し合い、リーディングの目的明確化と自己省察を含む反復学習、という間接ストラテジーが与える影響について共分散構造分析を用いた場合、有意差を伴う影響がみられるかどうか。

探索的因子分析の結果から、新しく加えた間接ストラテジーの質問尺度が十分な因子負荷量をもっていた（負荷量 0.4 以上）。具体的には、第 3 因子（復習・反復を含む計画的な学習: α＝0.73）と第 5 因子（他者との協力による目標達成への学習: α＝0.74）として検出された。よって、英文読解方略の具体的な指導を実践する前の段階であるが、既に学習者の間では、英語リーディング学習に対して、間接ストラテジーと判断される内容に対する意識があると言える結果となった。

この結果をふまえ、尺度値を用いた共分散構造分析の結果から、間接ストラテジーと英文読解方略との間に相関があり、間接ストラテジーが英文読解法方略に影響を与えるという仮説に基づいて立てた因果モデルに対する適合値は許容範囲であり、共分散構造における多重回帰法を用いた分析において、仮説を支持するといえる結果となったといえる。

具体的には、英文読解方略の因子間では、第 1 因子（トップダウン）と第 4 因子（文法を含めたボトムアップ）との間に有意な相関がみられ、両者が協力しあいながらリーディングにおける理解が進むことを示す interactive model を裏付ける結果となった。しかし、第 2 因子（英語リーディングの困難さ）と、第 1 因子（トップダウン）や第 4 因子（文法を含むボトムアップ）との相関が少なく、第 2 因子は独立している因子である傾向が見られる結果となり、英文読解方略は、学習者が感じる英語リーディングの困難さに十分に適用されていない可能性があるといえる結果となった。

間接ストラテジーが英文読解方略に与える影響については、第 3 因子（復習・反復を含む計画的な学習）が、第 4 因子（文法を含むボトムアップ）に有意な影響を与えていた。つまり、英語リーディングにおいて、語彙や文法に基づく読み方を向上させるためには、単に英文を読むだけでなく、未知語や文法に対して時間を置きながらも何度も復習する事が影響を与えると学習者が考えている可能性が示された。

また、第 5 因子（他者との協力による目標達成への学習）が、第 1 因子（トップダウン）に有意な影響を与えていることから、文章全体の内容を背景知識を活かしながら読み解く英文読解方略に対しては、他の仲間たちや教員などと話し合いつつ習得していこうとする意識がある可能性が示された。

本調査の結果からみられる教育的示唆として、以下の 3 点が持つ可能性について述べる。

1 点目として、本調査は、英文読解方略の習得を促す指導を行う前の段階で、共分散構造分析による多重回帰分析に基づく因果モデルの検証の結果、間接ストラテジーが英文読解方略に有意な影響を与えている結果となった。特にボトムアップと分類される英文読解方略の因子（第 4 因子）に対しては長期的な学習計画による復習と反復による学習（第 3 因子）が影響を与えている結果であった。この第 4 因子には、時制に注意する（11 番）、文法や構文の知識を使って、もう一度考えてみる（27 番）、などの下位尺度が含まれており、第 3 因子には、出来るだけ英語を読める機会が増えるように工夫している（45 番）、英単語が使われる状況を思い浮かべるなどして、知っている英語の単語や表現を増やそうとしている（42 番）、英文中の知らない単語について日を置いて英文を読み返して復習しつつ覚える（50 番）、などの下位尺度が含まれている。この結果から、学習者の中では、時制や文法、構文の知識の習得は、同じ内容であっても何度も反復して学習しつつ行う必要があり、さらに語彙の力とも関連があるとされている可能性が示され、実際の教室における指導において一つの示唆を示していると言える。

2 点目として、本調査でトップダウンと分類された英文読解方略は第 1 因子として検出され、これに対しては、他者との協力による目標達成の学習とされた第 5 因

子が有意な影響を与えているという結果となったが、具体的には、第 1 因子には、途中で、そこまでの内容を頭の中で、整理する（7 番）、書かれている場面や状況を思い浮かべようとする（14 番）、文章の中で、すでに出てきた内容を、そのあとに出てきた内容と関連づけることができる（22 番）、自分の知識や経験を、内容を理解するのに役立てる（25 番）などの下位尺度が含まれているが、それにたいして、第 5 因子には、英文を読んで分からない時はネイティブスピーカーや先生や友人に聞くようにしている(49 番)が含まれており、学習者が他者に質問を行う場合、英文内容の文脈や概要を読み取る方略が意識されていると言える。

　3 点目として、上記の内容をふまえつつ、本調査の探索的因子分析の結果、第 2 因子 (英語リーディングの困難さ) として検出された下位尺度に対しては、英文読解方略の他の因子との間に相関が少なく、間接ストラテジーが与える影響も少ない。第 2 因子には、質問紙上で英文を読んで難しくしている要因の項目に含まれている下位尺度として、単語の発音である（31 番）、文を文法的に理解することである（34 番）、英語と日本語の語順の違いである（33 番）、などが含まれているが、いずれも語彙や文法に関連するボトムアップと分類される内容であり、他の読解方略との相関が少なく、間接ストラテジーが与える影響も有意差が見られない。

　これらの EFL/ESL 環境下での英語リーディングにおいて語彙や文法に対して学習者が感じる困難さについては、Clarke（1980）、Cziko（1980）らが指摘する言語閾値仮説（堀場・荒木, 2002）と合致する結果となったと言える。

　EFL/ESL 環境下の英語リーディングは、第 1 言語によるリーディング力の影響を無視して論じることはできず、第 1 言語によるリーディング力がそのまま第 2 言語におけるリーディング力に反映されるとする言語相互依存仮説と、第 2 言語における基礎的な語彙や文法の習得が行われなければ、第 1 言語のリーディング力が第 2 言語にリーディングに反映されないとする言語閾値仮説の 2 種があるが、Grabe（2012）によれば、上記の 2 説は互いに反するものではなく補完し合う関係にあるとされる （Grabe, 2012, pp.129-151）。

　ただ、日本人英語学習者を対象にこの点の検証を行った Yamashita（2002）によれば、L2 リーディングに対して、L2 の語彙や文法の習得が持つ影響力が L1 の読解能力が持つ影響力よりも大きいという報告が行っており、日本人を対象とした外国語によるリーディングの阻害要因を研究した飯島（2004）は、日本人 EFL 学習者の読解阻害要因の多くはボトム・アップ処理能力に根本的な原因があるという主張を示しており、同じく足立・大石（2017）は、（未熟な読み手は）語彙、文法力が低いため、英文を一文ずつ読むことに注意を向けながらも意味を捉えられず、トップダウンストラテジー効率的に使用できない、とする視点を示している

が、本調査の結果もこれらの意見を裏付けるものと言える。

　本来、ストラテジーとは前出の Barnett（1988）の定義に従えば、読み手が英文への効果的に取り組み、意味をとらえるための知的活動であり、困難さに対処することをも目的に含まれているが、本調査では第 2 因子に含まれる英語リーディングの困難さの因子に対して、有意な相関を持つ英文読解方略や有意な影響を持つ間接ストラテジーが見られなかったが、言語閾値仮説に基づいた場合、このような語彙や文法において学習者の感じる困難さを克服することこそがストラテジーの指導において行われる必要があると言える。

　本調査では、これらの語彙や文法の習得に対して、学習者自身が、時間を置きながらも何度も復習する事が影響を与えると既に考えている可能性が示されたが、この第 2 因子の示す困難さの内容に有意な相関を示す因子や有意な影響が与える因子が無いという分析結果から、実践しきれていないという可能性も考えられる。この点は、今後の具体的指導のポイントになると思われる。

　最後に、本調査の限界点と今後の課題について述べる。まず、研究協力者（サンプル数）が少ない点が挙げられる。共分散構造分析には最低 100 名が必要とされているため（竹内・水本, 2014）、本研究の 81 名は十分な数とは言えない。ただ、本調査では、（1）探索的因子分析の結果、新しく加えた間接ストラテジーに属する下位尺度が従来の英文読解方略とは異なる因子として検出されたこと、（2）それらの因子の尺度値を用いた共分散構造分析の結果、許容範囲と言える適合値が得られたこと、また、（3）因子間の因果モデルの分析の結果、間接ストラテジーが（直接ストラテジーと分類される）英文読解方略に有意な影響を与えていること、という 3 点に基づき、英文読解方略に与える間接ストラテジーを分析する際に参考となりうる結果と判断した。

　また、探索的因子分析の結果における第 5 因子が下位尺度が 2 つのみとなっている。新しく加えた重要な下位尺度であるため因子として残したが、本来は 1 因子に 3 つの下位尺度が必要とされているため、この点も本研究の限界点と言える。

　上記の点を含め、サンプル数を増やしたうえでの再調査と、新しく加えた質問尺度に基づいた授業実践の結果に対する分析が挙げられるが、今後の課題としたい。

参考文献
足立望・大石晴美 (2017).「習熟度別リーディングストラテジー指導の効果」『学習開発学研究』(10), 57-63.
Barnett, M, A. (1988). Reading through context: how real and perceived strategy use affects L2comprehension. *The Modern Language Journal. 72(2),* 150-162. doi: 10.2307/328238
Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly, 20(3),* 463-494. doi:

10.2307/3586295

Carell, P., Devine, J. & Eskey, D. (Ed.) (1993). *Interactive Approaches to Second Language Reading.* (5th ed.) New York, USA: CUP.

Carell, P. (1989). Metacognitive awareness and second language reading. *The Modern Language Journal. 73(2),* 121-134. doi: 10.2307/326568

Carell, P., Pharis, B. & Liberto, J. (1989). Metavognitive strategy training for ESL reading. *TESOL Quarterly, 23(4).* 647-678. doi: 10.2307/3587536

Clarke, A, M. (1980). The short circuit hypothesis of ESL reading - or when language competence interferes with reading performance. *The Modern Language Journal, 64(2),* 203-209.

Cziko, A, G. (1980). Language competence and reading strategies: a comparison of first-and second-language oral reading errors. *Language Learning*, *30(1),* 101-116.

Grabe, W. (2012). *Reading in a second language.* (5th ed.). New York, NY: Cambridge University Press.

Hiromori, T. (2005). Metacognitive strategy instruction in L2 reading. *Research bulletin of English teaching. 2,* 1-19.

堀場裕紀江・荒木和美 (2002)「言語習熟度」天満美智子・荒木和美・深谷計子・堀場裕紀江・伊佐原均・小西正恵・宮浦国江・村杉恵子・成田真澄・尾崎恵子・田近裕子(津田塾大学言語文化研究所読解研究グループ)(編)『英文読解のプロセスと指導』(pp.166-184). 東京：大修館書店.

飯島博之 (2004)「L2 読解における阻害要因に関する考察」『茨城工業高等専門学校研究彙法』(39), 23-27.

