

Investigating the Relationship Between Students' Attitudes Toward English-Medium Instruction and L2 Speaking

Shungo SUZUKI
Tetsuo HARADA
Masaki EGUCHI
Shuhei KUDO
Ryo MORIYA

Introduction

The Department of English Language and Literature in the School of Education at Waseda University started to offer English-Medium Instruction (EMI) for upper-division content courses in 2015. Furthermore, in order for freshmen to prepare for such EMI courses, two content-based "bridge" courses in English for academic purposes (EAP) are also required. This curriculum revision has been undertaken in response to the results from the departmental needs analyses and a societal expectation in Japan (Harada, in press). As previous findings on English learning in Japan suggest (e.g., Sato & Lyster, 2012), one of the serious problems is the severely limited use of English inside and outside of class. Considering the current situation, the faculty members share the guiding principles behind the revision that the Department should maximize the opportunities for students to use English meaningfully.

As a curriculum development cyclically proceeds, the evaluation of the curriculum or program is a vital stage of it (e.g., Christison & Murray, 2014). The current study, therefore, examines which aspects of English use undergraduate students in an EMI course are satisfied and frustrated with in a classroom-based exploratory approach, for the purpose of finding some optimal ways of teachers' language support in EMI. This paper begins with a terminological discussion of EMI, followed by how this classroom-based study was carried out, and finally discusses the findings from the data with regard to pedagogical suggestions.

Literature Review

EMI is commonly defined as a non-language course in English as a foreign language (Hellekjær, 2010). In other words, it is an academic content course entirely conducted in English. Historically speaking, the origin of EMI is rooted in higher education in Europe. Since the beginning of the Bologna Process in 1999, a number of universities began to offer courses in English to promote internationalization in Europe (see Smit & Dafouz, 2012). In addition, students who attend EMI courses are expected to be as capable of learning the contents in English as in their L1s (Hellekjær, 2010). Thus, EMI in European universities takes place simply by changing classroom language to English.

EMI in Japan is, however, different due to the nature of EFL settings, except for the fact that it is likewise given not as a language course, but as a content course. Initially, while few universities in Japan (e.g., Akita International University) offer EMI to attract international students from all over the world, the student population that receives EMI in Japan tends to be virtually homogeneous, mainly comprising Japanese-speaking domestic students (Harada, *in press*). Accordingly, internationalization cannot necessarily be a primary goal in Japanese universities. Moreover, as the majority of the students are EFL learners who have not already acquired an ability to use English functionally, EMI students and instructors implicitly share the consensus that the use of English in EMI might facilitate the students' English learning to some extent. Therefore, the unique characteristic of EMI in Japan is the fact that most students may be Japanese-speaking learners of English, consequently suggesting that students and instructors consciously or unconsciously regard EMI as one of the ideal opportunities for English learning.

In order to better understand how EMI can contribute to L2 learning, its pedagogical characteristics should be revisited in terms of existing SLA theories. EMI encourages students to use English meaningfully to learn academic contents. This situation is quite similar to the concept of content-based instruction (CBI), where the content serves as a vehicle for meaningful L2 use (Lightbown, 2014). Moreover, as the focus is at least on the content in both EMI and CBI, their syllabi are organized around the content or topic. It should be noted, however, that EMI fundamentally does not offer any deliberate

language instruction, unlike CBI by its nature.

Such a unique organization of syllabi for EMI can be realized through EMI lesson structure in a similar way to task-based language teaching (TBLT). In EMI courses, students use English meaningfully for a clear outcome (i.e., understanding the academic content) without deliberate language instruction. According to the literature on TBLT (e.g., Ellis, 2003), EMI can thus be regarded as a series of unfocused tasks which facilitate incidental learning without any preselection of target linguistic features (for a comprehensive review see Ellis & Shintani, 2013). In order for L2 acquisition to take place through unfocused tasks, students have to pay attention to form as well as meaning (i.e., focus on form, FonF; see Long & Robinson, 1998). FonF requires an optimal condition where the meaning of language is transparent so that students could afford to pay attention to its form (Ellis & Shintani, 2013). These theoretical frameworks in the fields of CBI and TBLT suggest that EMI instructors in EFL settings understand what kinds of language support for transparent meaning (i.e., content learning in EMI) are prioritized for EFL learners. As effective language support should be based on the diagnostic assessment of learners' self-perceptions of failure and success in classroom tasks (van de Pol, Volman, & Beishuizen, 2010), the current exploratory study, therefore, addresses the first research question:

RQ 1: Which aspects of tasks in EMI are students satisfied or frustrated with?

