A Replication of Okada, Sawaumi, and Ito (2017): Effects of Viewing Speaker Videos by Proficiency Order on Japanese EFL Learners' Speaking Skills

Yasuko Okada
(okada92025@gmail.com)
Saitama Women's Junior College, Japan

Takafumi Sawaumi
(t.sawaumi@gmail.com)
Ryutsu Keizai University, Japan

Takehiko Ito
(take@wako.ac.jp)
Wako University, Japan

Abstract

Video-recording has been widely used in foreign language education to study learners' performance and thereby improve their language skills. Recent studies show that viewing not only their own recordings but also those of their peers enhances students' motivation for language learning. This study attempted to replicate Okada, Sawaumi, and Ito (2017), who examined how observing non-native speaker model videos by proficiency order affect students' public speaking skills. Participants were 31 Japanese university students enrolled in English communication courses. One group ($n = 19$) was shown more-proficient speaker videos first and less-proficient ones next, whereas the other ($n = 12$) was shown videos in the opposite order. Results indicated no significant effect on self-evaluation scores in either group. However, the third presentation was peer-rated significantly higher than the second for those who observed more-proficient models first and less-proficient ones next; these findings are not consistent with those of Okada et al. (2017), where students' third presentation was rated significantly higher when they observed less-proficient models first and more-proficient ones next. This internal replication study presents the accumulated results.

1 Introduction

The increasing emphasis on encouraging foreign language learners to improve their communication techniques has led to the growth of interest in how to teach these skills in English as a foreign language (EFL) classrooms in Japan, based on the proposal of the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) that Japanese students should be able to use the English language to thrive in a globalizing society (2014). However, a number of language learners experience general difficulty speaking in the target language in front of other students (Kawauchi, 2016; Young, 1990). This is likely related to the nature of the Japanese culture, in that Japan has often been considered a collectivist society in which people emphasize group goals and avoid showing their uniqueness in the group setting (Doi, 1971; Hofstede, 2011). Thus, Japanese learners of

foreign languages may become anxious about the risk of violating group goals when delivering oral presentations, or speeches, in front of other students in the classroom.

People in the United States, on the other hand, live in an essentially individualist culture, in which individual uniqueness is stressed over group goals (Hofstede, 2011). To succeed in such a society, American universities have widely offered courses or training to develop learners' public speaking skills and reduce speaking anxiety (Hancock, Stone, Brundage, & Zeigler, 2010). Japanese students also need training to improve their speech skills in English, for similar reasons; a recent survey conducted by MEXT (2015) reported that students lacked experience delivering speeches or presentations in English in the classroom. Therefore, to develop Japanese students' public speaking skills, it seems necessary or advantageous to offer public speaking courses at Japanese universities similar to those offered in the United States.

The use of video-recordings can play an active role in developing students' public speaking skills. Learners consulting videos can compare their progress to that of model speakers based on visual and audio information and assess both their language proficiency and public speaking skills thereby. Although these abilities and skills are developed through observing the performance of others (Bandura, 1971, 1977), there is evidence that we can substitute video models for live models in the language classroom with little or no loss of learning affordances. Okada et al. (2017) found that viewing non-native speaker model videos helps Japanese EFL learners compare their performance with that of their peers, and thereby helps develop their public speaking skills. Although that study provided experimental evidence that the viewing order of different model videos by proficiency may alter the effect on viewers' speaking performances, the generalizability of that evidence is unclear, as it is based on only one study, albeit a sound one. Thus, it seems necessary to replicate the study of Okada et al. (2017) in order to present solid evidence that EFL learners' language skills and public speaking skills can benefit from video-based observational learning and that the video observation order affects their public speaking skills. It should be noted that in language learning, "video recordings" usually refers to self-videos of students' performances; in this study, however, the videos are of non-native speaker models, while the students' own videos were also employed for self- and peer evaluation, but are not the focus of the present study.

2 Literature review

2.1 *Overview of Okada et al. (2017)*

Okada et al. (2014) applied the concept of observational learning, in which Bandura (1971, 1977) claims that children's behavior and attitude are learned through observing and modeling others' behaviors, to language learners and examined whether EFL learners were able to develop their speaking skills, in particular, using model videos. In the study, two groups of learners of different language proficiency levels were shown videos of more-proficient model speakers with the aim of improving their public speaking skills. Findings indicated that the low-proficiency learners became discouraged after viewing the models, whereas the high-proficiency learners become motivated by the observations, suggesting that learners should be provided with instruction that takes account of their language proficiency. Okada et al. argued that if learners are shown videos with appropriate models, based on careful consideration of the learners' level of proficiency, it may help develop the learners' language proficiency. Based on Okada et al.'s study, further research was conducted to examine the effects of video-based observational learning with more- and less-proficient speaker models (Okada et al., 2017). Observational learning methods have often been applied to develop students' oral presentation skills in the higher educational setting (De Grez, Valcke, & Roozen, 2014). It was expected that observing model videos, in lieu of live models, would help EFL learners enhance their public speaking skills

The qualitative findings from Okada et al. (2017) indicated that more- and less-proficient models respectively helped EFL learners improve their public speaking skills for different purposes, although difference in language proficiency was not the focus of the study. In another study, Adams

(2004), speculated that the use of peer models may lead to a social comparison effect and that "observers may [conclude] that acceptable standards of speech production and delivery [are] possible to achieve without native speaker ability" (p. 126). This idea was borne out by her experimental results using an expert model and a non-expert peer model with a group of graduate student learners who were at the upper-intermediate level of English and who had attained an English proficiency score of 6 on the International English Language Testing System (IELTS) scale. Therefore, it is probable that the viewing order of different model videos affects learners' public speaking to some degree.

Similarly, quantitative findings from Okada 1 et al.'s study (2017) indicate that learners who observed less-proficient models first and more-proficient ones next scored significantly higher than in the reverse order. Other studies also demonstrate that the sequence order could affect student performance in a foreign language (Chen, 2010; Shimizu, 2007; Winke, Gass, & Sydorenko, 2010). Nevertheless, it remains necessary to determine whether the findings from these studies are generalizable by conducting replications with other learners and in other contexts. In a subsequent study, Okada et al. (2018) replicated Okada et al. (2017) with different participants but the same procedure, gaining findings that were consistent quantitatively and qualitatively. Given that these studies (Okada et al., 2017, 2018) used regular classes and that participants were engaged in other language-related tasks besides model video observation, there is, however, a possibility that other factors, for example, the study being conducted in participants' regular classes, affected their public speaking skills.

