

Can Video-creation Project Work Affect Students' Oral Proficiency? An Analysis of Fluency, Complexity and Accuracy

August 2020 – Volume 24, Number 2

Ryan Spring

Tohoku University

<spring.ryan.edward.c4@tohoku.ac.jp >

Abstract

This study reports on how project-based language learning in which L1 Japanese EFL learners created short videos affected L2 oral proficiency. Students took short speaking tests before and after the class, and the fluency, complexity and accuracy of the pre- and posttests were measured to see which, if any, of these three aspects of proficiency would show improvement. The results indicated that participants made marginal progress in fluency, reducing their number of pauses and increasing their raw speech rate slightly, and significantly improved their syntactic complexity ($p < .001$, $d = 1.1$) and both syntactic ($p = .04$, $d = .48$) and pronunciation accuracy ($p = .002$, $d = .75$), but did not seem to make gains in lexical complexity. Overall, the results suggest that project-based learning can result in clear improvements in oral proficiency, meaning that it can be appropriately implemented in oral communication classes, but that the greatest gains are likely to be made in accuracy and syntactic complexity. However, it is still unclear whether different types of project settings will affect L2 oral proficiency in the same way.

Project-based language learning (PBL), a teaching methodology in which students learn or practice a foreign language by participating in project work in the target language, has been gaining increasing attention in foreign language teaching research because of its student-centered approach and focus on improving students' communicative competencies in a variety of settings from elementary school through university (e.g., Beckett, 2006; Beckett & Slater, 2005; Liu, 2016; etc.). However, while several studies suggest that PBL can help students to improve their oral communication in their target language (e.g., Kobayashi, 2006; Liu, 2016; Dooly & Sadler, 2016), other studies report little to no benefit to target language communication skills (Eguchi & Eguchi, 2006), and most offer no hard evidence of improvement, instead relying on surveys of student opinion or reporting what the teachers felt went well (Hong, 2019; Dooly, 2013; Foss et al., 2007; etc.). Therefore, it is still unclear whether PBL can enhance oral proficiency, and if it can, to what degree and in what specific ways. Here, an analysis of objective measures of speech, specifically

fluency, complexity and accuracy, can help to determine whether or not PBL leads to improvement in oral proficiency, and what kinds of improvement instructors can expect from such project-based work in the EFL classroom so that they can plan their classes and projects accordingly. This paper aims to fill this gap in the existing research by analyzing the pre- and posttest speaking test data of 40 L1 adult Japanese EFL students who participated in a PBL class in which they created short videos.

Literature Review

PBL is a form of project-based learning in which the completion of a project in the target language is used to help improve students' foreign language skills while also improving critical thinking and content knowledge (Beckett & Slater, 2005; Foss et al., 2007; Stoller, 2006). It is a highly communicative approach (Dooly, 2013; Eguchi & Eguchi, 2006) grounded in social constructivism that has been claimed to enhance a number of communication-based skills such as intercultural communication competence and social interaction (Godwin-Jones, 2013; Kobayashi, 2006) and willingness to communicate (Farouk, 2016; Liu, 2016). Furthermore, studies such as Hong (2019) have shown that students are generally receptive to the idea of learning and practicing their target language(s) through projects. However, there is some disagreement as to how much the implementation of PBL into an EFL class can improve specific language skills.

Despite the arguments advanced regarding the benefits of PBL, evidence showing that it has linguistic benefits in an EFL classroom is still weak, and the findings from different studies sometimes conflict. For example, Dooly and Sadler (2016) report that as a result of implementing PBL into EFL classes, learners improved their ability to produce target language structures orally, particularly those related to modality and creative reproduction, better than students who did not participate. However, their evidence was largely subjective in nature, based on the speculative observations of the researchers. Torres and Rodriguez (2017) also report that PBL increased EFL learners' oral production and helped develop their lexical competence, but their evidence was based on transcripts of interviews and post-treatment surveys that asked whether or not students felt they had improved. Conversely, Eguchi and Eguchi (2006) reported that an implementation of PBL through a magazine-creation project caused very minimal impact on students' communication skills, although again their data were based on post-treatment surveys of students' opinions of how much they felt they had learned. Thus, though there is some evidence that PBL can aid in communication skills, and specifically oral proficiency, the methods of data gathering have thus far tended to be subjective or focused on whether or not students felt they had learned (as opposed to whether or not they had actually learned), which may be the reason why there is not agreement as to how effective PBL is or which specific communication skills it affects.

However, another reason for the discrepancies in the aforementioned studies could be the fact that the project that each group of students was assigned to complete was different in each study. As noted by a number of PBL studies, the type of project and how it is implemented can greatly influence how well PBL can be actualized in the classroom and how much it will benefit students (Farouk, 2016; Foss et al., 2007; Tamin & Grant, 2013). One type of project that has been argued to be highly appropriate for PBL is video or short-film creation (e.g., Foss et al., 2007; Miller & Hafner, 2014) because it tends to garner student interest and is easily shared with the outside world, which means it meets the guidelines for a driving or burning topic (Farouk, 2016), and includes speaking roles, which are argued to be good for acquiring foreign language speaking skills and pronunciation (Hardison & Sonchaeng, 2005; Floss et al., 2007). However, again there is not much

objective evidence to support the idea that video-creation PBL specifically works to enhance oral proficiency. For example, Foss et al. (2007) suggested that video projects were valid as a form of PBL for intensive English programs, but offered no data to show that it actually improved any specific skill or proficiency. Furthermore, Miller and Hafner (2014) reported that video-creation PBL helped raise student motivation and their feelings toward the class, but did not actually measure how much specific language skills had improved.