By answering the first research question, the study describes which aspects language support should aim for, using the diagnostic assessment. Moreover, as L2 learners' self-evaluation and attitudes are closely related to speaking competence (e.g., Mak, 2011), it is beneficial to examine the complex relationship between students' attitudes toward EMI and L2 speaking performance for further revisions of the curriculum. Thus, the second research question is formulated with the scope limited to speaking tasks in EMI:

RQ 2: Which aspects of speaking performance are sensitive to students' self-evaluation of their performance in EMI?

The answer to the second research question clarifies which aspects of L2 speaking are likely to affect students' self-evaluation. For the first research question, a questionnaire survey was conducted to investigate their backgrounds

and self-evaluation of their tasks in EMI. Additionally, for the second research question, participants were voluntarily recruited from the class. Their speech data were elicited outside the class via a prompt similar to the tasks in EMI to assess their speaking performance, which we regarded as one of the crucial factors influencing their self-evaluation.

Methodology

Participants

Participants were recruited from an undergraduate course in the Department of English Language and Literature at Waseda University. Whereas 21 students were officially enrolled in the course, 15 students completed all the parts of questionnaire due to the absence through teaching practicums and job hunting (3 sophomores, 9 juniors, 3 seniors; 7 males, 8 females). In addition, seven out of the 15 students voluntarily participated in the speaking session outside the class. All the participants were Japanese speaking learners of English at an intermediate level of proficiency (*MTOEFL ITP* = 519.1).

Target EMI Course

The target EMI course was an undergraduate elective course about Content and Language Integrated Learning (CLIL). The instructor (the second author) was Japanese and had 16 years of EMI teaching experiences at university. Every lesson lasted 90 minutes, consisting of five major components: (a) reading assignments before the class, (b) a quiz, (c) two students' presentations, (d) a lecture from the instructor, and (e) group discussions during both the lecture and students' presentations. All the tasks were conducted in English. Before every class, students were required to read around a total of 15 pages from two textbooks written in English. At the beginning of the class, they individually answered a quiz, in which they were asked to define key terms and concepts from the assigned reading and to answer an open-ended question. According to the researcher's observation, the quiz took 15 minutes on average, and played a role of priming for the following classroom tasks. The lecture from the instructor lasted around for 30 minutes. The lecture was based on one textbook, covering relatively difficult issues and concepts. As for student presentations, two students were assigned as presenters every week, and each presenter had

around 20 minutes to make his/her presentation based on the other textbook (in total, 40 minutes). Each presentation included a couple of group discussion questions, which encouraged other students to discuss keywords and controversial statements from the textbook in small groups, and then to summarize their ideas to the whole classroom. According to the observation, group discussions in each presentation lasted around 5 minutes on average.

Instrument Development

Questionnaire. The questionnaire in the study was originally developed in response to the course instructor's (the second author) and four TAs' (all the other authors) concerns about the reality of classroom. It aimed for three major issues: (a) which aspects of EMI students were satisfied or frustrated with, (b) why they decided to take the EMI course (i.e., reasons and expectations to the course), and (c) who were likely to take the EMI course (i.e., students' background). The questionnaire consisted of five parts with the three aims reflected.