Ikeda, M. & Takeuchi, O. (2003). Can Strategy Instruction Help EFL Learners to Improve Their Reading Ability? : An Empirical Study Can Strategy Instruction Help EFL Learners to Improve Their Reading Ability?: An Empirical Study. *JACET Bulletin, 37,* 49-60. doi: https://dl.ndl.go.jp/info:ndljp/pid/10501411

Ikeda, M. & Takeuchi, O. (2006). Clarifying the differences in learning EFL reading strategies: An analysis of portfolios. *System. 34(3).* 384-398. doi:10.1016/j.system.2006.04.007

Ikeda, M. (2013). How do learners incorporate the metacognitive strategies taught in the classroom into their strategy repertoires? *Journal of foreign language studies. 8,* 115-131.

Isaji, T. (2003). Reading strategies for Japanese high school students who are not good at English. 『関東甲信越学会研究紀要』17, 1-11. doi:https://doi.org/10.20806/katejo.17.0_1

Janzen, J. (1996). Teaching Strategic Reading. *TESOL Journal, 6(1),* 6-9.

Kai, A. (2009). Achieving global coherence through retelling. *ARELE: Annual Review of English Language Education in Japan. 20,* 41-50. doi:https://doi.org/10.20581/arele.20.0_41

Kern, R. (1989). Second language reading strategy instruction: Its effects on comprehension and word inference ablity. *The Modern Language Journal. 73(2),* 135-149. doi:https://doi.org/10.1111/j.1540-4781.1989.tb02535.x

Kimura, T. (et al.). (1993). Effectiveness of Reading Strategy Training in the Comprehension of Japanese College EFL Learners. *JACET Bulletin, 24,* 101-120. doi: https://dl.ndl.go.jp/info:ndljp/pid/10500507

Kimura, Y. (1999). Metacognitive awareness training and reading comprehension of Japanese EFL learners. *JACET Bulletin, 30,* 45-58. doi:https://dl.ndl.go.jp/info:ndljp/pid/10501293

Klingner, J. & Vaughn, S. (1999). Promoting reading comprehension, content learning, and English acquisition though collaborative strategic reading (CSR). *The Reading Teacher, 52(7),* 738-747.

Klingner, J. & Vaughn, S. (2000). The helping behaviors of fifth graders while using collaborative strategic reading during ESL content classes. *TESOL Quarterly, 34(1),* 69-98. doi:

10.2307/3588097

Oxford, L. (1990). *Language learning strategies:what every teacher should know.* Boston, MA. Heinle & Heinle Publisher. (宍戸通庸・伴紀子(訳)(1994). 言語学習ストラテジー - 外国語教師が知っておかなければならないこと - 東京：凡人社)

Oxford. L. (2017). *Teaching and researching language learning strategies:self-regulation in context.* (2nd ed.)New York, NY. Routledge.

竹内理・水本篤. (編). (2014).『外国語教育研究ハンドブック：研究手法のより良い理解のために』東京：松柏社.

Taki, Y. (2013). Effects of explicit instruction in reading strategies on strategy use and English reading ability of Japanese EFL college students. *ARELE: Annual Review of English Language Education in Japan. 24,* 47-62. doi:https://doi.org/10.20581/arele.24.0_47

Yamashita, J. (2002). Mutual compensation between L1 reading ability and L2 language proficiency in L2 reading comprehension. *Journal of Research in Reading.* (1). 81-95.

Zimmerman, B. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81(3),* 329-339. doi:10.1037/0022-0663.81.3.329

付録. 下位尺度の平均値と標準偏差

	平均値	標準偏差
＊英文を効果的に読むために		
1 主語(〜は)と動詞(〜です, 〜する)に注意する。	4.35	0.80
2 1 文ずつ日本語に訳していく。	3.35	1.12
3 声には出さないが、心の中で、音読する。	4.02	1.02
4 文法に注意する。	3.76	0.91
5 英文の内容を自分の体験と比べながら読む。	2.30	0.99
6 もし辞書があれば、知らない単語の意味を、すべて辞書で調べる。	3.68	1.08
7 途中で、そこまでの内容を頭の中で、整理する。	3.58	1.08
8 文と文の意味上のつながりに注意する。	3.69	0.92
9 1 語ずつ、単語の意味を確かめたあとで、文の意味を考える。	3.12	1.12
10 知らない単語でも、どう発音するのかは、大体わかる。	3.36	0.83
11 時制(現在、過去、未来)に注意する。	3.99	0.80
12 できるだけ、日本語に訳さないで理解しようとする。	2.96	1.10
13 単語の意味を、英語の語順のまま、順番に理解していく。	3.33	1.04
14 書かれている場面や状況を、思い浮かべようとする。	3.73	1.02
15 書かれている内容の、細かい部分に注意する。	2.98	0.89
16 アルファベットを 1 文字ずつ読んだあと、単語として理解する。	1.58	0.85
17 知らない単語が多少あっても、とばして読み進める。	3.76	1.01

18 登場人物ごとに、内容を整理していく。	2.86	0.93
19 まず、始めから終わりまで、ざっと目を通してだいたい何が書いてあるのかをつかむ。	3.35	1.14
20 代名詞(he, she, it, they, this, that など)が何を示すのかに注意する。	3.87	0.82

＊英文を読むとき

21 文章全体の中で、重要な部分と、それらを説明したり補ったりする部分とを区別できる。	3.33	0.78
22 文章の中で、すでに出てきた内容を、そのあとに出てきた内容と関連づけることができる。	3.18	0.76
23 理解できているところ、できていないところを、自分で判断できる。	3.93	0.82
24 わからないところが、多少あっても、英文全体の意味を理解することができる。	3.29	0.90
25 自分の知識や経験を、内容を理解するのに役立てる。	3.14	0.97
26 その文を理解することをあきらめる。	2.54	0.88
27 文法や構文の知識を使って、もう一度考えてみる。	3.67	0.81
28 わからない部分を、日本語に訳してみる。	3.70	0.79
29 そのまま読み続けて、後からもう一度考えてみる。	3.99	0.90

＊知らない単語があるときは

30 前後関係や文脈からその単語の意味を推測しようとする。	4.05	0.85

＊英文を読んで理解することを難しくしているのは何だと思いますか？

31 単語の発音である。	2.04	0.95
32 1語1語の単語の意味である。	3.82	0.98
33 英語と日本語の語順の違いである。	3.13	1.26
34 文を文法的に理解することである。	3.82	0.91
35 文章が全体として、何を言いたいのかを理解することである。	3.77	1.03

＊英文を上手によむことができる人は、なぜ上手によむことができると思いますか？

36 内容の細かい部分に、注意してよむことがよくできるから。	3.51	1.00
37 各文の文型(SV, SVC, SVO, SVOO, SVOC)が何か、よく理解できるから。	3.86	0.88
38 前後の文との、意味上のつながりを理解することがよくできるから。	4.49	0.57
39 日本語に訳さないで、英語のままで理解することがよくできるから。	4.12	1.05
40 文章全体の論理展開に沿って理解することがよくできるから。	4.21	0.82

＊英文を読めるようになるために工夫していることはありますか？

41 英語リーディングのより良い読み手になるための自分の学習方法を持っている、あるいは探している。	3.30	1.05

42 英単語が使われる状況を思い浮かべるなどして、知っている英語の単語や表現を増やそうとしている。	3.63	0.90
43 英語リーディングの学習中に、間違いに気づいた場合、復習して次に生かそうとしている、	3.55	0.94
44 英語リーディングの勉強時間を確保できるようにスケジュールを作る。(英語リーディング授業の予習時間など)	2.80	1.08
45 出来るだけ英語を読める機会が増えるように工夫している。	3.46	0.94
46 英語のリーディングの明確な目標を持っている(TOEIC や TOEFL など)。	3.91	1.00
47 自分の英語のリーディングの進歩について普段から考える。	3.27	1.06
48 一緒に英語リーディングの学習ができる人を探すようにしている。	2.49	1.17
49 英文を読んで分からない時はネイティブスピーカーや先生や友人に聞くようにしている。	3.24	1.17
50 英文中の知らない単語について日を置いて英文を読み返して復習しつつ覚える。	3.02	1.14
51 英単語帳を作ったり、英単語帳を使用し、知らない単語を何度も読んで覚える。	3.50	1.07
52 英語の文法を集中的に学習した事がある。	3.68	1.05

航空機操縦士訓練生は英語で何ができることが望まれるのか
―日英 **PFB** 比較からみる必要な英語力―

西川　惠*・縄田　義直**

*東海大学
megumi-nishikawa@tsc.u-tokai.ac.jp
**航空大学校
nawata@kouku-dai.ac.jp.

要旨
国際民間航空機関によると、世界的な航空需要の増大に伴い 2030 年には現在の 2 倍以上の航空機操縦士（以下、操縦士）が必要とされている。この不足に応えるため、日本においても操縦士を養成する大学が増加し、訓練地も日本以外であるケースが増えている。訓練課程においては Post Flight Briefing（以下、PFB）という操縦訓練終了後に教官との言葉による振り返りを通し次回の操縦に臨む、という重要なポイントとなる場面がある。しかし、非英語母語話者である訓練生が PFB 時に英語で何ができる必要があるのか、という先行研究例はほとんど存在しない。そこで、本研究では日本語による PFB と英語による PFB を比較し、Bloom （Anderson et al., 2001）の学習モデルを基に英語で何ができることが望まれるのかを検証した上で、日本語母語話者である訓練生が PFB 時に必要な英語力とは何かを調査した。

キーワード：English for Specific Purposes, English for Flight Training, Pilot Trainees, Aviation English

1. 背景

　国際民間航空機関によると、世界的航空需要の増大に応えるには 2030 年までに 520 万人の操縦士が必要であるという。特に、成長著しいアジア・太平洋地域では年間 9000 人が必要とされ、日本においても年間 400 名以上の新規操縦士が必要となり、現在の供給可能数をはるかに超える事態となっている(国土交通省, 2014)。この操縦士不足という事態に取り組むべく、各国政府はそれぞれ対応を急いでおり、日本では航空大学校[注 1]が 2018 年より定員を 1.5 倍の 108 名に拡大し、また、2006 年に操縦士養成課程を設立した東海大学を皮切りに、操縦士養成を行う私立大学も増えてきており、その定員も増加傾向にある（表 1 参照）。

表 1．操縦士養成課程のある私立大学

	設立	定員	訓練所（海外）	訓練所（日本）
東海大学	2006	40 50	University of North Dakota（以下、UND）	湘南キャンパス
法政大学	2008	30 30	なし	本田エアポート・大分空港
桜美林大学	2008	30 40	CAE Phoenix Aviation Academy	多摩アカデミーヒルズ
崇城大学	2008	20 20	なし	熊本空港
千葉科学大学	2010	若干名 若干名	Hillsboro Aero Academy	神戸空港
第一工業大学	2014	10 30	Pan Am Career Flight Academy	佐賀空港
工学院大学	2019	若干名 若干名	Sky Creation	なし

（定員の欄は、上段が設立当初の定員、下段が 2020 年度予定の定員）

訓練機関が増加するのに伴い、訓練所も必ずしも日本国内とは限らず、教官が非日本語母語話者、媒介言語も英語であるケースが増えている[注 2]。一例として東海大学と航空大学校の比較を表 2 にまとめた。