2.2 Overview of the present study

The current study aimed to replicate Okada et al.'s (2017) findings with different Japanese EFL learners, using the same data collection procedure. As noted earlier, the prior studies (Okada et al., 2017, 2018) did not include enough participants for statistical reliability, due to the classroom-oriented nature of the research, which limited the student numbers to those in the site classes. Thus, close replication with more participants was viewed as desirable to ensure generalizability of the findings to the broader Japanese population.

The research questions for this replication study are twofold, as in Okada et al. (2018). First, we examined whether viewing non-native speaker model videos helped EFL learners enhance their speaking skills in the target language. Second, we investigated whether the model viewing sequence affected students' own subsequent presentations. It was expected that the results would replicate Okada et al.'s (2017, 2018) findings that model videos would affect learners' linguistic proficiency and public speaking skills positively across different modes of speaker proficiency levels and that learners would improve their public speaking skills more when viewing less-proficient model videos first and more-proficient ones next.

3 Method

3.1 Participants

The participants were 31 Japanese students in their first year of university, majoring in economics, who were enrolled in an English communication course in either Spring or Fall 2016. They had been studying English for at least six years, in high school and junior high school. At the beginning of the course, a total of 38 students (19 for each group) were asked to give permission for their recorded presentations and written data, such as self- and peer evaluations and reflection papers, to be used for future research or educational purposes. Permission was not obtained from seven students in group B,[1] and so data for this study were ultimately collected from 31 students (19 and 12 in Groups A and B, respectively). Students were informed that their participation would not affect their grade and that the study would not require any additional work or time.

Students were placed in sections according to their Test of English for International Communication (TOEIC) Bridge (ETS, 2017) scores; average scores for groups A and B were 101 and 108, respectively.[2] These scores fall into the A1 level on the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2017). To find out if the difference in average score between the groups was meaningful, a t-test was conducted; it indicated no statistically significant difference, $t(27.98) = 1.74$. $p = .09$. Accordingly, we considered that the students' English proficiency for two groups was equivalent.

The participants were taking the course due to their graduation requirement. The course met twice weekly over 14 weeks, each time for 90 minutes. For the student grades, 30% were assigned to oral presentations (graded by the instructor) and 15% to draft submission and participation in self- and peer evaluations (but not the scores given in the evaluations); the remaining 55% were based on other assignments.

3.2   Design and data collection procedure

This classroom research study was an approximate internal replication study, which repeated most parts of the original study by the original researcher (see Porte, 2012). In addition, using regular classes, a quasi-experimental, revised nonequivalent-groups pretest/posttest design was adopted to measure the effects of model video viewing on speaking performance in the EFL classroom. That is, this study examined the effects of showing video recordings (i.e. intervention) on two groups of learners who were not randomly assigned (i.e. non-equivalent groups), using self- and peer evaluation scores (i.e. pretest/posttest). It should be noted that there is a factor that may limit the consistency and rationality of this kind of replication study: the data from the two groups in the current study were collected at different times, while prior studies (Okada et al., 2017, 2018) gathered data from two groups concurrently.

The participants delivered three oral presentations during the course, because it was necessary to examine the effects of model video viewing (i.e. without watching the videos, after watching more- and less-proficient model videos, respectively). For each speaking performance, students spoke aloud from a prepared written script. The first, second, and third topics were "My favorite food," "My memorable event," and "My ideal trip," respectively.[3] Murphy (1992) points out that "[o]ne widely implemented procedure in the teaching of English as a second language (ESL) speech communication requires each student to take a turn standing at the front of the room as she or he delivers a preplanned oral presentation to an entire class" (p. 58). Though it must still be stressful for students to recite prepared material to the whole class, this memorization may be able to serve as a tool to internalize expressions they were taught and use these more and more naturally (Oanh & Hien, 2006). Thus, students used a three-paragraph written script (180–200 words) for the oral presentation, after writing it during a regular lesson and receiving feedback from the instructor.

As noted earlier, data from the two groups were collected at different times, but they were taught with a similar approach. There was a modest difference in session numbers, and thus the data collection procedure for the earlier semester will be explained in this section. In the performance phases of Sessions 7, 15, and 23, students' performances were video-recorded and evaluated by the instructor. In addition, immediately after delivering speeches in Sessions 15 and 23, participants were instructed to comment on their First and Second Reflections.

Self-/peer evaluations were conducted using a form in Sessions 8, 16, and 24, while participants were watching the recorded performances. Students evaluated their classmates and returned the completed forms to the respective speakers; the evaluation forms were collected by the instructor shortly thereafter for future analysis. In addition to evaluating the third presentation in Session 24, students were shown 30-second clips from each of their initial performances to help them consider how those performances might have evolved.

In Sessions 14 and 22, students were given opportunities to watch model videos and to identify speakers' weaknesses and strengths. In Session 14, Groups A and B viewed the more-proficient and

less-proficient models, in that order; the order was reversed in Session 22. After viewing and commenting on two or three clips, students were divided into groups of three or four, wherein they discussed the models for approximately 2–3 minutes.