One way to shed light on whether or not implementing video-creation PBL in the EFL classroom can boost students' speaking skills is through objective pre- and post-treatment measures of oral proficiency. Though there are various facets of speaking, oral proficiency is considered to be indicative of spoken communicative competence and L2 speaking ability, and when rating it objectively, there are three aspects of speech that have been widely shown to be highly correlated with subjective evaluation: fluency, complexity and accuracy (Lambert & Kormos, 2014; Lu, 2012; Skehan, 2009; Thai & Boers, 2016; Vercellotti, 2017; etc.). Furthermore, though some studies argue that these aspects are intertwined (e.g., Vercellotti, 2017), others assume that certain areas of L2 oral proficiency can improve separately of the others and that certain tasks will improve some aspects of it and not others (e.g., Skehan, 2009; Thai & Boers, 2016). Therefore, based on the aforementioned previous studies, though there is reason to believe that video-creation PBL can possibly improve learners' oral proficiency, there is currently a lack of objective data to verify this and elucidate which aspects of it (fluency, complexity or accuracy) actually improve. This study aims to fill in these gaps in the previous literature by answering the following research questions:

1. Do L1 Japanese EFL students participating in video-creation PBL improve their oral proficiency in ways that can be objectively measured?
2. If so, which aspects (fluency, complexity, or accuracy) are improved and which are not?

Methods

Participants

40 L1 Japanese EFL learners at Tohoku University, a large national university in Japan, agreed to participate in this study in accordance with the ethical guidelines of the university. All of the participants were second year students between the ages of 19 and 20, and had studied English for 7 years before the beginning of the class, including a combined 6 years in junior and senior high school, and two semesters of both an English reading and an English communication class prior to class enrollment. None of the participants had lived abroad or visited another country for longer than a month. Their average TOEFL ITP¹ score, taken 4 months prior to the class, was 511 ($SD=40.39$), indicating students were around a CEFR² B1 level. The participants were students in two different classes consisting of 15 females and 25 males, 18 of whom were engineering majors and 22 of whom were law majors. All learners were enrolled in a "Practical English Skills" class, taught by the author, which is a required class for graduation, although students are given a choice of instructor and class content. No participants were taking any other English classes or conducting any sort of additional English study during the semester, so it is likely that any gains in their L2 oral proficiency came from participation in the class. Due to this assumption and the fact that no assertions are being made in this paper that PBL, or the class described herein, is either superior or inferior to other teaching methods or classes, no control group was used. All students signed

informed consent forms agreeing to participate in this study and to provide speaking test data before the class began and after the class concluded.

Procedure

Students participating in the study were in one of two identically designed and operated classes taught by the author, a native speaker of English. The basic class design is represented in Table 1 and consisted of fifteen 90-minute sessions, held once per week for a total of fifteen consecutive weeks. During the first three sessions, the instructor taught five academic modules to the students that they would need and be expected to utilize during project work: how to conduct a meeting, useful phrases for discussions, how to write minutes of a meeting, how to give a short presentation, and how to conjecture the meanings of academic English words by using word parts (i.e., prefixes, suffixes and word roots, especially those originating from Greek and Latin).

After the third session, students formed groups of four to six members at their own discretion, resulting in one group of six, six groups of five, and one group of four, and began working on their projects, planning and creating a video, in the class. Project work and group meetings were expected to be conducted in English, as per the basic concept of PBL and because this was considered the majority of students' communicative (speaking and listening) practice. At the beginning of each project work session (with the exception of session number four, as it was the first project work session), one student from each group was asked to give a two-minute oral presentation about their group's progress based on their meeting during the prior session. After each group's presentation, a different group, chosen randomly by the instructor, was required to ask the presenter one question. After presentations, groups were given time to have meetings, work on their projects (e.g., write scripts, practice acting or narrating, etc.) and receive feedback from the instructor. The instructor monitored meetings to ensure that students were conducting them in English, helped students with their scripts and storyboards, gave pronunciation coaching to actors and narrators, and offered tips and advice regarding video editing. During each project work session, one student from each group had to write the meeting minutes, in accordance with the instructions given during the third session. These were given to the instructor for feedback and scoring, based on content and effort. The writer of the meeting minutes was then required to give the oral presentation at the beginning of the following session. As there were 10 project work sessions and generally five students in each group, almost every student (with the exception of smaller and larger groups) was therefore required to write meeting minutes and present on them twice over the duration of the class.

In addition to their project work, students were expected to learn some of the basics of video-creation. This content was taught through a flipped-classroom approach, by having the students read a chapter in a textbook created for the class that explained important points regarding how to create a short video and taught them academic vocabulary. There were a total of five chapters, the contents of the respective chapters being: how to work on a team, how to write a script, how to plan and create a good visual message, how to create a good spoken message through acting and narrating, and how to edit. The chapters were designed to be around the CEFR² B2 or B1 level and each chapter contained 20 academic words, taken from Coxhead (2000) and used several times in the respective chapter in context. These words were often comprised of the word parts that were taught to students during the second class. Quizzes were administered during the sessions following a reading assignment (sessions 5, 7, 9, 11 and 13) to check for comprehension and that the students had learned the meanings of the academic words. Students were also expected to discuss the contents of the chapters that they had read and how they would implement the information into

their project during their meetings, thus forcing them to integrate outside knowledge, a core tenet of PBL (Beckett & Slater, 2005).