The study presented in this article is part of our extensive classroom-based research, focusing on the first aim. Hence, the study focuses on part of the questionnaire, which asked students' self-evaluation of all the classroom tasks including reading assignments. Although the questionnaire items were created after Week 6 when the students got adequately accustomed to the lesson procedure, they seemed to have difficulties with group discussion in particular. The authors intentionally developed a detailed set of items about group discussions based on the *ACTFL proficiency guideline* (ACTFL, 2012), which had been developed based on the concept of communicative competence. As the EMI course offered a communicative situation which required both interpersonal and academic communication, the authors assumed that it was appropriate for the context of EMI. The target part of questionnaire adopted a 6-point scale and some of them were worded in an inverted scale.

Speaking test task. An argumentative task was used to elicit their speech outside the class. The speaking task reflected the characteristics of group discussion in EMI, where students were encouraged to express and justify their opinions in an academic manner (Suzuki, 2016). Task characteristics and conditions were specified to elicit the participants' upper limitation of

performance, following the literature on task performance (e.g., Skehan, 2014).

Data collection

Questionnaire. While a background language questionnaire was separately provided and collected in Week 7, the remaining parts were given in Week 8 and the students were encouraged to answer outside the class and submit to the instructor by Week 15. Although the difference in the collection time is one of the methodological limitations of the study, the authors adopted this way of data collection, considering that the participation in the questionnaire was voluntary and that securing the classroom time for the course content must be prioritized.

Speaking test task. Speech samples were elicited via an argumentative task around Week 14 to 15. For the procedure of the speaking task, the participants first planned the answer to the prompt for about two minutes and then performed their speech. The speaking performance was intentionally elicited outside the classroom individually to minimize the extraneous variables in the classroom such as the variation in content knowledge and the uneven dominance in speaking time, because eliciting their upper limit of performance was the most fundamental principle in the field of performance assessment (Ellis & Barkhuizen, 2005).

Data Analyses

Analysis of questionnaire items. In order to examine students' satisfaction and frustration with the in-class EMI tasks, mean scores were calculated for the questionnaire items directly related to the target attitudes (Items 11-45; $n = 35$) from 15 students. Due to the small sample size, the criteria for categorizing items were developed according to the median on the scale rather than the mean and standard deviation. As a 6-point scale was adopted, the median was 3.5 and the mean scores within 3.5 ± 1 were considered neutral about their satisfaction and frustration. Likewise, the mean scores 1 to 2.5 were regarded as frustrating, and the mean scores from 4.5 to 6 as satisfactory.

Analysis of speaking performance. The second aim of the study is to investigate the relationship between students' self-evaluation and dimensions of speaking performance. We analyzed the data from 7 participants who completed both the questionnaire and speaking test outside the class. For the analysis of self-evaluation score, items 27-38 ($n = 12$) relevant to the linguistic aspects of

performance in group discussion were used. There were two theoretical rationales for the analysis of these aspects. First, the speaking task reflected the characteristics of group discussion in EMI particularly (Suzuki, 2016). Second, the linguistic aspects of speech production can be measured by complexity, accuracy, and fluency (CAF) measures, which are regarded as the most comprehensive set of measures to capture the speaker's performance (e.g., Lambert & Kormos, 2014). Therefore, speech samples were assessed by several CAF measures. Following the previous CAF studies (e.g., Foster & Wigglesworth, 2016; Norris & Ortega, 2009), the developmentally appropriate measures were selected for each CAF domain as summarized in Table 1. The first author coded all of the analysis of speech units (AS-units) and clause boundaries, following Foster et al. (2000), and classified errors, following Foster and Wigglesworth (2016). Afterwards, the fifth author blind-coded 25% of them. According to Takeuchi and Mizumoto's (2014) criteria for Cohen's kappa, inter-coder agreements ranged from moderate to almost perfect (AS-unit boundary: $k = .804$, clause boundary: $k = .946$, and error classification: $k = .586$). The coding of the first author was included in further analysis.

Table 1.