Finally, in the optional interview phase, six students from one group and three from the other were recruited to answer questions regarding their experiences viewing model videos and learning the English language. The interviews were conducted outside of class time; so students were given a nominal amount of money in return for the participation in the interview. The interview was a dialogue between the student and the instructor in Japanese. Each interview took approximately 15–20 minutes and was conducted over the lunch break. The interviews were video-recorded and also recorded with an IC recorder. Total recording time was approximately 160 minutes for nine students. The flowchart of the data collection procedure is shown in Figure 1.
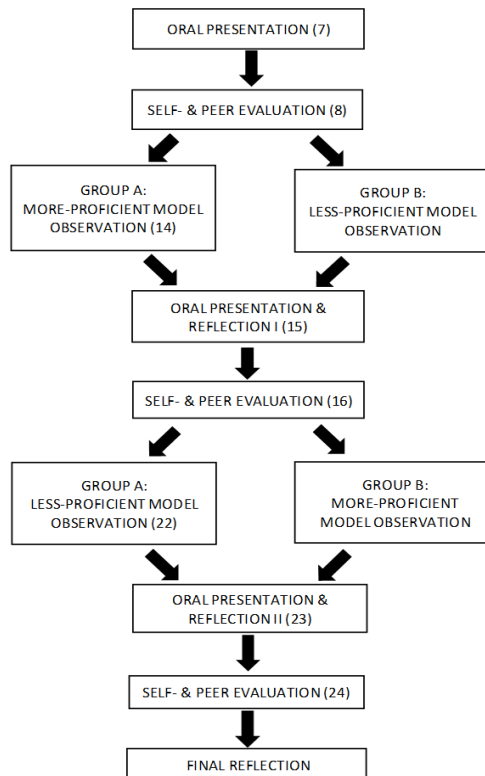
Fig. 1. The data collection procedure, adapted from Okada et al. (2017). (Numbers in parentheses are sessions in the course where the data were collected. Session numbers for model observation for Group B were not included because they varied slightly from those of Group A)

3.3   Materials

This study was conducted with the same procedure as in Okada et al. (2017, 2018); therefore, research materials (i.e. three reflection papers) were identical to those in Okada et al. (2017, 2018). All instructions in the evaluation and reflection papers were provided in Japanese.

3.3.1   Model video[4]

Model videos produced by Japanese EFL students from preceding years were selected by the first author. Of these models, eight more-proficient ones were selected, primarily from students in

a higher-level English class. These models demonstrated outstanding voice control and body language in their presentations, and therefore, were considered the best ones to enable other students to reach their performance targets. Less-proficient model videos were taken from lower-level English classes. These models' speech in their presentations was not adequate to assess their performance closely, but was enough to distinguish these models from high-proficiency ones. Each video lasted approximately 2–3 minutes.

3.3.2 Self-/peer evaluation

Self- and peer evaluations were used to examine how student evaluation scores had changed over the three oral presentations. To evaluate public speaking skills, the rating scale developed by Yamashiro and Johnson (1997) was used, with some modifications. The 14 elements adopted were: (1) projection, (2) pace, (3) intonation, (4) diction, (5) posture, (6) eye contact, (7) gesture, (8) introduction, (9) body, (10) conclusion, (11) topic choice, (12) language use, (13) vocabulary, and (14) purpose. These were grouped into four categories: (1) to (4) voice control, (5) to (7) body language, (8) to (10) content, and (11) to (14) effectiveness. This study adopted all four elements in voice control, two from body language (posture and eye contact), and all three in effectiveness from Yamashiro and Johnson's (1997) study, and added two original elements in body language: voice control (projection, pace, intonation, and diction), body language (posture, foot and hand position, eye contact, and facial expression), and effectiveness (topic choice, language use, and vocabulary). These elements were ranked on a 4-point Likert-type scale from 1 (weak) to 4 (great), as in Okada et al. (2017, 2018). The present study did not include the content elements, because participants had already submitted their drafts to the instructor and received feedback on these elements.

3.3.3 Reflection papers

First/Second Reflection: Participants were asked about how model videos had affected their speaking performance immediately after the second and third rounds of presentations.

Final Reflection: Participants were instructed to give information on their perceptions of the two different types of model videos. The form included three sections: In Section One, students responded regarding the difficulty they had encountered while preparing scripts for each topic, on a 5-point Likert-type scale from 1 (easy) to 5 (difficult). In Section Two, students rated the interventions' overall usefulness on a 5-point scale Likert-type from 1 (not useful at all) to 5 (very useful), to measure the effect of viewing model videos and students' recordings. In Section Three, students were asked to freely write down their thoughts regarding the two types of model videos.

3.3.4 Post interview

With several interview questions prepared in advance, a semi-structured interview was carried out with the participants. The following questions were asked:

- Did you feel reluctant to study English? If so, why?
- Did you think the instructions, such as using pronunciation symbols and word stress, were helpful in conducting performances? Can you explain why?
- Were you confident in delivering your performance? How was each of your performances?
- Did you enjoy watching the three kinds of video-recordings (model performances, peer performances, and respondents' own performances)? How were they helpful?
- Was the model video sequence (more-proficient video first and less-proficient video next) appropriate? How did each type of video-recording motivate you to deliver your performance?

### 3.4	Mixed-method analysis of data

For the quantitative analysis, self- and peer evaluated scores were analyzed to find out how students' presentations were assessed, using SPSS v. 22. Repeated-measures analysis of variance (ANOVA) was used to examine the effects of group (A vs. B [between-participants factor]) and presentation round (first vs. second vs. third [repeated factor]) on the self-/peer evaluated scores, as well as that of their interaction. After confirming high internal consistency, ratings for the 11 evaluation items were averaged as follows: voice control (Items 1–4), body language (Items 5–8), effectiveness (Items 9–11). The Cronbach's alpha coefficients for the three subscales were calculated separately for each presentation round (.86, .84, .86 [first]; .90, .81, .93 [second]; .93, .90, .97 [third]) and for peer rating sessions (.82, .66, .85 [first]; .78, .55, .80 [second]; .88, .81, .93 [third]). The Greenhouse–Geisser correction was used where the sphericity assumption was not met for the repeated factor.

## 4	Results

### 4.1	Quantitative results

First, the mean scores of self-rated voice control, body language, and effectiveness are summarized in Table 1. A two-way ANOVA (group × round) for self-rated voice control revealed a significant main effect for round, $F(2, 50) = 5.64$, p < .01, $\eta_p^2 = .18$. A Bonferroni post hoc test for the main effect showed that voice control self-evaluation for the third round of presentations was significantly higher than that for the first round (p < .01). No other effects were significant: group, $F(1, 25) = 0.47$, $p = .50$, $\eta_p^2 = .02$, and group × round, $F(2, 50) = 0.58$, p = .56, $\eta_p^2 = .02$.

Table 1. Mean Scores and SDs for the Self Evaluations

	Round 1			Round 2			Round 3		
	N	M	SD	N	M	SD	N	M	SD
Group A									
Voice	17	2.49	0.80	18	2.69	0.82	19	2.96	0.95
Body	17	2.41	0.72	18	2.79	0.73	19	2.87	0.81
Effect	17	2.76	0.96	18	3.00	0.81	18	3.09	0.93
Group B									
Voice	12	2.60	0.76	12	2.65	0.82	10	3.28	0.48
Body	12	2.69	0.87	12	2.79	0.67	10	3.43	0.51
Effect	12	2.61	0.72	12	2.94	0.83	10	3.37	0.48

Note. Data from absent students were discarded at each analysis stage. Voice = voice control, Body = body language, Effect = effectiveness.

