

Student Videos: A Replication Study Into Tech-Assisted TBLT

Martyn McGettigan

Hiroshima City University

Reference Data:

McGettigan, M., (2026). Student videos: A replication study into tech-assisted TBLT. In B. Lacy, M. Swanson, & H. Yoder (Eds.), *Language S: Learning, Teaching, Assessing – JALT 50 Years – Challenges and Perspectives*. JALT. <https://doi.org/10.37546/JALTPCP2025-07>

This paper describes a replication of a previous mixed-methods study (Chen, 2021) which looked at the effect of a group project involving the production of English-language videos on university students' speaking abilities. The original study, conducted in a Taiwanese university context, found a large effect on students' speaking scores and an overall positive perception of the video-making project on the part of the students. This replication aimed to reproduce the study in a Japanese context, comparing the results with those of the original study. The replication also showed a large effect, though the change in actual scores was smaller than in the original. The replication then extended the original study by examining skills beyond speaking and found a medium-sized effect on reading. The qualitative results were also found to be similar in both studies, with the students' impressions of the project being largely positive.

本論文は、大学生が英語による動画制作を行うグループプロジェクトを通じて、英語のスピーキング能力がどのように向上するかを検証した Chen (2021) の混合研究法による先行研究を日本の大学で再現したものである。台湾における元の研究では、スピーキングの得点に大きな効果がみられ、学生の評価も肯定的であった。本研究における再現でも、スピーキングに大きな効果が確認されたが、得点上昇幅は台湾での研究より小さかった。いっぽう、リーディングには中程度の効果が見られた。質的調査の結果は両研究で類似しており、学生は動画制作プロジェクトを有益な学習活動として受け止めていた。

The idea of replicating a previous study occurred to me as a novice researcher. My intention was to learn as much as possible about the process of conducting second language acquisition research while making a small contribution to the field. Task-based language teaching (TBLT) and the use of technology in the language classroom are two

areas of particular interest to me. Accordingly, I chose to replicate a study that combined both elements (Chen, 2021).

Chen conducted her study as action research in a compulsory English class at a Taiwanese university. Her 25 participants were divided into groups and asked to produce seven 5-minute videos in English on various pre-determined topics over the course of a single semester. She collected quantitative data through pre- and post-tests of the students' speaking abilities, as well as qualitative data through her own observations, students' written reflections, and their responses to a final questionnaire. The results of both types of data were impressive, showing a strong effect on students' speaking skills and an overwhelmingly positive reaction to the project from the participants.

The research questions for this replication are almost identical to those in the original study, with only the language background of the participants changed (Chen, 2021, p. 334):

- RQ1. What effect does technology-mediated task-based learning have on the English-speaking proficiency of a class of Japanese language-background undergraduate students?
- RQ2. How favourably does a class of Japanese language-background undergraduate students view the integration of technology-mediated task-based learning into their compulsory English classes?

Literature Review

The background of this replication study can best be described by beginning with the rationale provided in the original study. Chen (2021) begins by citing a number of references that illustrate the increasing role of technology in language teaching in general (González-Lloret, 2003; Reinders & White, 2010; Salaberry, 2000), before turning to sources that more specifically position such technology within a TBLT context (Thomas & Reinders, 2013). In particular, she adopts the term "technology-mediated TBLT" (TMTBLT), first used by González-Lloret and Ortega (2014, p. 1), defined as a

framework “whereby tasks and technology are genuinely and productively integrated in the curriculum according to learning-by-doing philosophies of language pedagogy, new language education needs, and digital technology realities.” Chen does not, however, define TBLT itself. For the purposes of this study, TBLT is defined as any language-learning activity that is:

- primarily meaning-focused,
- contains an information gap that must be bridged communicatively,
- has a clearly defined end goal, and
- requires learners to make use of their own resources (rather than being given, for example, set forms to use) (Ellis & Shintani, 2013).

Chen (2021) provides an extensive list of citations to support the effectiveness of TMTBLT, both in theory and in the results of actual studies. A more recent meta-analysis by Bryfonski et al. (2025) found a small-to-moderate effect for TMTBLT across 17 studies. A similar systematic review and meta-analysis by Bhandari et al. (2025) reported comparable results but also noted that TMTBLT remains a relatively new area and that the scope of research, both quantitative and qualitative, is still quite limited.