Summary of CAF measures used in the current study

CAF domain	Measure	Definition
Productivity	Total # of words	The total number of words produced for the speech excluding dysfluency words
Syntactic complexity	Clauses /AS-unit	The mean number of clauses per AS-unit (Norris & Ortega, 2009)
Lexical complexity	Measure of textual lexical diversity (MTLD)	The mean length of sequential word strings in a text that maintain a given type-token ratio value (see McCarthy & Jarvis, 2010)
Accuracy	Weighted clause ratio (WCR)	The mean score of clause ratings according to the degree of error seriousness (see Foster & Wigglesworth, 2016)
Fluency	Words per minute (WPM)	The mean number of words in speech per minute excluding dysfluency words (Ellis & Barkhuizen, 2005)

First, descriptive statistics were calculated for the CAF measures to interpret students' oral proficiency. To examine which aspects of CAF were sensitive to their self-evaluation, a series of non-parametric Spearman's rank order correlations were performed. Since the sample size ($n = 7$) was quite limited due

to the constraints of the classroom-based study, the significant correlational relationships are less likely to be detected based on p values (i.e., Type II error). In the study, the potential relationships, therefore, were interpreted based on the correlation coefficients. As the classroom-based L2 study possibly includes more influences of extraneous variables than the laboratory-based study, the current study employed more restrictive criteria on the effect size, following Plonsky & Oswald (2014). Thus, the coefficient r_s values will be interpreted as small-weak (.25), medium-moderate (.40), or large-strong (.60).

Results

We initially examined the students' relative satisfaction and frustration descriptively from the questionnaire data, and then investigated which aspects of speech production should be prioritized to mitigate their frustration with L2 speaking with regard to the CAF framework.

The results of descriptive statistics on the questionnaire items are summarized in Table 2. According to the predetermined criteria (see the Data analysis), participants were satisfied with four items: (a) the comprehension of directions in a quiz (Item 11; $M = 4.93$, $SD = 1.16$, $Range = 3-6$), (b) the effective non-verbal response in a group discussion (Item 35; $M = 4.53$, $SD = 1.06$, $Range = 2-6$), (c) the comprehension of assigned readings as a preparation activity for presentations (Item 40; $M = 4.50$, $SD = 1.31$, $Range = 2-6$), and (d) the identification of main points in assigned readings for the preparation (Item 41; $M = 4.50$, $SD = 1.17$, $Range = 2-6$). All these items were related to the comprehension skills rather than the production skills. In addition, the most striking finding was that no items showed their strong frustration. However, to identify their medium level of frustration more precisely, the three most frustrated items, all about group discussion, are now discussed: (a) the lexical retrieval (Item 30; $M = 3.13$, $SD = 1.13$, $Range = 1-5$), (b) the diverse use of lexical items (Item 31; $M = 3.33$, $SD = 0.98$, $Range = 2-5$), and (c) the maintenance of natural speech rate (Item 38; $M = 3.36$, $SD = 1.01$, $Range = 2-6$). These items indicated that they were relatively frustrated with their lexical and fluency performance in spontaneous speech production required for discussion activities.

Table 2.

Descriptive statistics of the questionnaire items

Tasks	Questionnaire items	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Quiz	Q 11. I can understand the directions well.	15	4.93	1.16	3-6
	Q 12. I am satisfied with my grammar use.	15	3.93	1.39	2-6
	Q 13. I am satisfied with my vocabulary use.	15	4.07	1.03	2-6
	Q 14. I am satisfied with the organization of my answer.	15	3.67	1.18	2-6
	Q 15. I can adequately understand the key terms and concepts to be defined in a quiz.	15	4.07	1.22	2-6
Group Discussion	Q 16. I can adequately answer questions from other students.	15	3.87	1.25	2-6
	Q 17. I can express my opinion on a given question or topic.	15	3.93	1.16	2-6
	Q 18. I can make an argument with clear reasons or evidence.	15	3.80	1.15	2-5
	Q 19. I can make my argument easy to understand by giving some examples.	15	3.80	1.21	2-6
	Q 20. When I cannot understand what others say, I can ask them a question.	15	3.53	1.55	1-6
	Q 21. I can grasp whether or not my opinion is successfully understood.	15	4.00	1.31	2-6
	Q 22. I can adequately communicate my experiences and simple facts in English.	15	4.07	1.22	2-6
	Q 23. I can adequately talk about familiar topics related to my daily life.	15	3.80	1.37	2-6
	Q 24. I can adequately communicate abstract matters (e.g., hypothesis).	15	3.47	1.19	2-5
	Q 25. I can connect several sentences along with my opinion.	15	3.87	1.19	2-6
	Q 26. I can coherently tell my story even if it is long.	15	3.40	1.18	2-5
	Q 27. I can speak with an appropriate word order.	15	3.53	1.19	2-6
	Q 28. I can use complex grammar such as relative pronouns if necessary.	15	3.67	1.23	2-6
	Q 29. I don't make grammatical errors which hinder communication.	15	3.73	1.10	2-6
	Q 30. I don't usually stop speaking due to the vocabulary problems.	15	3.13	1.13	1-5
	Q 31. I can use a variety of vocabulary to express my opinion.	15	3.33	0.98	2-5
	Q 32. I can use appropriate vocabulary following my intention.	15	3.53	1.06	2-5