表 2．東海大学と航空大学校の制度や使用言語の比較

	東海大学	航空大学校
飛行訓練地	アメリカ	日本（宮崎・帯広・仙台）
教官	主にアメリカ人	日本人
教官のソース	UND の課程を修了した学生	運航審査官・航空従事者試験官・航空会社・

		防衛省・航空大学校等
媒介言語	英語	日本語
英語習熟度	TOEFL®(iBT) 69 点以上（& GPA 2.5 以上）	TOEIC® 750（平均）
飛行訓練時間	FAA: 229 時間 JCAB: 40 時間	202 時間
取得資格	JCAB：事業用操縦士（多発）／ 計器飛行証明 FAA：自家用操縦士／ 事業用操縦士／ 計器飛行証明	事業用操縦士(陸上単発・陸上多発) 計器飛行証明

表 2 の教官のソースであるが、アメリカにおいて航空会社への就職を目指す者は、事業用操縦士・計器飛行証明を取得後、教官業務等で総飛行時間を重ねながら、中小、大手航空会社へとステップアップするのが一般的である。UND においても、卒業生の多くが（フライトインストラクターの資格を取得の後）教官として働いている。なお、日本においては、東海大学や航空大学校等で事業用操縦士・計器飛行証明を取得後、航空会社に就職して、そこで副操縦士としての訓練を行うのが一般的である。

2. 訓練生および訓練教官のニーズ
2.1 訓練生のニーズ
　日本の私立大学養成課程に所属しアメリカの訓練所にて飛行訓練中の日本人学生（57 人）へ行った英語ニーズ調査によると、英語 4 スキルのうちリスニング力が最も必要であると答えた学生は 71.9%、スピーキング力が 56.1%、これに対してリーディング力は 15.7%、ライティング力は 1.7%であった（Nishikawa & Nawata, 2019）。リスニングおよびスピーキング力がどのような場面で必要と感じたかという問いについては、「教官とのやり取り（レッスンの予約、飛行中の指示の聞き取りや応答、PFB 時間中のやり取り）」という答えが目立った。講義を理解する、マニュアルを読む、航空無線でのやり取り等はある程度要領を得れば慣れるという。

2.2 教官のニーズ
　アメリカの訓練所にて監督業務にあたる日本人飛行教官へ行ったインタビューによると、学生が身につけてほしい英語力として PFB 時にイメージングをしながらやり取りができる英語力が挙げられた（西川, 2019）。母語、外国語を問わず PFB 中に教官からの指導内容を理解し次回のレッスン時までに修正できるかが訓練の質を左右する、という。特に訓練の初期段階では基礎的技術を習得することになるので、進捗は言語力によるところが大きく、さらに、この段階での技術の質はその後の訓練の進捗に影響を及ぼす。日本人訓練生の場合、

PFB 終了後に日本語で質問に来ることが多いのだが、直接飛行に立ち会った教官に PFB 時に確認しないと答えられない内容のものがほとんどで、時間が経ってしまうとイメージングの共有も困難になり次回への対策が取れなくなってしまうという。
　このように、飛行訓練においてはイメージングが重要となる。特に、訓練の初期段階での有視界飛行方式（計器ではなく物標等を目視し位置を確認しながら飛行する方式）では、いつ、何が見えたのか、そのタイミングで、高度やスピード等はどの程度なのか、そこでどのような操縦動作を行うのか、を描写し、そして、実際に訓練中に起こったことや飛行中に教官から言われたことを PFB 時に思い出し、次回の訓練における操縦動作をイメージングすることがカギとなる。

3. 研究
3.1 研究課題
　それでは、英語で PFB を行う場合、やり取りはどのような内容なのであろうか。本研究では以下を研究課題とした。
　研究課題 1：日本人教官と日本人訓練生が日本語で行う PFB でのやり取りにはどのような特徴があるのか
　研究課題 2：米人教官と日本人訓練生が英語で行う PFB でのやり取りにはどのような特徴があるのか
　この課題の分析を通して、学生および教官のニーズに応え、訓練の質を高めるための方策の一助としたい。

3.2 方法
　日本人教官と日本人訓練生による PFB および米人教官と日本人訓練生による PFB をそれぞれ録画、文字起こし後、比較して分析した。

3.3 材料
　使用した材料の詳細は、表 3 にまとめた。教官と訓練生の組み合わせは表内にてアルファベットで表記した。

表 3. 英語 PFB および日本語 PFB の録画資料

	本数	教官	学生	訓練生の飛行時間
英語 PFB	5	A	a	約 20〜25 時間（単独飛行前）
		B	b	
		B	b & e	
		C	c	
		D	d	
日本語 PFB	4	A	a Group	約 20 時間（単独飛行前）
		B	b Group	
		D	d Group	
		E	e Group	

　撮影はソニー製のハンディカメラを使用した。英語

については 2018 年 6 月に UND の訓練所にて、日本語については 2019 年 5 月に航空大学校帯広分校の訓練所にて撮影した。英語による PFB 時間はそれぞれ約 10～20 分で、教官 1 名に訓練生 1 名、日本語は約 30 分で、教官 1 名に訓練生 2~3 名であった。

　訓練の初期段階を比較するため、飛行時間が約 20 時間の訓練生（教官の同乗無しで行う単独飛行前の訓練生）のやり取りを文字起こしし比較した。文字起こしは、英語については第一著者が行った後、訓練課程を修了し就職活動を終えた学生を雇用して完成させた。日本語については第一著者が行った後、第二著者が完成させた。録画内容と文字起こし原稿は、日本人飛行訓練教官 2 名により確認を行った。本稿では、その中から日本語と英語の PFB それぞれ 1 つを取り上げる。今回対象とした英語の PFB での訓練生の英語力は TOEFL®(iBT) 70 で、UND の留学要件は満たしているレベルであった。

4. 結果
4.1 日本語 PFB のスタイル
　日本語では訓練生が飛行中に起こったことを描写することから始め、教官がそれに応え、指導するというスタイルのやり取りであった。

<div align="right"><Japanese01></div>

---

1 教官：じゃあ今日のやろうか　じゃあ最初はスズキ（仮名）から

2 訓練生 1：まず　level off が　pitch が先行せずに level off した際に airspeed がちょうどよかったのに落としてしまうことがあって　ここは upwind に関しては風が考えられていなかったので upwind を伸ばそうという考えもなく crosswind の WCA も甘くて　経路が内側に流されてしまってしまいました　その指摘を受けて風を考慮して　その指摘後の周回ではちょっと意識して upwind を伸ばしたり WCA を取るようにして WCA を考慮して　WCA を取ることで abeam の見方が　見え方が変わっていることに気づくのが遅くてターンのタイミングが遅くなってしまって

3 教官：overshoot してたよね　何回も

4 訓練生 1：はい　WCA を取ることだけじゃなくて　気づくのが遅かったので　そういったことは次回にいかしていきます

5 教官：うん

これは PFB 開始時のやり取りで、教官が開始合図（発話 1）、訓練生は振り返り描写を行う（発話 2）。教官は肯定（発話 3）、それを受けて訓練生は次回への課題を確認（発話 4）、教官がその内容を肯定（発話 5）、というやり取りである。この後も基本的にはこの型でやり

取りは続いていく。

<div align="right"><Japanese02></div>

---

6 訓練生 1：あとはえっと　今日特に意識してたのは base turn で 91 knots キープで降下するということで　3 周目まではかなりうまくいっていて　flap handle スムーズで　大分 runway の見え方もいい感じにおろせていたので　そのイメージ通り　イメージフライトその通りには最初の方はなくて　その後の flare も　えっと final turn からも特に大きな修正は

7 教官：そうね 1~2 回目はよかったね　その辺はね で　<u>flare のところは</u>

8 訓練生 1：flare は最初が機軸がずれていて　2 回目が左に流され 3 回目はうまくいった感じだったんですけども　その bank が入ってしまってエリアの際に center line もちょっと甘いのか　ちょっとずれてもその細かい修正をするのに苦しみました

訓練生の描写は続き（発話 6）、教官は肯定するがここで問題点について焦点をあてる（発話 7 下線部）。訓練生は教官からの指摘について何が問題であったか描写しながら分析、評価をしている（発話 8）。

<div align="right"><Japanese03></div>

---

9 教官：そこだよね　サトウ（仮名）も同じようなとこがあったんだけど　彼は bank とか言い出したら　積極的に舵をきるような　スズキも<u>ずれた時に自分でこう直せればいいんでしょ　それがあって風で傾いた時に pitch up down up down した時に同時にずれていくんだよね　せっかく path 直してんのに直ったなと思ったらセンターがずれてしまったりとかさ　それが short final か接地前に起きてしまうと　もうつけられないから かつ power 絞って flare っていうのもあるわけだから　centerline で path で機軸をこれが flare のところ少なくともthresh または flare するまでに止まんなきゃいけないの 止まった状態に　止まった状態でつけられるだから power 絞って返しの姿勢をつくって待ってぽーんてつくとつくまでに一所懸命あわすんじゃなくてパワー絞るぐらい　centerline キープしてんだけど　最後 flare をしながら返しながら　bank とか軸線ももし狂ったら直したから安定して狂ったら直して狂ったら直してついていくっていう　ギリギリで接地ギリギリであえばいいっていうそういう考えじゃないんだけど　時間がないんだよ　狂った時にタイムリーにスムーズに直してあげないと</u>それ今日 2 人とも感じた　接地間際でね

訓練生の自己分析（発話 8）を受けて教官は冒頭で「そこだよね」と焦点を絞り、その後操縦動作の修正法を比較してイメージングさせながら詳細に説明する（発話 9 波線部）。

<Japanese04>

> 10 教官：スズキはなんだ後半が
> 11 訓練生 1：後半が崩れて short cut to downwind があって　まず最初　その base 崩れて　base まで後手 base まで後手に回っていたっていうのがあって
> 12 教官：shortcut っていうのは慣れた経路じゃないよね　normal crosswind downwind あって　自分が知ってる景色だと
> 13 訓練生 1：えっと落ち着いて　はい
> 14 教官：落ち着いてるよね　で　違った景色が見えた時に焦る　あれどこだどこだって探した　あれどこだどこだって　同じことが起きる西側でよその空港でも同じことが起きるじゃんと　ということは西側も一緒じゃん
> 15 訓練生 1：その西側でというよりかは今までの runway の見方を
> 16 教官：そうそう

話題は訓練の後半部分に移り、教官はそこで見られた問題点に焦点を絞る（発話 10）。訓練生は描写によって振り返り（発話 11）、教官は発話 12、13 で訓練生からの答えを引き出しながら課題を具体的に説明していく。訓練生はやり取りの構築から解決策にたどり着き（発話 15）、教官は肯定する（発話 16）。