The two-way ANOVA for self-rated body language also showed a significant main effect for round, $F(2, 50) = 6.03$, $p < .01$, $\eta_p^2 = .19$. A Bonferroni post hoc test for the main effect showed that body language self-evaluation for the third round of presentations was significantly higher than that for the first round ($p < .01$). No other effects were significant: group, $F(1, 25) = 2.54$, $p = .12$, $\eta_p^2 = .09$, and group × round, $F(2, 50) = 0.99$, p = .38, $\eta_p^2 = .04$.

The two-way ANOVA for self-rated effectiveness also found a significant main effect for round, $F(2, 50) = 5.95$, $p < .01$, $\eta_p^2 = .19$. A Bonferroni post hoc test for the main effect showed that self-evaluation effectiveness for the third round of presentations was significantly higher than that for the first round (p < .01). No other effects were significant: group, $F(1, 25) = 0.04$, $p = .85$, $\eta_p^2 < .01$, and group × round, $F(2, 50) = 0.55$, $p = .58$, $\eta_p^2 = .02$.

Taken together, the findings from self-rated voice control, body language, and effectiveness were consistent with those of Okada et al. (2017, 2018).

Table 2. Mean Scores and SDs for the Peer Evaluations

	Round 1			Round 2			Round 3		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Group A									
Voice	17	3.49	0.24	18	3.56	0.22	19	3.75	0.21
Body	17	3.38	0.17	18	3.47	0.14	19	3.67	0.16
Effect	17	3.56	0.13	18	3.67	0.09	19	3.81	0.11
Group B									
Voice	12	3.55	0.18	12	3.54	0.17	11	3.61	0.19
Body	12	3.62	0.09	12	3.51	0.12	11	3.67	0.11
Effect	12	3.75	0.08	12	3.61	0.11	11	3.73	0.09

Note. Data from absent students were discarded at each analysis stage. Voice = voice control, Body = body language, Effect = effectiveness.

The mean scores for peer-rated voice control, body language, and effectiveness are shown in Table 2. A two-way ANOVA (group × round) for peer-rated voice control revealed a significant main effect for round, $F(2, 52) = 10.22$, $p < .01$, $\eta_p^2 = .28$, but not for group, $F(1, 26) = 0.25$, $p = .62$, $\eta_p^2 = .01$. This main effect of round was qualified by a significant group × round interaction effect, $F(2, 52) = 4.06$, $p = .02$, $\eta_p^2 = .13$. Post hoc tests showed that Group A's peer-evaluated voice control for the third round of presentations was significantly higher than for the first and the second rounds (both $ps < .01$), but that this was not the case with Group B (see Fig. 2). These findings are not consistent with those of Okada et al. (2017, 2018),[5] since those studies showed that Group B's voice control peer evaluation for the third presentation was significantly higher than for the first and second rounds.
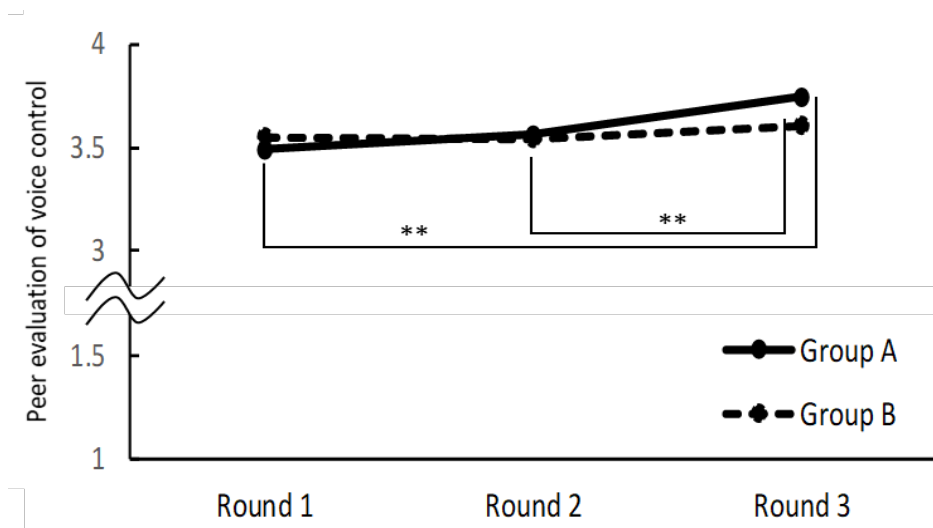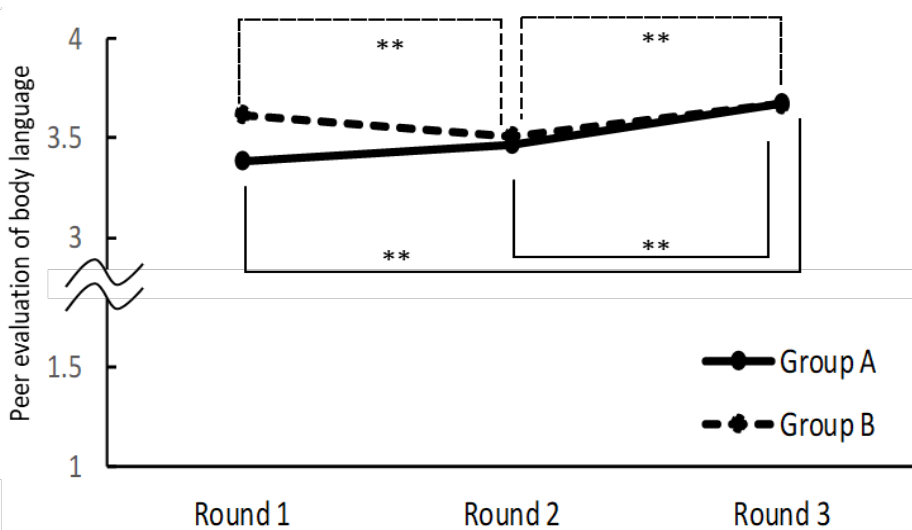