Group Discussion	Q 33. I can speak with intelligible pronunciation.	15	4.00	1.07	2-6
	Q 34. I can effectively use intonation to express myself.	15	3.73	1.16	2-6
	Q 35. I can respond to my peers by non-verbal cues such as nodding.	15	4.53	1.06	2-6
	Q 36. I can paraphrase peer's utterances to understand what the peer says.	15	3.60	1.12	2-5
	Q 37. I can naturally maintain the conversation with one or more peers.	15	3.53	1.19	2-6
	Q 38. I can maintain my talk without unnatural pauses.	14	3.36	1.01	2-5
Student Presentation	Q 39. Have you made your presentation in this course? (Yes or No)	-	-	-	-
	Q 40. I can understand the contents of the textbook for my presentation.	12	4.50	1.31	2-6
	Q 41. I can identify the important points in my assigned reading.	12	4.50	1.17	2-6
	Q 42. I can make Power Point slides for the student presentation well.	13	4.08	1.38	1-6
	Q 43. I am satisfied with my grammar use.	12	3.67	0.89	2-5
	Q 44. I am satisfied with my vocabulary use.	12	3.75	0.62	3-5
	Q 45. I am satisfied with my use of fixed expressions.	12	4.08	0.79	3-5

Next, we examined their CAF measure scores descriptively as an indication of oral proficiency, and then investigated their relationships with their self-evaluation scores statistically. As summarized in Table 3, their accuracy and syntactical complexity were substantively high, suggesting that they had sufficient knowledge of grammatical forms either declaratively or procedurally. On the other hand, wide standard deviation and range values show that their productivity, lexical diversity, and fluency varied greatly among the participants.

To address the second aim of our research, a series of non-parametric Spearman's rank order correlations were performed. Due to the limited number of the sample size (see the Data Analyses), this study focused on the values of correlation coefficients as an indication for the potential relationships between self-evaluation and speaking performance, following Plonsky & Oswald (2014). Although we could not draw a definitive conclusion due to the methodological constrains, two CAF measures indicated a medium effect size on the relationships with their self-evaluation scores: the total number of words ($r_s = -.54, p = .215$) and words per minute (WPM) ($r_s = .54,$

Table 3.

Descriptive Results of CAF Measures and the Correlations with Self-evaluation scores

CAF domain	Measure	<i>M</i>	<i>SD</i>	<i>Range</i>	Self-evaluation	
					<i>r_s</i>	<i>p</i>
Productivity	Total # of words	183.7	39.9	131-247	-.536	.215
Syntactic Complexity	Clauses/AS-unit	1.92	0.55	1.39-2.71	-.107	.819
Lexical Complexity	MTLD	52.74	9.06	37.37-64.75	0.000	1.000
Accuracy	WCR	0.894	0.052	0.821-0.964	-.179	.702
Fluency	WPM	72.6	28.6	34.57-119.25	.536	.215

$p = .215$). These results demonstrated that the less redundantly or the more fluently they could produce their speech, the less frustrated they were likely to feel in their speech production.