<Japanese05>

> 16（続き）教官：だから基本に戻るというか runway を見て飛んでいるんだよ　幅も turn して翼端よく見て判断すればいいんだよ　例えばね course in が中央よりそれまでは　風がなければ中央回っていくとかさ　どこでも普遍で使えるやり方ってのがある　それ狂ったって base で修正かけられるじゃん　今日修正してたじゃん　downwind の幅がおかしいけど　狂ったことはしょうがないんだからそういう気持ちになるより焦んないで　焦ってその後狂っていっちゃったんだね最後まで自分で修正しきれなかった　最後は持ち直したの
> 17 訓練生 1：その final turn が　base で修正しきったんですけども　その後しきってから final turn が遅れてしきったところで final turn 始まったんでその overshoot をまた修正しなくてはいけなくてそこで final stabilized せずにおろした

教官は続けて描写を交えながら修正法を説明する（発話 16 続き波線部）。説明の最後で、再び描写することで失敗の原因と起こってしまった過程を思い出させるようにし、それを受けて訓練生は操縦動作を再度詳細に思い出して描写する（発話 17）。

4.2　日本語 PFB の言語使用の特徴

　操縦動作の詳細な説明は、特に訓練の初期段階では主な学習内容となるので大変重要であり、イメージングが大きなカギを握る。タイミング、強さ、スピード、角度等、感覚的な表現が必要になるが、この説明の際、日本語ではオノマトペ等の使用が特徴的である（発話 18、21、23 点線部）。

<Japanese06>

> 18 訓練生 1：はい　その rudder　flare ギリギリの rudder 操作がぎゅっぎゅっじゃなくてじわーっとした rudder control ぎゅっぎゅっなんか踏み込んで踏み込んでみたいんじゃなくてじわーっとした rudder control
> 19 教官：take-off roll の話
> 20 訓練生 1：あと center をあわせに行く時の rudder
> 21 教官：私がじわーっと control しろといったのは例えば急激に踏んでた rudder control だから rudder pedal の踏み方は丁寧に気を付けてやってねって上空ではどういう風に言った
> 22 訓練生 1：その言ってないんですけど　意識としては　踏み込んで踏み込んで離して踏み込んでみたいな機軸をあわせる control だったので
> 23 教官：私が今日スズキに言ったのは全く踏んでない状況が上昇左旋回右左踏んじゃったから ball がぐっと右にいった　あれは全然踏み方が間違ってるよ
> 24 訓練生 1：はい

rudder とは方向舵のことで、これを足で踏むことで垂直尾翼後縁にある翼面をコントロールし機体の左右方向を操舵する。この踏み方についての説明の際に「ぎゅっ」「じわー」「ぐっ」等のオノマトペが見られた。オノマトペの使用は PFB 全体でも「ぽーん」「かくかく」「すーっ」等多くの使用が見られた。また、今回の教官だけでなく、他の教官でも使用が多く観察された。

4.3　英語 PFB のスタイル

　英語でも日本語と同様に飛行について教官が尋ねることから始まる。

1 Instructor: So how did you think of the flight?
2 Trainee: Good maybe
3 Instructor: Very good started off basically getting out of the GFK basically you the radio calls stuffs very happy the only small things I recommend once you reached 3500, turn on course then after that you conclude checklist starting leaning mixture all the after you had already in a Crockston area because if you just keep going on course you just waste time and your money just going north or south I guess all right?

訓練生の発話は見られたが非常に短い。教官は良かった点を挙げて褒め（発話 3 太波線部）、すぐに問題を挙げる（発話 3 下線部）。日本語では訓練生に描写させるのに対し、英語では教官が行うスタイルである。ここでの説明内容は操縦動作そのものというよりはチェックリストを確認するタイミングについてであった。説明後、理解を確認する（発話 3 二重線部）。

　英語では訓練生ではなく教官が描写を行い、描写の中で必ず良い点を褒め、併せて課題を明確に指摘、修正法を簡潔に説明する。理解の確認は要所で行う。基本的にはこの型で進んでいく。

3 (続き) Instructor : And the we are going to Crockstone you forgot to　　do Before Landing Checklist. Make sure … The first landing, flare was okay but you weren't on the centerline to left on the centerline, so we tried to do that and the second one, good on flare again not on the center line. Therefore on third one we tried to do don't touch the ground and you said that how you can use the rudder to maintain the directional control of the aircraft, right? So if you use too much rudder sweep to the right basically small correction so you can maintain the center line. And fourth, you were low and slow you go around but you could have done is the bring back uhh bring up to power so you maintain the same level and just come back to the glide rudder.

教官は描写を続けながら問題点を指摘する（発話 3 続き下線部）。この回は着陸を何度か練習したので、1 回ごとにどうであったかを教官が描写し、良かった点（発話 3 続き太二重線部）と併せてできなかったこと（発話 3 下線部）を説明、修正法を説明（発話 3 続き波線部）する。

4 Trainee: Can I after 200ft AGL?
5 Instructor: If it's quick one because what I saw was like 59(kts), so bump up the power really quick. I would say okay with that one. And fifth one, it was good approach but with my assistance and the 6th one, great approach

and also great round out with great flare　　　That was I really want in a stage. You probably got　　had　　so what changed? Before that you were not really flaring, so wa what
6 Trainee: From first to fifth, I forgot to see the threshold to…　I changed the
7 Instructor: So, vision, right? So basically you were coming from the round out you want to have the level out on to the cowl and once you were flaring your cowl right onto the threshold, right? All right, so　　landings are pretty good, so the 7th you were good but you wanted to adjust the　　on the center line so you got little bit to the left and…
8 Trainee: ***

教官は 6 回目に非常に良くなったことを挙げ（発話 5 太波線部）何が変わったのかと問いかけ（発話 5 点線部）、訓練生の答え（発話 6）を受けて大事な点を確認（発話 7 二重線部）、描写という形を取りながらどのような状態が望ましいのかを説明（発話 7 波線部）・確認（二重線部）する。

9 Instructor: But that's fine unless you are like too much on the first like the 1st and 2nd landing you were a little bit too much on the the the 7th 8th and 9th they were fine. If it was a stage check that will be satisfactory with that one. Umm once you land, don't use too much rudder, cuz that's how you gonna lose directional control, right
10 Trainee: That was good about
11 Instructor: Yeah, that was good. They're better way better and once coming back from GFK, what I foundout is you can set Ground frequency and also you can set Tower frequency. We are getting close to. So that's why I set it up for you, right? So before you before you do and also you got a little late on Approach briefing right?
12 Trainee: Yep
13 Instructor: You set all the way up to Billboard. So when you are coming back from GFK, what you need to do is you started up and listen to ATIS. Right after the ATIS you just do the Approach Briefing. Then don't allow you to how do you think are you able to enter the Airspace. Cause it's gonna change it might change, right? Even Tower might say alright enter left base or left downwind for runway 17L. It might change that it might never mind, extend your downwind from power plant before going to Billboard, finish up the Approach Briefing so that you know which runway you are gonna be using and how long it's gonna take to stop all that stuff. Right so overall, everything was great. Good job man.

教官は後半の着陸を褒めた後で一気に課題と修正法を説明し（発話 9、11、13 波線部）、最後に全体を褒め（発話 13 太波線部）ている。

<English05>

> 13（続き）Instructor: Do you have any questions? You have to ask at least 1 question
> 14 Trainee: Under 200AGL, if the airspeed is low and, path is red red red white. On the stage, I have to go around?
> 15 Instructor: On the stage I would do it
> 16 Trainee: Today's
> 17 Instructor: On the fourth one? Yeah if it was on the stage, but we are doing as a practice.
> 18 Trainee: Yes.
> 19 Instructor: And also, running out of time to excel so, that's all feel to landing on the stage if it looks a little bit wrong, just go around. They are gonna see go around better than landing
> 20 Instructor: All right? So that's good job

振り返りが一通り終了した後で教官は質問があるかと尋ね、訓練生は着陸すべきかゴーアラウンドして再度着陸の態勢を整えるべきか判断のポイントを質問する。これは日本語でも全く同様の質問を訓練生がしており初期段階では誰にとっても課題となる点と思われる。

4.4 英語 PFB の言語使用の特徴

日本語では操縦動作を説明する際にオノマトペの使用が見られたが、英語では多くの場合、副詞や形容詞等の使用で説明されていた（発話 5、9 網かけ部）。

<English06>

> 4 Trainee: Can I after 200ft AGL?
> 5 Instructor: If it's quick one because what I saw was like 59(kts), so bump up the power really quick. I would say okay with that one.
>
> 9 Instructor: But that's fine unless you are like too much on the first like the 1st and 2nd landing you were a little bit too much on the the 7th 8th and 9th they were fine. If it was a stage check that will be satisfactory with that one. Umm once you land, don't use too much rudder, cuz that's how you gonna lose directional control, right

5. 考察

日本語と英語では以下のような特徴が挙げられる。
・日本語では訓練生が描写（操縦動作、風景等）をするのに対し英語では教官が行う
・日本語では訓練生の描写に基づき、教官とのやり取りを通して次回への修正法を理解するのに対し、英語では教官の描写に基づき、質問に答えることで修正法を理解する
・日本語ではこのやり取りの中で比較、分析、評価、創造し理解する

・操縦動作の説明は、日本語ではオノマトペ等の使用が多いのに対し、英語では副詞や形容詞等を使ってなされることが多い

日本語と英語の比較を表 4 にまとめた。なお、今回紙面の都合上記載していないその他の PFB の文字起こしとその分析でもほぼ同様の傾向・結果が得られた。

今回分析対象とした英語の PFB は、日本語と比較すると所要時間は 3 分の 1 であるが、訓練生の総合的英語力が比較的高かったため、質問に短く答えているだけではあるものの、内容は理解しており質問で尋ねている内容も的を射ている。主に訓練生が描写する日本語でのスタイルに対して、英語では教官が主に描写を行っているが、要所での教官からの質問には的確に答えている。発話回数も所要時間が 3 分の 1 であることを考えると、短い返答ではあるが回数は少なくはない。よって、やり取りは成立していると思われる。

表 4. 日本語 PFB と英語 PFB のスタイルおよび言語使用の比較

	日本語（約 30 分）	英語（約 10 分）
やり取り	16 回	6 回
使用語彙	オノマトペ、感覚的表現の使用	副詞や形容詞
訓練生からの発話	訓練生 1：27 回 訓練生 2：11 回	9 回
発話内容	描写（操縦動作、風景）、比較、分析、評価、修正案の提案	教官からの質問に答える：yes/no 等（4 回） 質問する：パワーを入れるタイミング、良い・悪い点の確認（2 回）、着陸判断のポイント（1 回）
訓練生からの質問	訓練生 1：1 回（ラダーの踏み方） 訓練生 2：1 回（着陸判断のポイント）	3 回（パワーを入れるタイミング、良い・悪い点の確認、着陸判断のポイント）

やり取りは成立しているが、認知的学習という観点から higher order thinking （Anderson et al, 2001）が起こっているのか、という点では疑問が残る（図 1）。