Table 1. Class Progression.

Session	Content
1	Academic modules: How to conduct a meeting, phrases for discussion
2	Academic module: Using word parts to remember academic words
3	Academic modules: How to give a presentation, writing meeting minutes
4	Group work 1
5	Presentation 1, Chapter 1 quiz, Group work 2
6	Presentation 2, Group work 3
7	Presentation 3, Chapter 2 quiz, Group work 4
8	Presentation 4, Group work 5
9	Presentation 5, Chapter 3 quiz, Group work 6
10	Presentation 6, Group work 7
11	Presentation 7, Chapter 4 quiz, Group work 8
12	Presentation 8, Group work 9
13	Presentation 9, Chapter 5 quiz, Group work 10
14	Presentation 10, Group work 11
15	Final presentation, showing films in class

Because of the flipped-classroom approach and group-work oriented nature of the class, when students were absent, they could not exactly make up the group work conducted in the project work sessions. When students missed a session, they were instructed to contact their group members, review the meeting minutes and help their groups appropriately. Furthermore, they were still held responsible for the out-of-class readings and their portion of the project work, as determined by their group. Fortunately, very few students in the data set missed sessions (five students had one absence each and one of them was also late to another session), so the odds that absences severely disadvantaged students overall were quite low. The data set of students who had at least one absence was too small to determine if being absent had a significant impact on oral proficiency, so this aspect was not considered in the data analysis of this study.

The instructor provided some basic guidelines for the videos that the students created. They were allowed to be either documentary-style video descriptions of a topic that students were interested in, or a narrative-style short film. However, videos were required to be at least five minutes in length and the language was set as English. Furthermore, students were asked to include in their videos five academic words and one idiom that they did not previously know from a provided list. Finally, no inappropriate or potentially offensive material was allowed, as the videos were to be shown to their friends and peers in a viewing during the last session. The topic matter and content were left to the students to discuss with their groups and decide themselves. Examples of student videos included a video introducing and explaining the university campus, a story about a bear that could turn into a human, and a humorous, fake documentary about a traditional Japanese game. A webpage was also provided where students could communicate via a student forum, share their work with others and eventually enter their short video into a school-wide filmmaking competition.

Student assessment was based on five criteria that were explained in detail during the first session: participation, project work, quiz scores, a teamwork score, and the quality of the final project. Participation scores were given for each session by the instructor, with points being deducted for overuse of the L1, being tardy, or unexcused absence. Project work was based on the quality of the meeting minutes they wrote, presentations they gave, and a personal statement about what they had learned in the class written during the final session. Presentations were graded based on the content and presentation manner (i.e., body language, eye contact, appropriate voice volume and intonation) as taught during the second session, meeting minutes were assessed according to their adherence to the rules outlined in the textbook and amount written. Teamwork scores were based on the TEAM Q teamwork evaluation method created by Britton et al. (2017), which was implemented to help ensure that students were actively participating and help the instructor know which students were properly working on their projects and aiding in the discussions during the sessions. The same score was given to every group member in the same group for quality of the final product so as to encourage students to work together and utilize the knowledge given in their textbooks. These grades were used to keep students motivated to conduct project work in English and to assess their performance in the class, but were not actually used for analysis in this study.

All participants took pre- and posttests of speaking at the end of the first session and after the final session; the scores on these tests did not affect the student' final grades. All students were tested using the same sets of tasks. The pre-test and posttest sets of tasks were to give short monologues based on practice questions taken from the IELTS oral test and were therefore similar, but not identical. For example, in the pre-test students were asked to describe the area they lived in and in the posttest, they were asked to describe the home that they lived in. The tests were conducted by asking each participant to answer three questions orally, which were shown to them one at a time. After each question was presented, participants were allowed one minute to prepare before answering, and then they were asked to speak for up to two minutes. When participants felt that they had sufficiently answered the question, they were allowed to stop speaking, but non-answers were not allowed. A timer was used, and if participants were still talking after two minutes, they were asked to stop talking. Participants were made aware of this time constraint before beginning the tests. All participants' responses were recorded and then assessed objectively on measures of fluency, complexity and accuracy.

Data Analysis

Fluency. There are three dimensions of fluency that are generally considered in objective measures: speech fluency, breakdown fluency, and repair fluency (Tavakoli & Skehan, 2005). Studies such as Thai and Boers (2016) and De Jong and Wempe (2009) suggest that speech fluency can be considered in terms of how quickly someone speaks (i.e., raw speech rate and articulation rate). Breakdown fluency is generally measured as a function of how often one pauses in their speech (Skehan, 2014), and two common representative objective measures of this are the number of pauses in one's speech and the amount of time spent talking versus spent silent (i.e., phonation ratio) (Wood, 2010). Repair fluency is generally observed by looking at how quickly one delivers meaningful syllables (Lennon, 1990). Following these previous studies, five measures of fluency were considered in this study: trimmed speech rate, raw speech rate, number of pauses, articulation rate, and phonation ratio. The speech analysis software Praat (Boersma & Weenink, 2019) was used to conduct the calculations for the latter four measures for ease and objectivity. Raw speech rate is calculated as the number of syllables per second that participants spoke. The number of pauses was calculated via Praat, using the specifications set by Wood (2010), i.e., considering a threshold of

0.3 seconds to be a pause. Articulation rate is defined as the number of syllables per second for only active speech (in other words, the amount of time minus the amount of silence). Phonation ratio is the amount of time spent speaking divided by the total amount of time. Trimmed speech rate was calculated as the number of trimmed syllables that participants spoke per second. Trimmed speech refers to the number of meaningful syllables spoken by a participant (Lennon, 1990; Thai & Boers, 2016). This was found by transcribing each participant's speech excluding repeated syllables (e.g., "*I ... I ... P*"), false starts (e.g., "*I had a ... I mean I was a ...*"), filled pauses (e.g., "*uhhh*") and repairs (e.g., "*I want to go ... no, I WANTED to go*").