Discussion

Our first research question addressed which aspects of tasks in EMI students were satisfied and frustrated with. The most prominent finding from the descriptive analyses was that, according to our predetermined criteria ($M < 2.5$ on the self-evaluation scale), no questionnaire items showed any frustrating aspects of EMI. A possible explanation for this may lie in the timing of data collection. The questionnaire data were collected after nine lessons in total, suggesting that they had already been more or less confident in their accomplishment of in-class activities. As their successful experiences and behaviors were accumulated, they did not end up with unreasonably low self-evaluation (Bandura, 1997).

On the other hand, the descriptive results suggested that whereas they were satisfied with their comprehension skills in the EMI, they were relatively frustrated with spontaneous speech production. Extensive research in L2 psycholinguistics has claimed that while comprehension can be processed by either declarative or procedural knowledge due to the relatively adequate time available, speech production is largely dependent on procedural knowledge due to its spontaneity (e.g., Ellis & Barkhuizen, 2005). In addition to this theoretical consensus, group discussion in the EMI requires them to comprehend peers' utterances and conceptualize their own utterances simultaneously. In other words, less attentional capacity for their own speech production is available in production, so that they are forced to exclusively rely on

their procedural knowledge of L2 (Skehan, 2014). Thus, we could assume that they have less procedural knowledge available than declarative one, and consequently that they passively perceive more frustration with spontaneous speech production compared to comprehension.

The second aim of the study is to examine which aspects of speech production are sensitive to their self-evaluation, which is pedagogically relevant to language support in the EMI course. According to the preliminary analyses of speech, excessively high scores on accuracy and syntactic complexity indicated that students had attained sufficient declarative or procedural knowledge of grammatical forms. Moreover, we observed the variations among students in productivity, lexical complexity, and fluency. From these two findings, we could draw a potential conclusion that whereas they obtain quantitatively sufficient linguistic knowledge required for academic English speaking, students vary in terms of the efficiency of speech processing (i.e., procedural knowledge). This interpretation can be triangulated by the findings from our first research question.

According to a set of correlational analyses, two moderate potential correlational relationships were detected. First, the self-evaluation score was in a negative relation with the productivity measure. In other words, the less redundantly students could produce their speech, the more satisfied they would be with speaking performance. Therefore, their perceptions may have resulted from the efficiency of task accomplishment. From the perspective of learner characteristics, the students were intermediate-level English learners, so that they sometimes needed to elaborate or paraphrase some ideas which they could not express directly. They may have perceived their elaboration and paraphrasing uncomfortable due to their limited linguistic repertoire, even though such compensation strategies should be valued as strategic competence. Furthermore, as they were also university students, they knew much sophisticated and infrequent vocabulary in L1. Thus, they were more likely to notice the gap between their L1 and L2 lexical repertoire, resulting in their lowered self-evaluation in L2 performance. In sum, the negative relationship between productivity and self-evaluation indicates that students in the EMI seek an efficient way to express themselves, and as well-documented in SLA, that they notice the gap between their existing linguistic knowledge and the required one (i.e., noticing-the-gap), which potentially facilitates successful L2 acquisition in the communicative

contexts (e.g., Ellis & Shintani, 2013).

Second, the self-evaluation score was also in a positive relation with the fluency measure, suggesting that the more smoothly they produced their speech, the more satisfied they were with their performance. The theoretical consensus on L2 fluency is that fluency is a multidimensional construct by its nature (for a comprehensive review see Bosker, Pinget, Quené, Sanders, & de Jong, 2012). Fluency has three major subdomains: fluidity of speech (i.e., speed fluency), hesitation and pausing (i.e., breakdown fluency), and repetition and self-correction (i.e., repair fluency). Although it is traditionally regarded as the measure operationalized for speed fluency (e.g., Ellis & Barkhuizen, 2005), WPM is also indicative of the speaker's length of pauses (i.e., breakdown fluency) due to its calculation procedure. Accordingly, we could assume that WPM captures the speaker's speed and breakdown fluency. Studies on the perceptual sensitivity of fluency measures reveal that the sensitivity of both speed and breakdown fluency to perceptual fluency is empirically proved (see Bosker et al., 2012). Therefore, the positive relationship between self-evaluation and fluency has been found to be aligned with previous studies on L2 fluency.