日本語では描写を訓練生が自ら行いイメージングをしながら訓練中にできたこととできなかったことを比較し、できなかった操縦や動作についてはなぜできなかったのか、原因は何かについて考え（analyze, evaluate）、次回は問題点を解決できるよう対策や修正法を考え準備計画を立てる（create）ことが、教官とのやり取りを通して発生しているようである。

Higher order thinking skills

Create
Evaluate
Synthesize/analyze
Apply
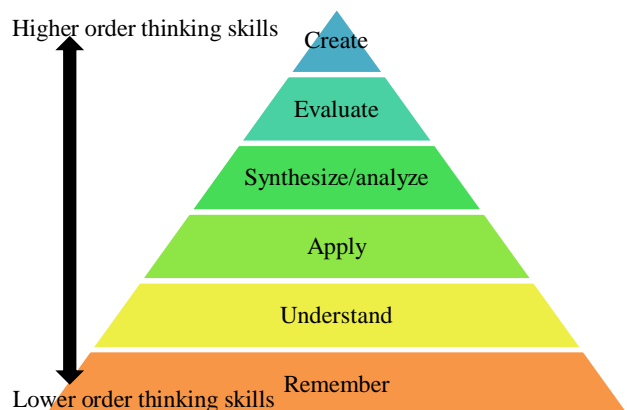Understand
Remember

Lower order thinking skills

図 1. 認知プロセスと思考スキル (Anderson et al., 2001 に基づいて著者が作成)

これに対して英語では、できたこととできなかったことの比較、何を修正すべきかについては質問と応答という形でできていたが、原因の追究と具体的な修正法、細かな操作法の説明等は非常に簡単に済まされていた。これは言語力によるものなのか、教授法の違いによるものなのかについては今回の分析のみでは判断できない。解明のためには、英語力の低い日本人訓練生や米人訓練生のケースを分析し比較する必要がある。

また、日本語では操縦手順や動作についての詳細な説明がオノマトペを使用し時間をかけてなされていた。これに対し、英語では形容詞や副詞の使用によりなされていた。

そもそも、英語やヨーロッパの言語にはオノマトペにあたるものが少ないといわれる（Tsujimura, 2013）ので、英語ではその使用がみられないのは言語学的に当然ともいえる。このような違いがもともと存在する上で日本語 PFB でオノマトペの使用がみられたのは、PFBでのトピックが操縦動作の説明であることが考えられる。操縦技能の習得のように、力加減や速さ等身体感覚とともに procedural knowledge（Anderson et al., 2001）を習得する場合には、オノマトペ等の使用による細かな感覚的表現が有効であると考えられる。Procedural knowledge（手続き的知識）とは"subject specific skills and algorithms" や "techniques and methods" "criteria for determining when to use appropriate procedures" 等の知識を意味し、"factual knowledge" や "conceptual knowledge" といった宣言的知識と異なり、実践を通して体得する種類の知識、スキルである（表 5 下線部）。例えば、スポーツ領域では身体の動き、スキル、技術等 procedural knowledge を習得することが大きな目標となるが、この指導の際にオノマトペが活用される（藤野，2008）。動きを表現する、体感する、イメージングを行う等にはオノマトペは非常に有効とされるからである。よって、運動技能の習得と同様に、操縦技能指導はスキル習得の側面が強いことを考えると、日本語 PFB で使用がみられたことは不思議ではない。ただし、これは教官の教授スタイルによるものである可能性もあるので今後多く

のケースを分析して比較する必要がある。

このように、英語の PFB では高次元学習が起こっているという例証が得られなかった。今回の訓練生は英語力が高く質問に応答する形で最低限の比較や確認はできていたが、より高次元の学習が起こっていたかは判断ができない。断定することはできないが、訓練の初期段階において高次元の認知的プロセスを経て自らを分析できていた可能性は低いと思われる。

表 5. Knowledge dimension （Anderson et al., 2001, p. 46 に基づいて著者が作成）

concrete	A. Factual knowledge	● knowledge of terminology ● knowledge of specific details and elements
	B. Conceptual Knowledge	● knowledge of classifications and categories ● knowledge of principles and generalizations ● knowledge of theories, models, and structures
	C. Procedural Knowledge	● knowledge of subject specific skills and algorithms ● knowledge of subject-specific techniques and methods ● knowledge of criteria for determining when to use appropriate procedures
abstract	D. Metacognitive knowledge	● strategic knowledge ● knowledge about cognitive tasks, including appropriate contextual and conditional knowledge ● self-knowledge

それでは、高次元学習を発生させるには具体的にどのような認知的プロセスが必要なのであろうか。高次元のプロセス（analyze / evaluate / create）を通して知識を学習するために必要な具体的認知プロセスは大きく8 つに分類（表 6）できるという（Anderson et al., 2001）。

表 6. 高次元学習に必要な認知的プロセス（Anderson et al., 2001, p.67-68 に基づいて著者が作成）

Analyze	Evaluate	Create
● differentiating ● organizing ● attributing	● checking ● critiquing	● generating ● planning ● producing

飛行技術に関する procedural knowledge をこれらの認知的プロセスを通して習得する場合、PFB 時に以下が英語でできる必要があると考えられる。
● 操縦時の描写（動作、機械操作、風景等）ができる
● 副詞や形容詞を使って操縦動作（速さ、力加減、タイミング等）を説明できる

- 飛行中の操縦動作について教官の指摘を理解しできたこととできなかったことを比較・分析することができる
- 描写・比較・分析・評価をもとに質問することができる
- 質問への答えを理解し、知識を組み立てなおし、次回への対策を創造することができる

　これらが英語でできれば、PFB 時に訓練生の方から積極的に描写や比較、分析、評価、創造をする機会が増える。例えば、事前にどの段階のレッスンではどのような描写や操縦動作の表現が多いのか等を状況別に整理し、対応する描写、比較、分析、評価の機能を持つ文法項目や語彙を学習しておけば内容理解もしやすく、質問もしやすくなるであろう。操縦技術の習得は振り返りやイメージングを PFB 中に正確にできるかどうかが大きなカギを握るとされるが、非英語母語話者の場合は英語力の差が習得技術の質や進捗に大きく影響すると考えられる（西川，2019）。よって、上記の点が英語でできるようになれば、より質の高い技術習得が可能になる可能性がある。英語力以外にも適性やカリキュラムの違い等も関係してくるが、少なくとも言語力による負荷を軽減することはできるだろう。また、英語力の低い訓練生にとっては大きな助けとなると思われる。

6.　まとめと今後の課題
　本研究では日本語と英語それぞれで行われる PFB 時のやり取りの特徴を比較し、Bloom's revised taxonomy（Anderson et al., 2001）に照らして望まれる力を検証した上で必要な英語力を考えた。特徴と比較については考察のセクションでまとめた。今回のケースでは、英語では高次元学習が起こっているという例証が得られなかった。断定はできないものの、起こっているとは考えにくいという結論に至った。
　本研究の問題点として、言語力によるものなのか教授法の違いによるものなのかについては今回の分析のみでは判断できない。解明のためには、英語を母語とする学生と教官とのやり取りも分析し、やり取りの回数や内容を比較する必要がある。英語母語話者同士でもやり取りの回数や内容が同様であれば教授スタイルであると考えられ、日本人訓練生が英語でできる必要があることについて再考の必要がある可能性もある。また、英語力の低い日本人訓練生ではどのようなやり取りになるのかについても分析し比較することで、より詳細に必要な英語力が検証できるであろう。今後は事例を増やすことで、より多角的に必要な英語力を検証していく。

注
[1] 航空大学校は、航空機の操縦に関する学科及び技能を教授し、航空機の操縦に従事する者を養成することを目的として、運輸省の付属機関として 1954 年に設置された日本における唯一の公的な操縦士教育訓練機関である。開校以来、航空輸送の中枢を担う操縦士を計画的・安定的に養成してきており、2015 年 1 月 1 日現在、本邦航空会社の操縦士（5686 人）の約 4 割（2266 人）が同校の出身者である。
[2] 東海大学の場合は、最初の 1 年間は日本で学科科目を受講し必要な国家試験を受験する。その後アメリカの自家用操縦士ライセンス課程を開始し、実機飛行訓練は全て北米の UND で行う。

参考文献
Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., … Wittrock, M. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. London: Longman.
藤野良孝. (2008). 『スポーツオノマトペ　なぜ一流選手は「声」を出すのか』東京：小学館.
国土交通省. (2014). 『交通政策審議会航空分科会基本政策部会　技術安全部会乗員政策等検討合同小委員会とりまとめ参考資料』https://www.mlit.go.jp/policy/shingikai/s304_jyouin.html　（最終閲覧日 2018 年 10 月 6 日）.
Nishikawa, M., & Nawata, Y. (2019, May). *Identifying the English language skills required by non-native English speaking pilot trainees during flight training*. Paper presented at the International Civil Aviation English Association Conference Tokyo 2019, Chiba, Japan. .
西川惠. (2019). 「非英語母語話者である航空機操縦士訓練生に必要な英語力とは何か」『大学英語教育学会 ESP 研究会年度レポート』vol.20, 5-15.
Tsujimura, N. (2013). *An introduction to Japanese linguistics*. Oxford: Blackwell.

日本の英語学習者を対象とした EAP ライティング教材研究
―EGP から EAP への接続に注目して―

マスワナ 紗矢子*・渡 寛法**・田地野 彰***

*目白大学
maswana@mejiro.ac.jp
**滋賀県立大学
watarihironori@gmail.com
***名古屋外国語大学
tajino@nufs.ac.jp

要旨
近年、日本の大学英語教育では、学術目的の英語（EAP）教育が注目されている。しかし、一般目的の英語（EGP）を主対象とする中学・高校の英語教育から、大学での EAP 教育への円滑な移行に配慮したライティング教材が不足しているという課題がある。このような背景のもと、本研究では、EGP と EAP との接続には何が重要かを考えるため、高校の英語教科書のライティング活動と、日本の大学で使用されている EAP ライティング教科書 10 冊を調査し、それぞれの特徴を分析した。分析の観点は、学習アプローチ、語彙・文法、批判的思考、文献資料の活用の 4 点とした。高校では文法に注目した文単位の作文が中心で、EAP ライティング教科書には、EAP に必要となる語彙・文法や批判的思考に関連するタスクが不足しており、これらを段階的に育成する教材が必要であることが示唆された。

キーワード：EAP、ライティング、教科書、高大接続

1. はじめに

2000 年代以降、日本において学術目的の英語（English for Academic Purposes: EAP）教育、その中でも特にライティング教育が注目されている。一方で、飯島他（2016）や大学英語教育学会 EAP 調査研究特別委員会（2018）が指摘しているように、日本の大学英語教育の文脈に適した教材の不足が課題となっている。「日本の文脈」については、高等教育での英語使用状況や英語教員の専門分野等、様々な観点からの検討が可能であろう。本論文では、英語教育のみならず教育全体の課題となっている高大接続の観点から考える。まず、高校までの一般目的の英語（English for General Purposes: EGP）と EAP は、理論上では接続しているとされるが（田地野・水光, 2005）、英語の目的や教育機関が異なるために、その内容やスキルの関連性に注意が払われることはあまりない。例えば、日本のアカデミックライティング授業では、欧米圏で開発されたライティング教科書がそのまま使われることも多い。EAP カリキュラムが進んでいる香港などでは、学生に合わせた教科書を独自に開発しているが（Chen, 2016）、日本で教材開発を行っている大学はいまだ少ない。