Chen (2021) further supports the rationale for her study by providing an overview of the literature on the effectiveness of video-making projects for language acquisition. She focuses particularly on the meaning-focused and collaborative skills these projects help develop. More recent scholarship has continued to support the effectiveness of such projects. Hsu (2024), for example, found a similar video-making project effective in boosting both proficiency and confidence, especially among lower-proficiency students. Other comparable studies (Tran et al., 2023; Hwang et al., 2024) have reported similar results. Khawa and Rizkiyah (2022) implemented another similar project explicitly positioned as a form of TBLT and found improvement in students’ speaking skills. None of these studies, however, found an effect on speaking nearly as large as that found by Chen (2021).

Finally, since this is a replication study, the rationale for undertaking it in this area of linguistics—and for choosing Chen’s (2021) study in particular—must be addressed. The ongoing proliferation of literature in this area, as detailed above, demonstrates that the topic remains relevant. Chen’s study also fulfils many of the criteria for suitability for replication outlined by Porte and McManus (2019). The original study is relatively recent, produced arguably exceptional results (a very strong effect on speaking scores), and was possible to reproduce closely with the available resources. Furthermore, as Polio (2012) argues, replication studies can be particularly valuable when they exactly replicate the methodology or instruments used in the original study while changing another variable

(such as the language background of the participants). This study aimed to replicate Chen’s (2021) original in exactly that way, by examining a different group (Japanese rather than Taiwanese language-background university students) but using the same methodology and instruments as far as possible.

Methodology

This section describes the methodology followed in Chen (2021) and in this replication, noting any cases where it was necessary to diverge from the original. The original study was conducted with a compulsory English class of 25 Taiwanese language-background university students, aged 19–21, all majoring in Applied English. The participants in this replication totaled 17 students, all from a Japanese language background. They were also enrolled in a compulsory English class and fell within the same age range but were Information Science majors. Despite the difference in major, this was the most advanced English class available for the replication. The class was studying a textbook aimed at CEFR B1-level students, although no formal level test result was available until the pre-test was administered at the start of this replication.

In the original study (Chen, 2021), students were divided into five groups and asked to produce seven 5-minute videos in English over one semester. They were given topics, but otherwise had complete freedom in terms of content. The topics were: storytelling, food, travel, movies, horror stories, “bubbling,” and a topic of their choice. In the replication, there were four groups, and the same topics were used except for “bubbling.” It was unclear what this topic referred to and attempts to contact the original researcher were unsuccessful. Therefore, the topic was replaced with “AI.” Additionally, in the replication, the time frame was extended to two semesters. This was partly due to the need to cover compulsory textbook material in the class (no such requirement was mentioned in the original study) and partly in response to one of the student suggestions in Chen (2021) for more time per video.

In both studies, quantitative data were collected through pre- and post-tests. The original study (Chen, 2021) used the Test of Spoken English (TSE) provided by the Educational Testing Service, a speaking-only test with scores on a 0–30 scale. However, early in this replication, it was discovered that the TSE is no longer available. A replacement was found in the form of the Versant English Placement Test by Pearson. This test provides individual scores for speaking, listening, reading, and writing on a 10–90 General Scale of English (GSE). Qualitative data in both studies were collected through teacher observations and reflective comments written by the student

participants. The teacher observations followed a set form, in which the teacher recorded weekly how well the groups were working together, using English, etc., and each group received an overall positive or negative result for each point at the end of the project. The students were free to write about whatever they wished in their reflections.

In the first class, the student participants were given a clear explanation of the research project. It was made clear to the students how the data gathered from the course activities would be used. Students were given the option to withdraw consent for their data to be used, or for the videos they produced to be shown to third parties, without any effect on their grade. No student decided to opt out, although one student ultimately dropped out of the course mid-way through and all data related to that student was excluded from the final study.

Results

Quantitative Results

This section presents the quantitative results from both the original study and the replication, followed by additional results unique to the replication. Table 1 shows the mean pre- and post-speaking test scores side by side. As can be seen, the actual improvement in speaking scores in the replication was not as large as in the original study, although a large effect size was still found. To corroborate the effect size, Cohen's d was calculated for the replication, yielding a value of 1.02, which indicates a large effect.