Complexity. According to Skehan (2009), two measures of spoken complexity need to be accounted for: syntactic and lexical. Syntactic complexity was calculated by determining the number of clauses per analysis of speech (AS) unit in each participant's trimmed speech. AS units were used instead of t-units because the latter were originally based on L1 written language, and thus can be problematic when segmenting spoken data, especially for L2 learners, whereas the latter were specifically created for handling difficult segmenting cases (Foster et al., 2000). An AS unit is described by Foster et al. (2000) as a speaker's utterance that has at least one independent clause, or sub-clausal unit, along with any associated subordinate clauses, and their guidelines for AS units were followed (i.e., including counting coordinated verb phrases). Lexical complexity refers to the range of vocabulary that a speaker uses (Skehan, 2009). While there are several ways to examine lexical complexity (e.g., number of words beyond the most frequent 2000 – Thai & Boers, 2016; a calculation of lexical diversity known as D – see end note 3 or McKee et al., 2000 for a detailed description), Lu's (2012) multivariate analysis of several different lexical complexity measures indicated that only nine were clearly correlated with higher scores on professionally rated oral L2 English speech tests³. Several of his measures were adopted for this study because the data analyzed by Lu (2012) were similar to the data used in this study (short monologues about a given topic) and can be calculated through software ensuring objectivity and inter-rater reliability (in that the exact same score will always be given to the same response). Though Lu (2012) distinguished nine measures that were highly correlated with higher lexical complexity, several were calculated very similarly. For example, Lu identifies both the average number of different words in 10 sets of 50 words selected randomly from the monologue (NDW-ER50) and the average number of different words in 10 sets of random 50-word sequences, i.e., stretches of speech containing 50 words (NDW-ES50), as indicative of lexical complexity, but they are both ways of estimating the number of different words a speaker is likely to use per 50 words. For the purposes of this study, one of each such pairs or groups of similar measures is sufficient to indicate change in lexical complexity, so one representative variable (the one with the highest reported Fisher's Z and lowest *p*-value) for each type of measurement was selected for analysis: the overall number of different words in the monologue (NDW), the estimated number of different words (NDW-ES50; as explained above), the corrected type-token ratio (CTTR; the number of different words divided by the square root of two times the total number of words), and the corrected verb variance (CVV1; the number of different verbs used divided by the square root of two times the total number of verbs used). These four measures were automatically calculated through the online program provided by Lu (2012), set to American English, because Japanese EFL students learn American English in junior high school and high school.

Accuracy. Two measures of accuracy were conducted in this study: syntactic accuracy and pronunciation accuracy. Syntactic accuracy was adopted from works such as Tavakoli and Skehan (2005), Thai and Boers (2016), and Vercellotti (2017) as the percentage of error-free clauses in participants' trimmed speech. This measure was utilized because there is less variation in clause

length than AS unit length, and thus it is less likely to punish speakers who attempt longer utterances than measures such as the ratio of error-free units to total number of units (e.g., Robinson, 2001). The number of clauses was calculated with the syntactic complexity calculator provided by Lu and Ai (2015) for ease and inter-rater reliability, and the number of error-free clauses was determined by two native speakers of English, who marked errors in the participants' trimmed speech transcripts. When discrepancies between their judgments were found (6 times out of 80 transcripts; e.g., when a participant used the pronoun "he" when probably referring to his mother), the researcher consulted with them and a collaborative decision was made as to whether or not it should be counted as an error. While pronunciation accuracy has not been looked at in many studies of objective speech analysis, it has been considered in other lines of work such as Suter (2006). However, such studies generally use the subjective judgments of multiple native speakers. In order to check pronunciation objectively and to avoid problems with inter-rater reliability, pronunciation accuracy was checked in this study utilizing automatically generated subtitling software. Following Spring (2020), participants' speech was uploaded to YouTube and then the automatically generated English subtitles were saved as text files. Two research assistants listened to the same files and transcribed them. Their texts were checked and when there were discrepancies (4 times out of 80 transcripts), the author listened and made final judgment. Pronunciation accuracy was calculated as the number of words that YouTube's automatically generated subtitles could accurately predict as compared to the human created transcripts, divided by the total number of words spoken, which has been shown to be a reasonably reliable measure of objectively assessing EFL pronunciation (Spring, 2020).

Statistical Analysis. After the measures of fluency, complexity and accuracy were calculated as described above, dependent t-tests were performed on the pre- and posttest data for each measure to determine if statistically significant improvement had been achieved. Cohen's d was then calculated as a measure of effect size for measures showing statistically significant differences. Additionally, the length of speech of the pre- and posttest data was compared (as seconds spoken and number of words spoken) as this can potentially influence measures of complexity, fluency and accuracy, but no significant differences were found (number of words: $p=.17$, seconds: $p=.16$).