Fig. 2. Peer evaluation of voice control as a function of group and round (** p < .01)

A two-way ANOVA (group × round) for peer-rated body language showed a significant main effect for round, $F(2, 52) = 28.01$, $p < .01$, $\eta_p^2 = .52$, but not for group, $F(1, 26) = 3.20$, $p = .09$, η_p^2

= .11. This main effect of round was qualified by a significant group × round interaction effect, $F(2, 52) = 10.28$, $p < .01$, $\eta_p^2 = .28$. Post hoc tests showed that Group A's body language peer evaluation for the third round of presentations was significantly higher than for the first and the second rounds (both ps < .01), whereas Group B' body language peer evaluation for the second round was significantly lower than for the first and third rounds (both $ps < .01$; see Fig. 3). These findings are not consistent with those of the earlier studies (Okada et al., 2017, 2018). For example, in Okada et al. (2017), Group B's second and third peer evaluations were significantly higher than the first. Similarly, in Okada et al. (2018), Group B's third peer evaluation was significantly higher than the first and third rounds, and its second round was significantly higher than the first round.

Fig. 3. Peer evaluation of body language as a function of group and round ($p < .01$)**

A two-way ANOVA (group × round) for peer-rated effectiveness found a significant main effect for round, $F(1.56, 40.44) = 32.14$, $p < .01$, $\eta_p^2 = .55$, but not for group, $F(1, 26) = 0.13$, $p = .72$, $\eta_p^2 < .01$. The main effect of round was qualified by a significant group × round interaction effect, $F(1.56, 40.44) = 36.49$, $p < .01$, $\eta_p^2 = .58$. Post hoc tests showed that Group A's effectiveness peer evaluation for the third round of presentations was significantly higher than for the first and the second rounds, while the one for the second round was significantly higher for than the first round (all $ps < .01$), whereas Group B's effectiveness peer evaluation for the second round was significantly lower than those for the first and third rounds (both $ps < .01$, see Fig. 4). These findings were not consistent with those of Okada et al. (2017, 2018), where it was shown that Group B's third peer evaluation was significantly higher than the first and second ones.
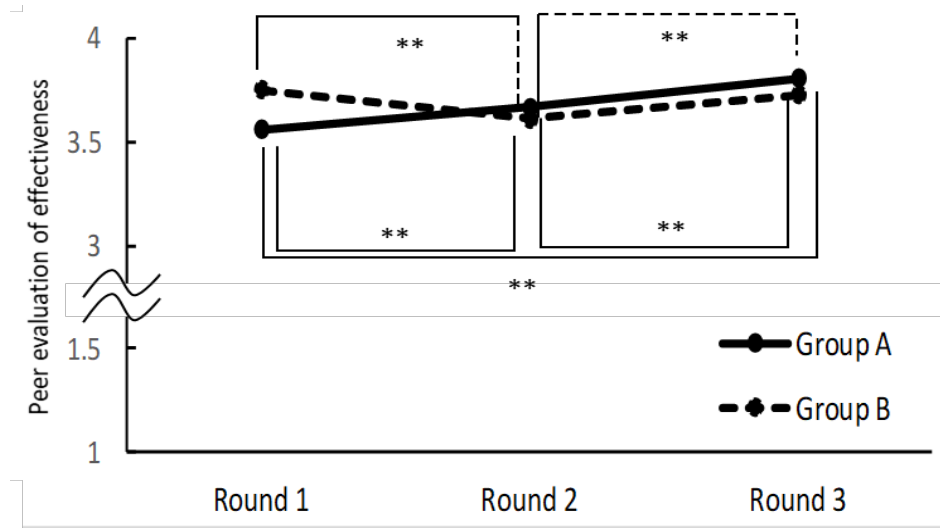
Fig. 4. Peer evaluation of effectiveness as a function of group and round (*p* < .01)**

4.2 Textual analysis

Reflection papers were analyzed using Text Mining Studio 5.1, by NTT DATA Mathematical Systems Inc. Text mining techniques can help to classify data automatically and enable users to "locate relevant information within the retrieved documents in a more timely and efficient manner" (Ananiadou et al., 2010, p. 3830) so as to analyze unstructured textual information, such as open-ended responses. The interview data can further support the findings from textual and text mining analyses and address the research questions of this mixed methods research study.

Table 3. High-frequency words in the First Reflection

	Word	Meaning	Total (A + B)
1	スピーチ	*speech*	37
2	見る	*to watch*	34
3	思う	*to think*	25
4	人	*person*	19
5	意識	*consciousness*	18
6	話す	*to speak*	17
7	サンプルビデオ	*model video*	16
8	良い	*good*	15
9	できる	*to be capable*	14
10	声	*voice*	13
	発音	*pronunciation*	13

Table 4. High-frequency words in the Second Reflection

	Word	Meaning	Total (A + B)
1	思う	*to think*	34
2	見る	*to watch*	31
3	スピーチ	*speech*	29
4	良い	*good*	21
5	サンプルビデオ	*model video*	19
6	人	*person*	18
7	ビデオ	*video*	17
8	発音	*pronunciation*	14
9	意識	*consciousness*	13

The results of the word frequency analysis demonstrated that the most frequently used words in the First Reflection were, in order, "speech" (37 tokens), "to watch" (34) and "to think" (25), and in the Second Reflection, "to think" (34), "to watch" (31) and "speech" (29) (see Tables 3 and 4). These words appeared in various contexts in the reflections. The words "to think" and "to watch" were also used frequently in Okada et al.'s study (2018), consistent with the present finding, and "to watch" was also used frequently in their study (2017).