ライティング教育を考える上で、学生の文章力は重要な要素である。一般的に日本人学習者は、英語ライティングが苦手であると言われている。例えば、文部科学省（2017）が行った英語力調査テストでは、調査対象と

なった中学 3 年生の「書くこと」の無得点者の割合は 11.0%、高校 3 年生では 15.1% であったと報告されている。日本人学習者が英語ライティングを苦手とする理由として、高校までの英語教育は文法訳読と和文英訳が中心であることが挙げられる。自分の考えをまとまりのある英文で表現する学習機会が限られていることに加えて、大学生の多くは大学入学以前に日本語でもライティング教育を受けていないことが指摘されている。島田（2012）は、高校までの国語教育では、読解に重きが置かれており、学生はまとまった分量の文章を書く経験や指導を受けていないと述べている。よって、英語ライティングの基本であるパラグラフライティングを大学で初めて学び、躓く学生も多い。

このような背景と大学教育では過去の学習経験が活かされていないという指摘（渡辺・島田, 2017）を踏まえ、EGP と EAP のつながりの部分（図 1 重なりの部分）に注目した研究が必要であろう。この部分は、高校の英語学習と大学初年次の EAP（より正確には一般学術目的の英語 English for General Academic Purposes: EGAP）とみなすことができる。
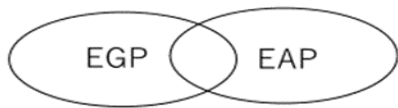
図1.　EGP と EAP の関係

　より良い EAP ライティング教材開発を目指す場合、この重なりの部分に焦点をあてること、すなわち高校までのライティング教育を念頭に置きながら、現在使用されている EAP ライティング教科書を検討する必要がある。そこで本研究では、以下の研究課題を設定した。

RQ1　現在高校で行われているライティング指導にはどのような特徴があるか。

RQ2　現在大学で使用されている EAP ライティング教科書にはどのような特徴があるか。

　以上の調査結果に基づき、ライティングにおける EGP から EAP への接続について考察を行う。

2.　先行研究

2.1　高校の英語ライティング指導

　高校ではそもそも「ライティング」という科目がなく、現在英語のライティング指導が主に行われている科目は「英語表現」である。英語表現は、基本的な言語規則に基づいて、様々な場面に応じて適切に話すことや書くことができるようになることを目的としており、スピーキングとともにライティング指導が行われている。高校の英語教育では、パラグラフライティングなどの指導も一部では実践されているが（仲川, 2017; 大井, 2008）、多くの高校英語では、一文単位の和文英訳や、文法学習としての英作文がライティング活動の中心である（Kobayakawa, 2011）。

　「英語表現」について、孫工・江利川（2019）は、「高等学校学習指導要領解説外国語編・英語編」（文部科学省, 2009）で示されている「批判的思考力を養うことをねらいとして内容を構成する」という点が不十分であることを指摘している。「英語表現Ⅰ」教科書 11 冊のライティングを含むすべての設問題材を分析したところ、言語形式の設問は 8 割を超えるが、批判的思考の設問は平均して 15%であった。

　上記のように高校におけるライティングやライティングを扱う教科書に関する研究はいくつか見られるが、体系的な英語ライティング教育に向けた EGP から EAP への接続に注目した研究は見られない。

2.2　EAP ライティング教材

　アカデミックライティングには、大学授業のレポートから、学術論文まで様々なジャンルがある。そのため、EAP ライティングについては、ジャンル・アプローチに基づいた研究や教材開発がなされてきた（例 Swales

& Feak, 2000, 2012）。一方で、EAP 教育は文脈によって必要となるスキルやタスクが大きく異なるため（Ning, 2019; Stoller, 2016）、一律的な EAP 教材評価リストや教科書分析法の開発は困難である。英語圏でもカリキュラムにおけるライティング教育の位置づけや内容、EAP ライティングの捉え方は異なり（Tardy & Jwa, 2016）、ライティングの書き方にも違いがあることが示されている（Purves, 1986）。さらに日本などの EFL 環境では、学生の英語力やニーズも英語圏とは異なってくる。本研究では、アカデミックライティングは、英語での専門学習や学術研究に必要なライティングとする。日本の英語教育における EGP から EAP への接続に関連する観点には以下のようなものがある。

2.2.1　アプローチ

　EAP ライティングの教科書は、修辞、ジャンル、アカデミックリテラシーと大きく 3 つのアプローチに分けられる（Tribble, 2009）。修辞的アプローチは、Exposition・Argumentation・Classification など文章のパターンを学習し、一文からパラグラフ、そしてエッセイへと進み、プロセス・アプローチに基づく。近年、ピア・レビューやルーブリックがプロセスに基づくライティング学習には有用であると言われている（Berg, 1999; Wang, 2014）。ジャンルに注目したアプローチは、ジャンルの目的に沿ってディスコースコミュニティの言語使用を学ぶものである。最後は、議論や発表などを含む学問修得の基礎となるアカデミックリテラシーを軸としたアプローチである。

2.2.2　語彙と文法

　EAP 研究では、コーパスに基づく研究成果を教材に反映すべきであると言われている（Harwood, 2005; Stoller, 2016）。教科書にも応用できる Coxhead（2000）の Academic Word List（AWL）や Browne, Culligan, and Phillips（2013）の New Academic Word List など、アカデミックライティングで使用頻度の高い語彙リストが発表されている。語彙のみならず、コロケーションや多く使用される文構造などの研究結果を取り入れた EAP 教材の重要性も述べられている（Biber & Gray, 2010; Hyland, 2008）。EAP ライティングでは EGP ライティングに比べて語彙や文法がより高度で複雑化するため、指導の重要性も高くなる。

2.2.3　批判的思考

　EAP ライティングでは論理的、客観的に書くことが必要であり、批判的思考が根底にあるとされる。楠見・田中・平山（2012, p. 69）は批判的思考を「論理的、客観的で偏りのない思考であり、自分の推論過程を意識的に吟味する反省的思考である」と定義している。英国の EAP 教員で構成される学会 BALEAP は、知識を単に描写するのではなく、変化させる（knowledge transforming）能力として批判的思考を捉えている。その

ため EAP 教材には、知識を変化させるタスクとプロセス、インタラクション活動が含まれていることが重要である（BALEAP, 2008）。また、書くためにはリーディング（reading for writing）が必要であると近年認知されている（Hirvela, 2016）。これは批判的思考を身に付けるためのプロセスの一つとも言われ、次節の文献資料の使用と深く関わるものである。

2.2.4　文献資料の使用

アカデミックライティングでは文献資料などを適切に引用することが求められる（Swales, 1987）。自分で情報を探す、読む、読んだものを正しく引用する、言い換える、要約するというのは、大学では汎用的なアカデミックスキルであり、リーディング能力や批判的思考力とも関連する重要なライティングスキルであると言える。日本の EAP 教員は、当該スキルは EGAP のライティング授業で身に付けておくべきものと捉えている（マスワナ・田地野, 2018）。

以上の観点以外にも考慮すべき点はあるが、本研究では EAP ライティングの基本と考えられる上記 4 点（アプローチ、語彙と文法、批判的思考、文献資料の使用）に絞って検討を行う。

3．方法
3.1　教材
3.1.1　高校の教科書

「英語表現 I」は高校生の約 7 割が履修している（大井, 2015）。当該科目で使用されている教科書のうち、平成 31 年度東京都立高等学校で採用校数上位 3 冊を取り上げた。教科書は『Revised Vision Quest English Expression I Standard』（VQS）啓林館、『Select English Expression I New Edition』（SL）三省堂、『Revised Vision Quest English Expression I Advanced』（VQA）啓林館である（東京都教育委員会, 2018）。

3.1.2　EAP ライティング教科書

教科書の特徴分析を行うため、国内 EAP カリキュラムに関する先行研究（飯島他, 2016; 渡・マスワナ, 2017）をもとに EAP 教育を実施している大学において大学 1 年生レベルで使用されている市販教科書 10 冊を収集した（教科書のリストは付録参照のこと）。6 冊は英語圏の出版社、4 冊は国内出版社から刊行されている。分析対象となった教科書を指定している授業シラバスの授業目標は"to acquire basic academic writing skills"などとなっており、基本的なアカデミックライティング技能の習得を目指す授業で使われている。

3.2　分析
3.2.1　高校の教科書

教科書内で、どの学習内容・スキルを対象とする設問が扱われているかについて、文法設問、語彙・表現設問、ライティング設問を抽出し、そのページ数の割合を算出した。複数に分類されうる設問については、設問の主たる活動に基づき分類した。その後、ライティング設問にはどのようなものがあるかを精査した。EAP 教科書の 4 観点についても次節の通り分析した。第一著者が分類を行った後、英語表現 I 指導経験のある研究者が分類を確認した。

3.2.2　EAP ライティング教科書

2 章で挙げられた 4 点について分析を行った。

観点 1 の「アプローチ」は、Tribble（2009）に基づき分類した。プロセス・アプローチの一環で用いられることの多いピア・レビュー活動と教科書のルーブリック提供の有無も調査した。ルーブリックには、セルフチェックリストも含めた。

観点 2 の「語彙と文法」関連問題がどの程度含まれているか、語彙および表現学習のための設問のページ数と文法問題のページ数をそれぞれ調べた。文法説明のみのページは除いている。

観点 3 の「批判的思考」は、リーディング設問を抽出した。抽出されたリーディング設問が内容理解のみ、もしくは批判的思考を問う問題であるかを検討した。

観点 4 の「文献資料の使用」は、文献資料の収集を必要とする設問、引用をさせる設問について抽出した。引用については、別種の設問で引用が求められている場合は、引用としてもカウントしている。サマリーとパラフレーズについては、提示された文章か、学習者自身で探した文章か、に関わらず設問として抽出した。

以上、EAP ライティング指導経験のある著者らが別々に分析したのち、相違点については再検討を行い、結果を統一した。

4．結果
4.1　高校のライティング指導
4.1.1　教科書の全体的特徴

表 1 は「英語表現 I」の教科書 3 冊のタスク概要である。VQS と VQA は同じシリーズの異なるレベルであり、設問内容は異なっていても教科書の構成が同じことから同一の数値が出たと思われる。表 1 で示した設問以外の部分（それぞれ 52%）は主に、文法説明、スピーキング活動である。表現することを中心とした授業であるものの、文法が中心に置かれ、ライティング設問は 5%〜13%と文法設問に比べるとかなり少ない割合となっていた。

表 1. 設問の概観

	文法設問	語彙・表現設問	ライティング設問	全ページ数
VQS	24%	11%	13%	120
SL	35%	8%	5%	75
VQA	24%	11%	13%	120

注：付録を除いたページ数

文法設問はすべて、基本的には文レベルの練習と定義されるエクササイズ（Ellis, 2009; 高嶋, 2000）であった。習得を目指す文法項目が明らかで、その定着を図るための練習である。例えば、VQS において不定詞を学ぶレッスン（p. 53）では、

「日本語に合うように、（　）内の語句を並べかえて英文を完成させなさい。」
「音楽を聴くことは楽しい。（ fun / is / it / listen / music / to / to ）.」といった設問に続き、
「あなたの将来の夢について、3つの文を書いてみよう。」とあり、例として「My dream is to be a vet. I like animals very much. I want to help sick or injured animals.」が提示されている。目的とする文法項目の不定詞が明示された文法学習と捉えることができる。各レッスンにこのような形式の設問が出されていた。

4.1.2　ライティング設問

教科書 VQS と VQA には各レッスンに「Expressing」というページがあり、ライティング設問が出されている。ライティング設問のトピックは多岐にわたり、VQS と VQA では、自己紹介、興味、旅行、スポーツ、観光、交際、娯楽、校則、英語学習、文化、社会問題、悩み事、SL では発明品、職業、日本文化、結婚といったものであった。ライティング設問は、特定の言語表現の習得を意図したものであった。

例えば、VQA（p. 68）では以下のメールを書くというタスクを通して、勧誘・申し出る機能表現を学ぶことが目的とされていた。

Write an email to ask Emily out. Add the details – activity, date, place, etc.