In addition to the issues around operationalization, the finding could be also explained in terms of L2 development. According to several CAF studies, the internalization and modification of linguistic forms (i.e., complexity and accuracy) are followed by the consolidation of them with fluency (see Housen, Kuiken, and Veddar, 2012). As mentioned above, students were well equipped with declarative knowledge about linguistic forms, so that they may have been in the current process of L2 fluency development. Thus, the variation in fluency measures was observed due to the individual differences in the rate and degree of consolidation.

Pedagogical Implications

The findings have a couple of implications for EMI in universities in EFL settings. First, the questionnaire data suggests that cumulative experiences with using English in academic contexts have a significant impact on students' self-evaluation. Therefore, the curriculum should abundantly offer preparatory courses such as EAP and CBI courses with the primary focus on language development before they take EMI courses. The contents in EAP courses could be cognitively demanding due to its academic nature, resulting in quite limited attentional capacity for language

processing. In order for L2 acquisition to take place in meaning-oriented contexts, students should pay attention to form aspects of language (i.e., FonF). The prerequisite for FonF is the condition where the contents of the EAP course (i.e., meaning) are clearly understandable. Thus, the curriculum should provide a variety of EAP or CBI courses in terms of the topic difficulty and the relative weight of focus on content and language, so that students can select appropriate courses according to their proficiency levels. The appropriate correspondence of the courses to their English proficiency enables them to process the language successfully.

Second, both the questionnaire and speech data indicate that whereas they have much declarative knowledge about the linguistic forms, students have insufficient procedural knowledge due to their limited use of English in Japan. Thus, EMI instructors in EFL settings should create the pedagogical situation where students can use specific linguistic forms repetitively with the primary focus on meaning. In this sense, academic contents in EMI and CBI easily offer such a kind of situations. For instance, as in-class tasks in EMI are often based on assigned readings before the class, students can process certain forms with multiple modalities such as listening through the lecture and speaking in group discussion, consistently focusing on the same topic. The lesson structure possibly allows students to process specific linguistic items repetitively on different occasions, promoting the proceduralization of the linguistic knowledge (Sato & Lyster, 2012; Suzuki, 2016).

Last, from the results of correlational analyses, we could assume that students seek to find an efficient way to express detailed meaning they want to convey. To address this issue, the EMI instructors can provide two different types of vocabulary lists as supplementary materials. They could make one list of content-specific vocabulary to convey the sophisticated meaning and to support their lexical diversity, while they could have available to EMI students the other list of frequent formulaic sequences in academic contexts to secure temporal and cognitive capacity for elaborated and creative language expressions (Skehan, 2014).

Conclusion

Two broad conclusions can be drawn from the findings of our study. Initially, according to the questionnaire data, the students in the EMI course tend to be frustrated with spontaneous speech production rather than reading and listening

comprehension. Spontaneous speech production requires them to express themselves in response to their interlocutors, resulting in the situation where they must rely only on procedural knowledge. In contrast, comprehension can be processed by either declarative or procedural knowledge. Therefore, students could make the use of what they have learned through college entrance examinations and English major content courses. Second, the productivity and fluency aspects of L2 speech potentially influence speakers' own self-evaluation of spontaneous speech (i.e., group discussion in EMI), as the effect sizes of correlations indicated. The concise speech seems to be valued in academic discourses compared to the redundant one. Students could also be sensitive to their fluctuation in the efficiency of language processing due to the lack of proceduralization of linguistic knowledge. Thus, students might consider faster speech rate the desired trait of L2 speaking in EMI.