By using "*to think*," Japanese students may have avoided making statements too assertively, as they expect to soften their thoughts and opinions with this expression (Okada et al., 2018). Almost all the students in Group A stated that model videos affected their presentations positively for different purposes, as evidenced in these excerpts:

> After watching the model videos, I practiced my speech just like the model speakers did. I tried to give my speech without hesitation. Model videos were very helpful. (Student 11 [S11])

> We watched model videos last week. I thought that they seemed to enjoy delivering their presentations although they looked nervous. (S13)

These show that learners made an effort to deliver speeches for their own purposes, not just for the course requirement. Moreover, as with Okada et al.'s prior studies (2017, 2018), students find it important to pronounce English words correctly. In the present study, six students in Group B discovered the necessity of improving either vocal or visual features of the presentation to give a positive impression to the audience:

> When viewing average model videos, I thought that a good speech was that the speaker used appropriate speaking volume and pausing, in addition to good pronunciation. (S112)

In addition, words pertaining to sounds, that is, "to speak" "voice" and "pronunciation," appeared in Table 3 to explain specific aspect of students' presentation. After viewing the more-proficient models, a student in Group A learned how tone of voice would affect his subsequent oral presentation:

> I realized that not only the speaking volume of voice, but the speaking tone of voice is also important. The tone made a difference in the overall impression and affected the difficulty of understanding the speech, whether it was a bright or dark tone. (S4)

As this student indicated, tone of voice is an essential factor in communication, and the audience may have a different impression of the speaker depending on which tone he or she uses while talking. This excerpt suggests that the student was aware of the importance of using an effective tone of voice when speaking.

It was also evidenced that giving presentations multiple times provides students with opportunities to improve their public speaking skills, as in Okada et al. (2018). During the interview, participants in both groups reported that delivering oral presentations multiple times helped them get accustomed to delivering speeches:

> The third round was the final. We watched several models, some of which contained weaknesses. I became aware of how we had made negative aspects of presentation, such as not being able to make the audience understood clearly. I realized we should speak louder and at an appropriate pace to make the audience understood better. After reflecting on my first and second presentations, I wanted to overcome my weaknesses for the third presentation round. (S9)

> I became nervous for the first presentation round. Because I was so nervous, I wondered if I had memorized my speech. Well, I forgot what to say during my presentation. I realized that giving a presentation is different from practicing. However, I was confident that I did better for the second presentation round. (S111)

Similarly, a student in Group B reported that she focused on improving her eye contact skills for the third presentation, because she had found it difficult to use these skills:

> I used to give a speech staring at one point [in space]. Exceptional model speakers seemed to look over the audience, as if they had been talking to them. Therefore, I delivered my speech while keeping what I had learned from the models in mind. (S110)

This statement demonstrates that the student's experiences of delivering multiple presentations not only provided an opportunity to realize her own weaknesses but also motivated her to improve them.

Another student, however, reported that less-proficient models were more helpful in encouraging him to deliver speeches:

> Exceptional model videos deprived my motivation because they looked so perfect, but average models performed only a little better than we did, so I may be able to achieve that level with further effort. This encouraged me to practice my speech harder. (S15)

Thus, watching more-proficient model videos may not be necessarily beneficial for students learning a language; students may instead be best able to learn from models whose language proficiency is most equivalent to theirs. As Hodgson (2014) claims, language instructors should use successful non-native speaker models as the "motivational target" (p. 129).

Table 5. High-frequency words in the Final Reflection

	Word	Meaning	Total (A + B)
1	見る	*to watch*	54
2	上手	*exceptional*	39
3	思う	*to think*	34
4	良い	*good*	33
5	ビデオ	*video*	32
6	レベル	*level*	30
7	スピーチ	*speech*	29
8	人	*person*	23
9	平均レベル	*average level*	22
10	点	*point*	14

Finally, high-frequency words in the Final Reflection (see Table 5) were quite similar to those in the First and Second Reflections. As with the prior studies (Okada et al., 2017, 2018), the small number of participants meant that only a limited range of words appeared in the data. In addition, one student in Group A stated during the interview that the video viewing sequence might not be so important:

> … So I don't think that the viewing sequence order does matter. If you want us to have a positive impression of the video, then it might be better to watch an average model first and excellent one next. It would probably give us a good impression of the models. (S19)

Next, finally, the student responses regarding the difficulty of the three topics and the usefulness of the video types will be briefly reported. In Section One, students ($n = 30$)[6] responded that the topics were equally difficult overall; M and SD values for each were: "My favorite food" ($M = 2.77$, $SD = 1.25$), "Memories of high school club activities" ($M = 3.03$, $SD = 1.33$), and "My ideal trip" ($M = 3.17$, $SD = 1.42$). There was no significant difference in difficulty between the three topics, $F(2, 58) = 0.73$, $p = .49$, $\eta_p^2 = .03$. These results were consistent with those in Okada et al. (2018). In Section Two, the perceived helpfulness of each video type was as follows: model videos ($M = 3.90$, $SD = 1.00$), own videos ($M = 3.83$, $SD = 1.15$), peer videos ($M = 3.97$, $SD = 0.93$). There was no significant difference in the helpfulness of videos either, $F(2, 58) = 0.19$, $p = .83$, $\eta_p^2 = .01$. The results of both sections were consistent with those of Okada et al. (2018).

5 Discussion

The purpose of the present study was to replicate the findings of Okada et al.'s research (2017, 2018) in order to verify whether viewing non-native speaker models improved students' public speaking skills and whether the viewing order of model videos affected the quality of their oral presentations in the EFL classroom. The findings can be summarized as follows: (a) while overall self-evaluated scores improved over the course of the research project, the difference was not statistically significant; (b) for peer evaluation, students who watched the more-proficient models first and the less-proficient ones second gained significant increases in peer-evaluated score on the third presentation round; and (c) both more- and less-proficient model videos had a positive effect on subsequent oral presentations; less-proficient models allowed learners to make meaningful adjustments and more-proficient videos helped learners imitate the models' strengths.

Findings from the qualitative analyses showed that, as in the other studies (Okada et al., 2017, 2018), both more- and less-proficient model speakers stimulated learners' interest in delivering good oral presentations. Students learned from more-proficient models by trying to imitate the ways the speaker pronounced the language and looked over the audience. In contrast, less-proficient model speakers seemed to provide some hints on ways of improving learners' speaking skills by raising awareness of oneself and others. Thus, despite being non-native speakers, the models successfully provided participants with encouragement and confidence (Adams, 2004).