Dear Emily,
How are you doing? I'm fine………………………………
………………………………………………………………………
………………………………………………………………………
………………………………………………………………………
I'm looking forward to your reply.
Your friend,

これは先述の文法問題とは異なり、言語項目が具体的には明示されておらず、エクササイズというよりは、Ellis（2009）の言うフォーカスされたタスク（focused task）と考えられる。一方で、SL ではライティングの設問は少なく、スピーチ原稿を書くことが主なライティング活動となっていた。職業についてのライティング（p. 39）では、以下のように一文の空欄を埋めていくものである。

①なりたい職業 I want to be ＿＿＿＿＿＿＿＿＿＿＿.
②理由の数 I have ＿＿＿＿＿＿ for choosing this job.
③理由 1 First, ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿.

④理由 2 Second, ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿.
⑤理由 3（最後に）Finally, ＿＿＿＿＿＿＿＿＿＿＿.
⑥結びのことば That's why ＿＿＿＿＿＿＿＿＿＿＿.

順序立てて表現するという学習目的はあるものの、自分のなりたい職業とその理由を述べる活動がメインであり、これもフォーカスされたタスクといえよう。

ライティング設問を検討したところ、EAP 教科書のアプローチが該当するものはなかった。文法と語彙問題については、ライティング設問よりも豊富であった。批判的思考に関しては、自分の意見や経験を述べることが目的となっており、「論理的、客観的で偏りのない思考」といった大学で求められるような批判的思考の定義を用いた場合には該当する設問はなかった。文献資料の使用についても、高校の教科書でのライティング設問では必要とされていなかった。

4.2　大学の EAP ライティング
4.2.1　教科書の全体的特徴とアプローチ

教科書の概要とアプローチを表 2 に示す。分析対象の教科書は、J2 を除き 2010 年以降に出版されたものである。平均ページ数は 153 であるが、欧米圏の教科書は、日本の教科書よりもページ数が多い傾向であった。扱われているトピックは身近な話題から科学まで多岐にわたり、E5 のみが Engineering, Medicine など学術的な分野に関連付けて提示していた。

表 2. 教科書概要とアプローチに関する結果

	E1	E2	E3	E4	E5	E6	J1	J2	J3	J4
出版年	2011	2014	2019	2012	2014	2017	2010	2008	2013	2019
ページ数	107	279	211	183	146	119	213	87	95	86
アプローチ	R	R	R	R	G	R	G	R	R	G
ピアレビュー	○	○	○	○	-	-	○	-	○	○
ルーブリック	-	○	○	○	-	-	-	-	-	○

R=修辞的アプローチ、G=ジャンル・アプローチ

アプローチについては、修辞的アプローチに分類されるものが 7 冊と多く、パラグラフの修辞パターン（例 Cause and Effect, Classification）に基づいてそれらを順に習得する構成となっていた。Tribble（2009）が修辞的アプローチはプロセス・アプローチに基づいているとした通り、教科書もプロセスライティングを基本としていた。ピア・レビューについては、修辞的アプローチを使っている 7 冊のうち 5 冊が扱っていたが、ルーブリックについては、3 冊のみであった。ピア活動やルーブリックは、近年日本でも重要性が強調されているが、すべての教科書には用いられていなかった。E5・J1・J4 はジャンル・アプローチに分類されるもので、パラグラフの書き方といった基本的なライティング技能についても扱っているものの、その指針は、ESP ジャンルの考え方に基づいたものであった。

4.2.2 語彙と文法

語彙・表現と文法設問に関する結果を表3に示す。語彙・表現と文法のタスクが10%以上含まれていたのは、すべて英語圏の教科書であった。語彙や表現の設問では、類義語・対義語や特定の機能表現（例 比較表現）、共起など様々なものが各ユニットに付随していた。語彙と表現、文法タスク以外の部分は、主にライティングプロセス（ブレインストーミング・アウトライン・ドラフト作成・読み返し・修正）に関する活動やモデル文章とその解説であった。

表3. 語彙と文法に関する結果

	E1	E2	E3	E4	E5	E6	J1	J2	J3	J4
語彙・表現設問	4%	10%	14%	1%	28%	0%	0%	0%	1%	1%
文法設問	4%	11%	14%	0%	9%	0%	0%	0%	0%	3%

注：付録を除いたページ数で算出

語彙と表現の設問が14%を超えるE3とE5は、Coxhead（2000）のAWLを参照していることが教科書に明記されていた。文法は、主語と動詞の一致や独立節と従属節といったものから、特定のパラグラフパターンで使われやすい文法、提示された文章の文法的誤りの指摘などの設問を含んでいた。アカデミックライティングで必要という観点での語彙や文法設問ではあるが、一文単位のものや空欄に適切な語を入れる等、日本の高校生に馴染みのある形式であった。国内の教科書の語彙・表現と文法設問は、非常に限られたものであった。

4.2.3 批判的思考

リーディング設問と批判的思考に関して見ていく（表4）。

表4. リーディング設問に関する結果

	E1	E2	E3	E4	E5	E6	J1	J2	J3	J4
モデル文章分析	○	○	○	○	○	○	○	○	○	○
批判的リーディング	-	-	-	-	○	-	○	-	-	-

全教科書にリーディング設問があり、モデル文章の構造分析と内容理解を目的としていた。以下は典型的な設問例である。

・Underline the topic sentence.
・According to this paragraph, what are the fundamental steps in the process of doing well in class?（E2, p. 104）

一方で、内容に対する議論などの設問は少なかった。EAPでは批判的思考の重要性が強調されつつあるが、今回分析対象とした教科書でそうした設問を含んでいたのは2冊であった。例えば、E5では広告には赤色が効果的であるという文章について、構成を確認する設問に続いて、「Do you believe the statements made in the article? For example, do you believe that red increases the appetite and attracts attention? Why or why not?」（p. 33）という設問があり、文章の内容について自分なりに判断することが求められていた。

J1にはcritical readingについてのセクションがあり、批判的に読むためのポイントが説明され、実際に文献を選び批判的な読みを練習する設問があった。

4.2.4 文献資料の使用

最後に、文献の使用に関する結果を示す（表5）。

表5. 文献資料の使用に関する結果

	E1	E2	E3	E4	E5	E6	J1	J2	J3	J4
文献収集	-	-	-	-	-	○	○	-	-	○
文献の引用	-	-	-	○	○	○	○	-	-	○
サマリー設問	-	○	-	○	○	○	○	-	-	-
パラフレーズ設問	-	○	○	-	-	-	○	-	-	-

文献資料の収集については、3冊が扱っているのみであった。パラグラフライティングを中心にした教科書では、トピックに対して自分の意見を出して、理由を整理することに重点が置かれている。リサーチに関連したスキルを扱った教科書では、引用とサマリーはそれぞれ5冊、パラフレーズについては3冊のみ設問がつけられていた。

ジャンル・アプローチを採用しているJ1とJ4では、リサーチペーパーを書くことを念頭に、剽窃に関して注意喚起を行い、情報収集と引用が比較的詳しく解説されていた。E4・E5・E6については、サマリーパラグラフを書く活動のなかで引用が説明される程度にとどまっていた。

5. 考察

本研究では、日本の英語学習者を対象に、EGPとEAPを接続するEAPライティング教材開発のため、高校と大学の教科書の特徴分析を行った。高校では「ライティング」という科目がないため、EAPライティング教科書との直接的な比較はできないが、それぞれのポイントを議論した上で提案を行う。

高校の教科書では、特定文法項目の使用や定着を目的とした和文英訳などの設問が多く、ライティングでは自分の意見を表現することに焦点がおかれ、批判的思考や文献資料の使用に関する設問は見られなかった。今回調査したEAP導入段階の教科書の多くは、パラグラフライティングの習得を目的とし、プロセスライティングのアプローチで、ターゲットとなる修辞パターンに基づいて学習内容が構成されていた。プロセスライティングで重要なピア・レビューやルーブリックを取り入れていない教科書が散見され、この点において発展の余地が見られた。

本研究が調査した国内のEAPライティング教科書では、文法説明や文法設問、語彙や表現練習の取り扱いが

非常に少なかった。アカデミックライティングは独立したスキルと捉えられていること、そして上級者向けであるため語彙・文法の設問は必要ないと考えられていることも推測できる。しかし、語彙・表現の学習は、アカデミックなレポートや論文作成では一層必要となり、現行の教科書では不十分といえる。実際に、教科書のほかに、独自で開発した学術語彙集を用いて学生の語彙力を高める試みを行う大学も近年では見られる（大学英語教育学会 EAP 調査研究特別委員会, 2018）。海外の教科書は語彙や文法を扱っていたが、主語と動詞の一致など基礎的な項目が多く、学生の既習事項とのミスマッチが見られた。アカデミックライティングにおいても文法や語彙の重要性は変わらないため、EAP ライティングに関連の深い練習問題をいくつか含むことが提案できる。

批判的思考を促すようなリーディング設問は限定的であった。Comparison/ Contrast などで、他者の意見を踏まえつつ、自分の意見を論理的に述べるライティングタスクは見られた。一方で、他者の文章を批判的に検討するような設問は 2 冊だけであり、Hirvela（2016）の言う EAP で必要となる Reading から Writing への連続性について意識している教科書は少ない。批判的思考については様々な捉え方があり、例えば、孫工・江利川（2019, p. 53）では以下の設問を批判的思考設問として扱っていた。

あなたが環境保護のためにしていることについて、下線部分を言い換えて話しましょう。
I recycle my waste paper in order to save forests [in order not to destroy forests.]