Results

Fluency

The results of the two measures of fluency are reported in Table 2. Dependent t-tests showed that participants improved their raw speech rates and number of pauses significantly, but not their articulation rate, phonation ratio, or trimmed speech rates. This suggests that with regards to speech fluency, participants made utterances overall more quickly, but that this was likely due to a decreased number of pauses, and not as a result of pronouncing syllables more quickly. Furthermore, the data also suggest that though participants paused less frequently in their posttests, the length of pauses were probably slightly longer. Finally, a comparison of trimmed speech rates and raw speech rates indicates that many participants must have been using more "meaningless" fillers as a communication strategy, or self-correcting in their posttests than in their pretests. Looking at the actual utterances made by participants, this idea seems to be supported, as there seemed to be more occurrences of filler words such as "ummmm" and "well" in the transcripts of the posttest data, as well as self-corrections and restarts such as "... and I live in the condo.... I HAVE lived in a condo since I was born" and "...so I like the... so I like the location". Overall, considering the fact that (1) only two of the five measures of fluency showed significant

improvement, and (2) the effect sizes were not particularly large for either of the measures that showed statistical significance, it seems that participants only made marginal improvement in their fluency. Furthermore, the qualitative data presented here suggests that most of the changes in fluency were related to an increase of self-corrections and restarts.

Table 2. Fluency scores of pre- and posttests.

Measurement	Pretest mean (SD)	Posttest mean (SD)	Test of significance
Raw speech rate (syllables per second)	2.27 (0.67)	2.47 (0.58)	$t(39) = -2.137, p=.039, d=.48$
Articulation Rate (syllables per second excluding pauses)	3.82 (0.34)	3.73 (0.38)	$t(39) = -1.565, p=.127, n.s.$
Number of pauses	74.1 (49.09)	63.03 (41.38)	$t(39) = -2.164, p=.037, d=.25$
Phonation Ratio (percentage of time spent making utterances)	.65 (.13)	0.67 (0.12)	$t(39) = 1.283, p=.207, n.s.$
Trimmed speech rate (syllables per second)	2.25 (0.73)	2.42 (0.67)	$t(39) = -1.718, p=.094, n.s.$

Complexity

The results of the measures of syntactic and lexical complexity are shown in Table 3. Dependent *t*-tests determined that participants made significant improvement in syntactic complexity (clauses per AS unit) and verb variance (CVV1), but not in overall number of words (NDW) or corrected token-type ratio (CTTR), and it seems that the estimated number of different words (NDW-ES50) actually decreased. These results, coupled with the fact that syntactic complexity showed the largest effect size, suggest that while participants grew to use more complex sentence structures, e.g., sentences with multiple clauses, and a wider range of verbs, there is no evidence to support the idea that they used a wider range of vocabulary in general. Examples from pre- and posttest transcriptions also support this notion. For example, in the pretest participants were more likely to use single clause sentences such as “*I live in an apartment*” and “*I admire the seniors of my club activities*”, but grew to use more multiple clause sentences in their posttests such as “*The home that I live in is located in XXXX Prefecture*” and “*When I was in third grade at high school, I felt disappointed because my mother didn’t understand my path*”. Though participants did improve their CVV1, indicating that they used a wider variety of verbs during the posttest, the fact that two of the measures of lexical complexity (NDW and CTTR) did not show statistically significant improvement, and one other (NDW-ES50) actually decreased suggests that there was probably not an overall tendency to improve in this area.

Table 3. Complexity scores of pre- and posttests.

Measurement	Pretest mean (SD)	Posttest mean (SD)	Test of significance
Syntactic complexity (clauses per AS unit)	1.45 (0.23)	1.63 (0.28)	$t(39) = -4.876, p < .001, d = 1.1$
Total number of different words	119.28 (39.19)	121.73 (39.57)	$t(39) = .997, p = .325, n.s.$
Estimated number of different words (NDW-ES50)	36.45 (2.51)	35.28 (2.02)	$t(39) = -2.73, p = .009, d = .52$
Corrected type-token ratio (CTTR)	5.25 (0.75)	5.08 (0.79)	$t(39) = -1.579, p = .122, n.s.$
Corrected verb variance (CVV1)	2.32 (0.51)	2.62 (0.51)	$t(39) = 3.667, p = .001, d = .82$

Accuracy

The results of the two measures of accuracy as measured in the participants pre- and posttests are shown in Table 4. Dependent t-tests revealed that both measures showed statistically significant improvement, with syntactic accuracy exhibiting a moderate to low effect size and pronunciation accuracy showing a medium-large effect size. Thus, the participants showed slight improvement from the pretest to the posttest in their accuracy in general as well.

Table 4. Accuracy scores of pre- and posttests.