Furthermore, analysis of the reflections and interview provided evidence on the various factors, such as observing their own and their peers' video recordings and students' experience of multiple speeches that affect learners' oral presentations. EFL learners were able not only to improve their public speaking skills but also to learn how to assess these skills (in themselves and others) while delivering speeches, conducting self-/peer evaluation, and filling out reflections. In addition, this study integrated model video observation and evaluation of students' video-recorded presentations into the teaching of public speaking, which may have stimulated students' interest in further improving the target language and raised their self-confidence for subsequent rounds, regardless of their proficiency. In particular, learners with low proficiency in the target language may find it more practical and feasible to give oral presentations in an EFL context after such an intervention. In Makino's study (2016), while Japanese EFL learners failed to enhance their self-efficacy through developing grammar knowledge and reading skills, there was significant development based on improvement in speaking skills. Makino speculated that learners would become more confident if their paralinguistic features (e.g. rhythm, tone of voice, gesture) developed, and in this light viewed

speaking as a form of active learning that enhances students' self-efficacy. The current study endorses Makino's view along with the view that it is becoming increasingly important to practice public speaking skills in the EFL context, because it can help learners to find solutions to future problems as global citizens (Yamashiro & Johnson, 1997).

While findings from earlier studies (Okada et al., 2017, 2018) indicated that learners (Group B) who watched less-proficient models first and more-proficient ones next scored significantly higher in peer evaluation scores than those who did the reverse, this was not the case in this study; instead, participants (in Group A) who observed more-proficient models first and less-proficient ones next improved their peer evaluation scores. In the earlier studies, it is likely that students in Group B were peer-rated lower than those in Group A in the first evaluation round. For instance, for the original study (Okada et al., 2017), the first-round mean scores of voice control and body language for Group B were 3.17 and 3.01, whereas those for viewing in the opposite order were 3.19 and 3.18, respectively (while the results for effectiveness showed the opposite pattern). Similarly, in the first replication study (Okada et al., 2018), the first-round mean scores for voice control, body language, and effectiveness in Group B were 3.05, 3.07, and 3.34, whereas those in Group A were 3.34, 3.34, and 3.52, respectively. It was speculated in that paper that Group B's lower peer-evaluated scores may have helped improve scores in subsequent evaluation rounds compared to those of Group A. Likewise, in this study, it is probable that the ceiling effect[7] occurred in the first round of peer evaluation for Group B, but not Group A, partly because students in Group B may have delivered excellent presentations in the first round and/or may have evaluated their peers' speeches leniently. Thus, findings from peer evaluation analyses were inconsistent with those of previous studies. If the results had been consistent with earlier studies (Okada et al., 2017, 2018), this study could have suggested that viewing speaker model videos is an important factor that affects students' speech performance. Nevertheless, the findings of this study indicate other factors that affect students' speech performance.

First, it is important to explain that group dynamics might have affected the results of this study, even with the same teacher and instruction, although these factors did not appear in both qualitative and quantitative data. In Group A, 19 students took the course, all of whom participated in this study. The first peer evaluation scores for this group were lower than those of Group B, but their scores gradually increased in the second and third peer evaluations, which suggested that students had a positive influence on their learning. In Group B, on the other hand, there were only 12 participants, although the same number of students had enrolled for the course. The presence of both participants and non-participants in the group may have influenced participants' motivation to learn in the classroom. Students' motivation probably changes depending on other students in the same group, as Chang (2010) indicates the importance of shaping learners' motivation for learning a foreign language in the classroom.

Individual differences may also have affected this study's findings, which are different from those of earlier studies (Okada et al., 2017, 2018). Japanese university students usually take EFL courses in their freshman or sophomore years, regardless of their specialties. The participants in this study were majoring in economics and had to take the course as part of their graduation requirements, and it is thus likely that some students wanted to develop their English proficiency through the course, while others just wanted to receive credit. Students' goal orientation, such as performance and mastery goals in speaking, should be further explored, as discussed in Okada, Sawaumi, Ito and Fujii (2016). It is speculated that students in Group A more interactively engaged in delivering presentations and class activities than those in Group B. It is worth mentioning that it is important for learners to creatively and actively participate in the language learning process (Dörnyei & Ryan, 2015, p. 5). With the quantitative and qualitative data collected in this study, it was not possible to examine how group dynamics and individual differences as influential factors affected learners' motivation in classroom learning.

This study has some limitations, similar to those of Okada et al. (2017, 2018). First, this study was a classroom-based replication and did not include a sufficient number of participants. However, the accumulated findings may nevertheless contribute to further study of video viewing effects on

students' performance in foreign language learning. Second, it was necessary to enhance internal validity by using teacher evaluation in addition to student evaluation. Finally, models were selected by the first author of the study; in future work, it would help increase internal validity for instructors or researchers to select models collectively, on the basis of a set rubric.

6   Conclusion

In sum, the present study replicated the basic qualitative results of the previous studies; however, the quantitative results were not consistent with the previous studies (Okada et al., 2017, 2018). This is sufficient to let us conclude that model video viewing order has some degree of positive impact on students' oral presentations, but that other factors may also affect peer-evaluated results in particular, at least in the Japanese culture and society, where it is more valuable to achieve group goals than to show one's uniqueness (Yamaguchi et al., 1995). As American universities do, Japanese universities should offer more public speaking training courses to improve students' speaking skills.

This study has indicated that it would be beneficial for Japanese EFL learners to improve their public speaking skills and self-confidence by delivering speeches, evaluating their presentations based on the video-recordings, and observing non-native speaker models. However, although this study saw encouraging results for language learning from model video viewing, the participants were limited to Japanese university EFL learners; thus, it will be useful to conduct a study to examine whether observation of non-native models is also effective in other contexts (e.g. different nationalities, different age groups), in order to support the generalizability of the findings of this replication and the earlier studies. Furthermore, this study provided participants with opportunities to watch their own and their classmates' video-recordings, in addition to non-native speaker model videos. The effect of the model video viewing might have been more distinguishable, if the study had not used students' own and peers' video-recordings.

Nevertheless, irrespective of the limitations, viewing non-native speaker model videos can improve students' oral presentations and enhance their public speaking skills and confidence, in Japan and potentially also in other countries.

Acknowledgement

Notes

This is a revised and expanded version of a paper presented at the 15th Asia TEFL International Conference and the 64th TEFLIN International Conference, Yogyakarta, Indonesia, 13–15 July 2017.

[1] Although these students took the course and completed the measurement instruments (self- and peer evaluation forms and reflection papers), this study did not use their data for analysis.