Whereas our study has offered a few potential insights into EMI courses in EFL settings, the findings should be interpreted cautiously with regard to several methodological limitations. First, the time for the administration of the questionnaire varied among our participants to avoid disrupting the regular class. Thus, some participants who submitted the questionnaire earlier might have lower self-evaluation scores, because they had fewer experiences of English use in the EMI class than the others. Second, the speaking data were obtained outside the class in order to both elicit their upper limitation of performance and control for extraneous variables in the classroom. Their actual English use in the EMI, however, may have more transparent information on students' actual frustration in the classroom. Third, the study focused on one single EMI course and its instructor, so that the findings may have been influenced by the instructor's personal traits. Finally, as the sample size was quite limited, we could only refer to the potential relationships between their self-evaluation and CAF measure scores from the effect sizes of correlational analyses. Therefore, further research is called for to capture more detailed nature of EMI and its effects on students' self-evaluation and English learning with a larger number of participants with the collaboration with multiple EMI instructors.

References

- American Council on the Teaching of Foreign Languages (ACTFL). (2012) *ACTFL proficiency guidelines*. Washington, DC: Author. Retrieved from https://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
- Bosker, H. R., Pinget, A.-F., Quene, H., Sanders, T., & de Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 7(30), 159–175.
- Christison, M. A., & Murray, D. E. (2014). *What English language teachers need to know Volume III: Designing curriculum*. New York, NY: Routledge.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Ellis, R., & Shintani, N. (2013). *Exploring language pedagogy through second language acquisition research*. New York: Routledge.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring Spoken Language: A Unit for All Reasons. *Applied Linguistics*, 21(3), 354–375.
- Foster, P., & Wigglesworth, G. (2016). Capturing Accuracy in Second Language Performance: The Case for a Weighted Clause Ratio. *Annual Review of Applied Linguistics*, 36, 98–116.
- Harada, T. (in press). Developing a content-based English as a foreign language program: Needs analysis and curriculum design at the university level. In M.A. Snow & D.M. Brinton. (Ed.) *The content-based classroom*. Ann Arbor, MI: University of Michigan Press.
- Hellekjær, G. O. (2010). Lecture comprehension in English-medium higher education. *Hermes - Journal of Language and Communication Studies*. 45, 11–34.
- Housen, A., Kuiken, V., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency – Investigating complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins.
- Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task-based

- L2 research: toward more developmentally based measures of second language acquisition. *Applied Linguistics*, 35(5), 607–614.
- Lightbown, P. (2014). *Focus on content-based language teaching*. Oxford, England: Oxford University Press.
- Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15–40). Cambridge, UK: Cambridge University Press.
- Mak, B. (2011). An exploration of speaking-in-class anxiety with Chinese ESL learners. *System*, 39 (2), 202–214.
- McCarthy, P. M., & Jarvis, S. (2010). MTL, D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42 (2), 381–92.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555–578.
- Plonksy, L., & Oswald, F. L. (2014). How Big Is “Big”? Interpreting Effect Sizes in L2 Research. *Language Learning*, 64(4), 878–912.
- Sato, M., & Lyster, R. (2012). Peer Interaction and Corrective Feedback for Accuracy and Fluency Development. *Studies in Second Language Acquisition*, 34, 591–626.
- Skehan, P. (2014). *Processing perspectives on task performance*. Amsterdam, the Netherlands: John Benjamins.
- Smit, U., & Dafouz, E. (2012). Integrating content and language in higher education: An introduction to English-medium policies, conceptual issues and research practices across Europe. *AILA Review*, 25, 1–12.
- Suzuki, S. (2016, June). *Examining the pedagogical potential of English-medium instruction for second language development in speaking: A longitudinal study*. Paper presented at Task-Based Language Teaching in Asia (TBLT in Asia), Kyoto, Japan.
- Takeuchi, O. & Mizumoto, A. (2014). *The handbook of research in foreign language learning and teaching (Revised version)*. Tokyo: Shohakusha.
- van de Pol, J., Volman, M., & Beishuizen, J. (2010). *Scaffolding in teacher-student interaction: A decade of research*. *Educational Psychology Review*, 22(3), 271–296.