高校、大学それぞれの段階でどのようなタスクが、批判的思考の育成に必要になるかをより詳しく検討する必要がある。

文献資料の使用や引用スキルはアカデミックライティングでは必須とされるものの、パラグラフライティングの教科書では取り扱いが比較的少ないことが判明した。海外の教科書は段階的なシリーズであることが多く、文献資料の使用は、次のレベルでの導入となっている（例 E1, E2）。日本では近年、いわゆるコピペや剽窃が大きな社会問題にも発展しており、EAP 導入段階でも、学生に正しい引用の仕方を学ばせる必要があるだろう。

全体的に英語圏の教科書は国内の教科書に比べ語彙や文法タスク、ピア活動を取り入れており、教科書自体も重厚になっていた。英語圏の教科書は、学生のライティング力やニーズに合わせ、教員の裁量でタスクを取捨選択できるようになっているのに対し、日本の教科書は各ユニットが 90 分の授業時間内で完結するように構成される傾向が見られた。教科書の説明も日本語と英語で書かれており、日本の教育現場や教員のニーズをより反映していると言える。

本調査では、分析観点を全て網羅した教科書はなかったが、どの観点にも、複数当てはまる教科書があった。同じアプローチでも、強調しているポイントに違いが見られ、教科書の特性を確認することができた。1 冊の教科書で、すべての項目を網羅することが良いのか議論の余地はあるが、EAP 教材開発におけるチェックリストとして、この 4 観点が活用できることが今回の調査結果から示唆された。

6. おわりに

本稿では、高校と大学の教科書を分析し、高校でのライティング学習を考慮に入れた EAP ライティング教材の開発に向けて考察を行った。今後の EAP 教科書開発について 2 つの示唆ができる。第一に、EAP 語彙と文法の導入である。語彙内容や文法項目の検討は必要であるが、高校までの英語学習で馴染みが深いエクササイズであり、学習者にも有益であろう。高校の教科書では、表現力の基盤として、文法が重視されていたが、文法学習と大学でのパラグラフライティングには隔たりもある。これまでの日本の英語教育では、ライティングにつながる「文法」ではなく、「文法問題」が得意な学習者が育成されていることが指摘されている（田地野, 1999）。EAP ライティング教科書では、大学入試で採用されているような選択問題ではなく、アウトプットにつなげる文法学習の機会を提供する必要がある。

第二に、ライティング指導についての研究の知見を反映させた教材開発が求められる。今回調査したライティング教科書では、ライティング評価ルーブリックは半数以下の教科書のみで見られ、リーディング設問を使用して批判的思考を取り入れたものは限定的であった。これらに関する研究は盛んに行われており、最新の研究結果を取り入れた教科書開発が望まれる。現在、高校では学習指導要領に基づき「主体的・対話的で深い学び」、いわゆるアクティブ・ラーニングが導入されつつある。今後、大学のライティング教育においても学生の自律的学習を促す上で、ピア・レビューなどの研究知見が活用できるであろう。

最後に、本研究では分析した教科書の数や分析項目が限定的であったことが課題として挙げられる。また、今後は EAP ライティング教科書を使用する学習者のニーズや英語力についての調査を行い、教材研究を深めていく必要がある。

参考文献
BALEAP. (2008). *Competency framework for teachers of English for Academic Purposes*. (https://www.baleap.org/wp-content/uploads/2016/04/teap-competency-framework.pdf)（2020 年 2 月 21 日）

Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing, 8*(3), 215–241.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes,9*(1),2–20.

Browne, C., Culligan, B., & Phillips, J. (2013). *The New Academic Word List*. (http://www.newgeneralservicelist.org)（2020 年 2 月 21 日）

Chen, J. (2016). *EAP curriculum evaluation and English Across the Curriculum. Invited workshop*, JACET Summer Seminar, Kyoto, Japan, August 2016.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly 34*(2), 213–238.

大学英語教育学会 EAP 調査研究特別委員会. (2018). 「大学英語教育の質保証に向けた EAP カリキュラム実態把握調査」研究成果最終報告書（2014 年度～2017 年度）. (http://www.eiken.or.jp/center_for_research/pdf/bulletin/vol99/vol_99_17.pdf)（2020 年 2 月 21 日）

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics, 19*(3), 221–246.

Harwood, N. (2005). What do we want EAP teaching materials for? *Journal of English for Academic Purposes, 4*, 149–161.

Hirvela, A. (2016). Academic reading into writing. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for Academic Purposes* (pp. 127–138). Abingdon: Routledge.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4–21.

飯島優雅・渡辺敦子・マスワナ紗矢子・渡寛法・堀晋也・高橋幸・金丸敏幸・田地野彰・寺内一. (2016). 「日本の大学における学術英語カリキュラムの現状と課題―実態調査結果を踏まえて―」『京都大学高等教育研究』22, 95-98.

Kobayakawa, M. (2011). Analyzing writing tasks in Japanese high school English textbooks: English I, II, and Writing. *JALT Journal, 33*(1), 27–48.

楠見孝・田中優子・平山るみ. (2012). 「批判的思考力を育成する大学初年次教育の実践と評価」『Cognitive Studies』19(1), 69–82.

孫工季也・江利川春雄. (2019). 「高校英語教科書における社会問題を批判的に考える力の扱われ方―「英語表現Ⅰ」用教科書の設問題材分析を通じて―」『和歌山大学教育学部紀要』69, 51–56.

マスワナ紗矢子・田地野彰. (2018). 「アカデミックライティング技能と知識に関する英語教員の意識調査」. ESP 1 Day Conference at Takachiho University. 2018 年 2 月 20 日.

文部科学省. (2009). 「高等学校学習指導要領解説外国語編・英語編」(https://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2010/01/29/1282000_9.pdf)（2020 年 3 月 4 日）.

文部科学省. (2017). 「平成 29 年度英語教育改善のための英語力調査 事業報告」. (https://www.mext.go.jp/a_menu/kokusai/gaikokugo/1403470.htm)（2020 年 1 月 8 日）.

仲川浩世. (2017). 「高校教科書と大学教材内のパラグラフ・ライティングタスクの分析」『広島大学大学院教育学研究科紀要』第二部第 66 号, 103–110.

Ning, N. (2019). The suitability of EAP textbooks to the learning needs in Chinese context—From a pre-use perspective.

Education Quarterly Reviews, 2(2), 305–320.

大井恭子編. (2008). 『パラグラフ・ライティング指導入門―中高での効果的なライティング指導のために』東京：大修館書店.

大井恭子. (2015). 「4 技能試験時代のライティングに必要な指導と評価とは」『英語教育』64, 10–12.

Purves, A. C. (1986). Rhetorical communities, the international student, and basic writing. *Journal of Basic Writing, 5*(1), 38–51.

島田康行. (2012). 『「書ける」大学生に育てる―AO 入試現場からの提言』東京：大修館書店.

Stoller, F. L. (2016). EAP materials and tasks. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for Academic Purposes* (pp. 577–591). Abingdon: Routledge.

Swales, J. (1987). Utilizing the literatures in teaching the research paper. *TESOL Quarterly, 21*(1), 41–68.

Swales, J. M., & Feak, C. B. (2000). *English in today's research world: A writing guide*. Ann Arbor: University of Michigan Press.

Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students*. Ann Arbor: University of Michigan Press.

田地野彰. (1999). 『「創る英語」を楽しむ』丸善.

田地野彰・水光雅則. (2005). 「大学英語教育への提言―カリキュラム開発へのシステムアプローチ―」竹蓋幸生・水光雅則（編）『これからの大学英語教育』(pp. 1–46), 東京：岩波書店.

高島英幸編. (2000). 『実践的コミュニケーション能力のための英語のタスク活動と文法指導』東京：大修館書店.

Tardy, C. M., & Jwa, S. (2016). Composition studies and EAP. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for Academic Purposes* (pp. 56–68). Abingdon: Routledge.

東京都教育委員会. (2018). 「平成 31 年度使用都立高等学校及び中等教育学校（後期課程）用教科書教科別採択結果」. (http://www.metro.tokyo.jp/tosei/hodohappyo/press/2018/08/24/documents/05_00.pdf)（2019 年 11 月 7 日）.

Tribble, C. (2009). Writing academic English—a survey review of current published resources. *ELT Journal, 63*(4), 400–417.

Wang, W. (2014). Students' perceptions of rubric-referenced peer feedback on EFL writing: A longitudinal inquiry. *Assessing Writing, 19*, 80–96.

渡辺哲司・島田康行. (2017). 『ライティングの高大接続―高校・大学で「書くこと」を教える人たちへ』東京：ひつじ書房.

渡寛法・マスワナ紗矢子. (2017). 「大学におけるアカデミック・ライティング教育の動向―シラバスの共起ネットワーク分析から―」『東アジア英語教育研究』6, 41–50.

付録
教科書リスト

E1: Zemach, D. E., & Ghulldu, L. A. (2011). *Writing essays: From paragraph to essay*. Macmillan.

E2: Oshima, A., & Hogue, A. (2014). *Longman academic writing series 3: Paragraphs to essays* (4th ed.). Pearson Education.

E3: Folse, K. S., Vestri, E., & Clabeaux, D. (2019). *Great writing 3: From great paragraphs to great essays* (5th ed.). National Geographic Learning.

E4: Taylor, M. A., & Kluge, D. E. (2012). *Basic steps to academic writing: From paragraph to essay*. Cengage Learning.

E5: Burgmeier, A., & Lange, R. (2014). *Inside writing 1*. Oxford University Press.

E6: Boon, A. (2017). *Writing for fluency and accuracy*. Cengage Learning

J1: 田地野彰, ティム・スチュワート, デビッド・ダルスキー編 (2010).『Writing for Academic Purposes 英作文を卒業して英語論文を書く』ひつじ書房.

J2: 神保尚武, ケイト・エルウッド, 森田彰, 渡辺洋一, 山田茂, レオニード・ヨッフェ. (2008).『効果的なパラグラフの書き方』南雲堂.

J3: ケイト・エルウッド, レオニード・ヨッフェ. (2013).『効果的な英文エッセイの書き方』南雲堂.

J4: 深山晶子, 幸重美津子, 尾鍋智子, 村尾純子, アシュリー・ムーア. (2019).『3-Point Academic Writing: Organization, Content, Language』金星堂.

【Postscript 編集後記】

　昨年の JAAL in JACET proceedings 創刊号に続き、第2号をお届けすることができました。今回も教材開発から授業研究まで、英語教育に関する幅広いトピックの論文が 15 本、読み応えのある 1 冊となりました。3 月の発行にむけてタイトなスケジュールにも関わらず熱心に査読に当たっていただいた先生方に心より感謝申し上げます。発行にあたり、内藤永先生、飯島優雅先生、渡辺敦子先生、Web 担当の荒木瑞夫先生には特にご尽力いただきました。

　2020 年は新型コロナウイルスで大変な幕開けとなりました。こうして JAAL in JACET proceedings 第 2 号を無事に発行できたことに安堵しております。"Everything that happens twice will surely happen a third time." 第 3 号の発刊にご期待ください。

（渡寛法・加藤由崇）