Measurement	Pretest mean (SD)	Posttest mean (SD)	Test of significance
Syntactic accuracy (% of error free clauses)	0.55 (0.15)	0.59 (0.15)	$t(39) = -2.125, p = .04, d = .48$
Pronunciation accuracy (% of words accurately transcribed)	0.85 (0.09)	0.90 (0.05)	$t(39) = -1.718, p = .002, d = .75$

Discussion

The results of this study suggest that the participants made demonstrable improvements in their oral proficiency through the video creation PBL class described in this paper. This is in line with research that has suggested that PBL can help to improve the communicative abilities of participants (e.g., Dooly & Sadler, 2016; Kobayashi, 2006; Liu, 2016) and research that has specifically suggested that video creation PBL can help to improve participants' speaking ability (e.g., Dooly & Sadler, 2016; Foss et al., 2007; Miller & Hafner, 2014). Specifically, the results suggest that participants' fluency, accuracy and syntactic complexity improved to some degree, but that little to no improvement was seen with regards to lexical complexity. This could be due in part to the nature of how these areas of spoken proficiency generally develop (Vercellotti, 2017) and in part because PBL is a highly communicative approach (Dooly, 2013) and understanding others and being understood by them in oral meetings and discussions is a large focus of a PBL class. This is important, because many EFL curricula, especially in Asian countries, are shifting to more communicative approaches, focusing specifically on speaking skills and oral proficiency (Kobayashi, 2006; Liu, 2016; etc.), and this study suggests that PBL has a place as one teaching method within such a context. Furthermore, indicating which aspects of oral proficiency can be improved through PBL and other methods will help instructors and educational institutions find the best fit for their students' specific needs.

With regards to fluency, the data reported in this study showed that participants' raw speech rate improved and the number of pauses they made decreased, but their trimmed speech rate, articulation rate and phonation ratio did not. When taken together with the effect sizes and qualitative data, it seems that though participants actively attempted to continue speaking more on their posttests, they were not able to produce content more quickly. This appears to be due the fact that they took longer pauses, albeit fewer, and seemed to be using more filler words and self-corrections. A decrease in pauses and an increase in filler words suggests that they perhaps became better at oral strategies that aid communication (Santos et al., 2016), but that this did not necessarily aid in expressing ideas more quickly. This could be due in part to the communicative nature of PBL. Throughout the class, students had to have meetings and discuss their ideas in groups to complete their project, so explaining themselves clearly is important, but doing so quickly is perhaps not as vital. The increased number of restarts and self-corrections also suggests that participants perhaps improved their ability to monitor their own language performance, which can be considered a general improvement in speaking ability that would also make sense if they were most concerned with communicating, i.e., clearly expressing themselves.

Participants in this study also showed improvement in the measure of syntactic complexity, but it does not seem that they made strong gains in lexical complexity. One reason for this could be that according to Vercellotti (2017), lexical complexity does not show linear growth in the same way that accuracy and fluency tend to. Instead, Vercellotti (2017) suggests that lexical complexity dips first and then later exhibits a steeper increase. The results of this study seem to match hers in this way. Another possible reason as to why lexical complexity did not show the same level of improvement could be due to the communicative nature of PBL. Using syntactically more complex sentences allows for the communication of more complex ideas and precise language and is more appropriate in a communicative context (Faigley, 1980). For example, the use of a relative clause allows for a much higher degree of specification than a simple adjective (e.g., *Please give me the yellow pencil* versus *Please give me the pencil that the teacher lent us last class*). However, though using a wider variety of words (i.e., higher lexical complexity) can aid communication, it is not always necessary for it, and could in some cases hinder it, if one were to use words that others did not know. Though using more lexically rich expressions may warrant higher perceived oral proficiency by native speakers of English (Lu, 2012), in the sessions, students most often communicated with other foreign language learners, not native English speakers. When speaking to other learners, using a large amount of fringe vocabulary could potentially impede others' understanding, leading to a breakdown in communication. Thus, making use of a larger vocabulary may not have been seen as especially important or helpful for students in the class, which could be one of the reasons that less improvement was seen in this area. This can perhaps explain the suggestions of Eguchi and Eguchi (2006) that increased exposure to native speakers of the L2 might have enhanced students' communicative abilities, but seems to indicate that the exact proficiency that would be bolstered most by such interaction would be lexical complexity.

With regards to accuracy, participants exhibited some degree of improvement through the class, as indicated by the significant differences in pre- and posttest measures, but with varying effect sizes. This finding is not surprising given the earlier mentioned finding that students seemed to improve their ability to self-monitor their speech, and calculations of semantic accuracy are based on trimmed speech. Furthermore, as Vercellotti (2017) showed, gains in both spoken fluency and accuracy often occur in a linear fashion along with one another, so this could be an indication of some overall boost in spoken proficiency. Another reason that accuracy in particular improved could be that in order for students to be understood when having discussions in class, they needed

to have a certain degree of accuracy in their spoken output. If too many grammatical or pronunciation errors occurred, it is possible that other group members could not understand what they were saying, which would lead them to redouble their efforts to speak more accurately. However, these gains could also be due in part to the nature of the project chosen for this class. To create their short videos, students had to write scripts, which were then checked and corrected by a native speaker, and actors and narrators were also coached on pronunciation by the instructor. Students then had to practice saying their lines multiple times both before and during filming. Such activities place heavy focus on both syntactical and pronunciation accuracy as they needed to be sure that they were saying their lines correctly, i.e., using the correct words from the script, and pronouncing them understandably, as predicted by Hardison and Sonchaeng (2005). To test this notion, a two-way ANOVA test with one repeated measure was conducted on participants who were known to have taken speaking roles, but it showed no significant interaction between taking a speaking role and improvement in pronunciation ($F[1, 23]=0, p=1$). Admittedly, the sample size in the ANOVA test was quite small, and some of the participants' exact roles were not explicitly given in the data, so it is still unclear as to why participants improved on these metrics of accuracy with only the data from this study.