[2] The average score for Group A was obtained from 18 students, because one had missed the test, and the average score for Group B was from 12 students who participated in this study.

[3] Three topics were selected for the speeches, assuming that the students were likely to be familiar with them.

[4] Among the participants, the term "average level" was used as an alternative to "less-proficient" to avoid negative impressions of the video clips. Likewise, "excellent level" was used as an alternative to "more-proficient" to avoid implying by contrast the term "less-proficient."

[5] As with this study, Okada et al.'s studies (2017, 2018) used Groups A and B, of which the former watched more-proficient models first and less-proficient ones second, while the latter watched in reverse order.

[6] One student missed the class in which the Final Reflection was completed.

[7] Ceiling effect is a phenomenon that occurs when test items are not challenging for a group of people and the test score will not increase further as the group has already reached the highest score.

References

Adams, K. (2004). Modelling success: Enhancing international postgraduate research students' self-efficacy for research seminar presentations. *Higher Education Research and Development, 23*(2), 115–130. doi:10.1080/0729436042000206618

Ananiadou, S., Thompson, P., Thomas, J., Mu, T., Oliver, S., Rickinson, M., … McNaught, J. (2010). Supporting the education evidence portal via text mining. *Philosophical Transactions of The Royal Society A, 368*(1925), 3829–3844. doi:10.1098/rsta.2010.0152

Bandura, A. (1971). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1977). Analysis of modeling processes. In A. Bandura (Ed.), *Psychological modeling: conflicting theories* (pp. 1–62). Chicago: Aldine-Atherton Press.

Chang, L. Y. (2010). Group processes and EFL learners' motivation: A study of group dynamics in EFL classrooms. *TESOL Quarterly, 44*(1), 129–154.

Chen, M. L. (2012). Effects of the order of reading text or viewing a film and L1/L2 captions on reading comprehension. *Perceptual & Motor Skills: Learning and Memory, 115*(1), 18–26.

Council of Europe (2017). *Common European framework of reference for languages*. Retrieved from https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions

De Grez, L., Valcke, M., & Roozen, I. (2014). The differential impact of observational learning and practice-based learning on the development of oral presentation skills in higher education. *Higher Education Research and Development, 33*(2), 256–271.

Doi, T. (1971). *Amae no kozo*. [The anatomy of dependence]. Tokyo: Kobundo.

Dörnyei. Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York: Routledge.

Educational Testing Service. (2017). *The TOEIC Bridge Test*. Retrieved from https://www.ets.org/toeic_bridge

Hancock, A. B., Stone, M. D., Brundage, S. B., & Zeigler, M. (2010). Public speaking attitudes: Does curriculum make a difference? *Journal of Voice, 24*(3), 320–307.

Hodgson, K. M. (2014). Mismatch: Globalization and native speaker models of linguistic competence. *RELC Journal, 45*(2), 113–134.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Reading in Psychology and Culture, 2*(1).

Kawauchi, C. (2016). University student perspectives on English language classroom anxiety: Majors, gender differences, and changes. *Bulletin of the Institute of Foreign Language Education, Kurume University, 23*, 15–40.

Makino, M. (2016). Relationship between speaking practice and self-efficacy toward English learning in remedial English class. *Studies in English Language Teaching, 39*, 1–15.

Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT). (2014). *English education reform plan corresponding to globalization*. Retrieved from www.mext.go.jp/a_menu/kokusai/gaikokugo/__icsFiles/afieldfile/2014/01/31/1343704_01.pdf

Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT). (2015). *Plans on the promotion of improvement of students' English abilities*. Retrieved from http://www.mext.go.jp/en/news/topics/detail/__icsFiles/afieldfile/2016/10/19/1378469_001.pdf

Murphy, J. M. (1992). Preparing ESL students for the basic speech course: Approach, design, and procedure. *English for Specific Purposes, 11*, 51–70.

Oanh, D. T. H., & Hien, N. T. (2006). Memorization and EFL students' strategies at university level in Vietnam. *The Electronic Journal for English as a Second Language, 10*(2), 1–21.

Okada, Y., Sawaumi, T., & Ito, T. (2014). Different effects of sample performance observation between high and low proficiency English learners. In Aishah M. K., W. M. Chan, S. W. Chi, K. N. Chin, S. Klayklueng, S. Kumar Bhatt, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Comps.), *Knowledge, Skills and Competencies in Foreign Language Education. Proceedings of the 6th Centre for Language Studies International Conference* (pp. 394–413). Retrieved from http://www.fas.nus.edu.sg/cls/CLaSIC/clasic2014/Proceedings/okada_yasuko.pdf

Okada, Y., Sawaumi, T., & Ito, T. (2017). Effects of observing model video presentations on Japanese EFL learners' oral performance. *Electronic Journal of Foreign Language Teaching, 14*(2), 129–144.

Okada, Y., Sawaumi, T., & Ito, T. (2018). How do speech model proficiency and viewing order affect Japanese EFL learners' speaking performances? *Computer-Assisted Language Learning-Electronic Journal, 19*(2), 61-81.

Okada, Y., Sawaumi, T., Ito, T., & Fujii, T. (2016). Tassei mokuhyo riron kenkyu no gaikan to eigo oral purezenteshon shido eno shisa. [Implications of achievement goal theory research for teaching oral presentation skills in English]. *Research Bulletin Liberal Arts Nihon University College of Economics, 82*, 59–72.

Porte, G. (2012). Introduction. In C. A. Chapelle & S. Hunston (Eds.), *Replication research in applied linguistics* (pp. 1–18). Cambridge: Cambridge University Press.

Shimizu, M. (2007). Effects of inference question sequence on reading comprehension. *Annual Review of English Language Education in Japan, 18*, 51–60.

Winke, P., Gass S., & Sydorenko T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, 14*(1), 65–86.

Yamaguchi, S., Kuhlman, D. M., & Sugimori, S. (1995). Personality correlates of allocentric tendencies in individualist and collectivist cultures. *Journal of Cross-Cultural Psychology, 26*(6), 658–672.

Yamashiro, A. D., & Johnson, J. (1997). Public speaking in EFL: Elements for course design. *The Language Teacher, 21*(4), 13–17.

Young, D. (1990). An investigation of students' perspectives on anxiety and speaking. *Foreign Language Annals, 23*(6), 539–553.