Though the results of this study seem to suggest that video creation can have a positive, measurable effect on students' oral proficiencies, it should be noted that there are several limitations. First, a number of variables were used in this study (12 in total), which could increase the odds of type I errors. Though the data were taken from the same group of participants, the variables were all different measures of the same data, so methods of correcting type I errors, such as a Bonferroni correction, could not be utilized. Therefore, caution should be exercised in interpreting the data, and for this reason, the results should not be understood on the basis of the p-values alone, but in conjunction with the effect sizes, qualitative data, and a wholistic view of which measurements showed increases. In other words, it is not sufficient, for example, to say that speech fluency improved simply because the raw speech rate showed statistically significant improvement. Rather, conclusions about the improvement in fluency should be drawn from the overall results of the data relating to fluency. With this in mind, it only seems reasonable to think that fluency marginally improved, because only two of the five measurements showed significant differences, and the effect sizes on the variables that did improve were not very large. Similarly, syntactic complexity showed the clearest improvement according to the data, whereas one should not conclude that lexical complexity improved, as only one of the measurements showed significant increases, but half did not, and one actually showed a significant decrease. Finally, accuracy can be said to have clearly improved, as both measures showed statistically significant differences between pre- and posttests with reasonable effect sizes. The next limitation that should be mentioned is the fact that learners were studying English as a foreign language and that the class only lasted 15 weeks; only so much improvement can reasonably be expected in this time period. The metrics in this study indicate small but positive changes in oral proficiency, so while the results are hopeful, continued study and experience with the language would likely be required for more noticeable improvement. Next, it should be mentioned that the settings of the project and the class in general may influence the outcomes, as suggested by Foss et al. (2007) and Farouck (2016). For example, if students were to participate in PBL with native speakers of English, rather than other foreign language learners, there is the possibility that their spoken vocabularies might increase as well, due to the fact that native speakers will use a wider range of vocabulary, as suggested by Eguchi and Eguchi (2006). This is important because the use of a more varied vocabulary is correlated with higher oral proficiency rating by native speakers (Lu, 2012), so working to bolster this metric in participants will also be vital in the future. Furthermore, the gains in pronunciation accuracy found in this study

may have been partially due to the fact that students had to practice and read lines for video creation, and received pronunciation coaching from a native English-speaking instructor. Had a different project been used in which there was little to no performance aspect, e.g., writing a paper together or making a webpage together, students might not have focused on their pronunciation as much, and the same types of gains might not have been seen, although this is still unclear. Furthermore, there are a number of other confounding factors that could have influenced students' improvement such as individual effort and group size (Foss et al., 2007). However, the sample size of this study, while enough to draw conclusions about whether or not improvement could be seen from the class, is not large enough to make statistical analysis of these various factors. Therefore, future research should work to find if there are any correlations between certain aspects of the project work and specific gains in oral proficiency. For example, other studies on using video creation for PBL could aim to discover if there was more or less improvement in certain (or all) aspects of spoken proficiency depending on the type of video topic, how often students spoke in class, and their role in their group.

Conclusion

With regards to the research questions, this study found that (1) using short video creation as the project in a PBL class was effective in improving the general oral proficiency of L1 Japanese university students, and (2) that participants improved their fluency marginally (measures of raw fluency and number of pauses improved, but not articulation rate, phonation ratio, trimmed-speech rate) and their syntactic complexity and accuracy significantly (measures of both syntactic and pronunciation accuracy improved), but not their lexical complexity (though verb variance improved, type-token ratio and overall different words did not, and the estimated number of different words declined). The improvements are thought to be partially due to how L2 oral proficiency develops, and partially to the communicative nature of PBL. While this study was unable to account for factors that might have influenced the amount of individual improvement (i.e., video topic, group size, group role), and did include several variables, it provides a first look at how PBL can have an objectively measured, positive impact on students' oral proficiency and warrants further study in the future.

Notes

1. The TOEFL ITP, short for Test of English as a Foreign Language Institutional Testing Program, is an internationally recognized standardized test of English proficiency (see http://ets.org/toefl_itp/about for more details). [[back](#)]
2. CEFR, short for the Common European Framework of Reference of Languages, is an international standard used to describe foreign language ability. [[back](#)]
3. The nine variables that Lu (2012) identifies as best correlated with higher scores on oral L2 English speaking tests are: the number of different words (NDW), the number of expected different words from 50 random words (NDW-ER50) and from random 50 word sequences (NDW-ES50), the corrected type-token ratio (CTTR), the root type-token ratio (RTTR), mean segmental type-token ratio (MTTR-50), D measure (a measure of type-token ratio based on McKee et al., 2000), squared verb variation (SVV1), corrected verb variation (CVV1). [[back](#)]

Acknowledgements

I would like to thank the participants for agreeing to take part in this study and the reviewers and editors for their suggestions and contributions to this paper.

About the Author

Ryan Spring is an associate professor in the Institute for Excellence in Higher Education at Tohoku University. His research interests include applications of cognitive linguistics to second language acquisition and teaching and the use of multimedia in EFL teaching. He currently serves as the vice-president of the East Japan chapter of the Association for Teaching English through Multimedia.

References

- Beckett, G.H., & Slater, T. (2005). The Project Framework: a tool for language, content and skills integration. *ELT Journal*, 59(2), 108–116. <https://doi.org/10.1093/eltj/ccl024>
- Beckett, G.H. (2006). Project-based second and foreign language education: theory, research and practice. In G.H. Beckett & P.C. Miller (Eds.), *Project-Based Second and Foreign Language Education: Past, Present and Future* (pp.3–18). Information Age.
- Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer. <http://www.praat.org>
- Britton, E., Simper, N., Leger, A., & Stephenson, J. (2017). Assessing teamwork in undergraduate education: A measurement tool to evaluate individual teamwork skills. *Assessment & Evaluation in Higher Education*, 42(3), 378–397. <https://doi.org/10.1080/02602938.2015.1116497>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 43(2), 212–238. <https://doi.org/10.2307/3587951>
- de Jong, N., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, 41, 385–390. <https://doi.org/10.3758/BRM.41.2.385>
- Dooly, M. (2013). Promoting competency-based language teaching through project-based language learning. In M.P. Cañado (Ed.), *Competency-based language teaching in higher education* (pp.77–91). Springer. https://doi.org/10.1007/978-94-007-5386-0_5
- Dooly, M., & Sadler, R. (2016). Becoming little scientists: Technologically-enhanced project-based language learning. *Language Learning & Technology*, 20(1), 54–78.
- Eguchi, M., & Eguchi, K. (2006). The limited effect of PBL on EFL learners: A case study of English magazine projects. *Asian EFL Journal*, 8(3), 207–225.
- Faigley, L. (1980). Names in search of a concept: Maturity, fluency, complexity, and growth in written syntax. *College Composition and Communication*, 31, 291–300. <https://doi.org/10.2307/356489>

Farouck, I. (2016) A project-based language learning model for improving the Willingness to communicate of EFL students. *Systemics, Cybernetics and Informatics*, 14(2), 11–18.

Foss, P., Carney, N., McDonald, K., & Rooks, M. (2007). Project-based learning activities for short-term intensive English programs. *Asian EFL Journal*, 23. <https://www.asian-efl-journal.com/monthly-journals/project-based-learning-activities-for-short-term-intensive-english-programs/>

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–375. <https://doi.org/10.1093/applin/21.3.354>

Godwin-Jones, R. (2013). Integrating intercultural competence into language learning through technology. *Language Learning & Technology*, 17(2), 1–11.

Hardison, D., & Sonchaeng, C. (2005). Theatre voice training and technology in teaching oral skills: Integrating the components of a speech event. *System*, 33, 593–608.

Hong, C. (2019). Learning culinary English through food projects – what do students think? *The Asian ESP Journal*, 15(3), 99–124.

Kobayashi, M. (2006). Second language socialization through an oral project presentation: Japanese university students' experience. In G.H. Beckett & P.C. Miller (Eds.), *Project-Based Second and Foreign Language Education: Past, Present and Future* (pp.71–94). Information Age Publishing.

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, 607–614. <https://doi.org/10.1016/j.system.2004.01.001>

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387–417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>

Liu, X. (2016). Motivation management of project-based learning for business English adult learners. *International Journal of Higher Education*, 5(3), 137–145. <https://doi.org/10.5430/ijhe.v5n3p137>

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190–208. <https://doi.org/10.1111/j.1540-4781.2011.01232.x>

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16–27. <https://doi.org/10.1016/j.jslw.2015.06.003>

McKee, G.T., Malvern, D.D., & Richards, B.J. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15(3), 323–337. <https://doi.org/10.1093/lc/15.3.323>

Miller, L., & Hafner, C.A. (2014). Taking control: A digital video project for English for science students. In D. Nunan & J.C. Richards (Eds.), *Language learning beyond the classroom* (pp. 212–222). Routledge.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22, 27–57. <https://doi.org/10.1093/applin/22.1.27>

Santos, N.M.B., Alarcon, M.M.H., & Pablo, I.M. (2016). Fillers and the development of oral strategic competence in foreign language learnings. *Porta Linguarum*, 25, 191–201.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532. <https://doi.org/10.1093/applin/amp047>

Skehan, P. (2014). The context for researching a processing perspective on task performance. In P. Skehan (Ed.), *Planning and Task Performance in a Second Language* (pp. 111–141). John Benjamins.

Spring, R. (2020). Using multimedia tools to objectively rate the pronunciation of L1 Japanese EFL learners. *Teaching English through Multimedia: ATEM Journal*, 25, 113-124.

Stoller, F. (2006). Establishing a theoretical foundation for project-based learning in second and foreign language contexts. In G.H. Beckett & P.C. Miller (Eds.), *Project-Based Second and Foreign Language Education: Past, Present and Future* (pp.19–40). Information Age.

Suter, R. (2006). Predictors of pronunciation accuracy in second language learning. *Language Learning*, 26(2), 233–253. <https://doi.org/10.1111/j.1467-1770.1976.tb00275.x>

Tamin, S.R., & Grant, M.M. (2013). Definitions and uses: Case study of teachers implanting project-based learning. *Interdisciplinary Journal of Problem-Based Learning*, 7(2), 72–101. <https://doi.org/10.7771/1541-5015.1323>

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp.239–273). John Benjamins. <https://doi.org/10.1075/llt.11.15tav>

Torres, A.M.V., & Rodriguez, L.F.G. (2017). Increasing EFL learners' oral production at a public school through project-based learning. *Profiles: Issues in Teachers' Professional Development*, 19(2), 57–71. <https://doi.org/10.15446/profile.v19n2.59889>

Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects of fluency, complex, and accuracy. *TESOL Quarterly*, 50(2), 369–393. <https://doi.org/10.1002/tesq.232>

Vercellotti, M.L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics*, 38(1), 90–111. <https://doi.org/10.1093/applin/amv002>

Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. Continuum

Copyright rests with authors. Please cite TESL-EJ appropriately.