Table 1

Mean Pre- and Post-Test Speaking Scores for the Original and Replication Studies

Original Study		Replication Study	
TSE Score (0-30)		Versant Speaking Score (10-90)	
Pre-test (M)	8.3	Pre-test (M)	29.65
Post-test (M)	12.52	Post-test (M)	32.18
	$t = 16.04$		$t = 4.21$
	$p = >0.001$		$p = >0.001$
			Cohen's $d = 1.021$

Since the Versant Placement Test is a four-skills test, the pre- and post-test results were also examined for listening, reading, and writing. No effect was found for listening or writing. For reading, however, a small but significant effect was observed. Upon closer inspection, one group of students was identified as an outlier in terms of reading scores. This group had been noted for the creativity of their videos and their avoidance of simply reading scripts over still visuals (a common approach in the other three groups). When this outlier group was removed, the effect on reading became large. It should be noted, however, that this reduced the sample size to $N = 12$. Table 2 shows the mean reading scores from the pre- and post-tests, both with and without the outlier group.

Table 2

Mean Pre- and Post-Test Reading Scores With and Without the Outlier Group

All participants ($N = 17$)		Without Outliers ($N = 12$)	
Versant Reading Score (10-90)		Versant Reading Score (10-90)	
Pre-test (M)	34.88	Pre-test (M)	35.83
Post-test (M)	37.47	Post-test (M)	42.08
	$t = 1.325$		$t = 3.626$
	$p = 0.006$		$p = 0.005$
	Cohen's $d = 0.321$		Cohen's $d = 1.047$

Qualitative Results

As noted above, qualitative data were collected through three methods: teacher observations, weekly student reflections, and a final questionnaire. Teacher observations were conducted weekly but recorded as final binary results (check mark or no check mark). Tables 3 and 4 present the final results for these observations in both the original study and the replication, respectively. The results were quite similar. The most notable difference was that no groups in the replication received a check mark for "Speaks English more frequently." Efforts to encourage the use of English during planning and editing phases were rarely successful.

Table 3
Original Study Teacher Observations

Student performances	1	2	3	4	5
Actively participates in group discussions	✓	✓	✓	✓	✓
Ask for help when difficulties were encountered		✓	✓		✓
Shows enthusiasm toward the video making tasks	✓	✓	✓	✓	✓
Shows creativity	✓	✓	✓	✓	✓
Speaks English more frequently		✓	✓		✓
Cooperates well with the other members of the group	✓	✓	✓	✓	✓
Tries to speak English more frequently with better grammar and pronunciation in the videos made		✓	✓		✓
Pays attention to other group's video performance	✓	✓	✓	✓	✓
Gives fair video evaluations		✓	✓	✓	✓

(Chen, 2021 p. 339)

Table 4
Replication Study Teacher Observations

Student performances	1	2	3	4
Actively participates in group discussions	✓	✓	✓	✓
Ask for help when difficulties were encountered				
Shows enthusiasm toward the video making tasks	✓	✓	✓	✓
Shows creativity	✓	✓		✓
Speaks English more frequently				
Cooperates well with the other members of the group	✓	✓	✓	✓
Tries to speak English more frequently with better grammar and pronunciation in the videos made	✓	✓		
Pays attention to other group's video performance	✓	✓	✓	✓
Gives fair video evaluations	✓	✓	✓	✓

The second source of qualitative data was weekly student reflections. In both the original study and the replication, positive comments significantly outnumbered negative ones. The comments were coded and analysed in the same manner in both studies, based on matrices adopted from Miles and Huberman (1994). Tables 5 and 6 present the positive reflections from each study, while Tables 7 and 8 present the negative ones.

Table 5
Positive Comments from the Original Study

N	Facebook posts	Code
53	I really appreciate the efforts of my teammates.	Collaboration
45	Making videos are an interesting way to practice speaking in English.	Perceptions
39	The video tasks have brought everyone in my group closer.	Collaboration
35	I learn something new every time I watch other group's videos.	Learning achievement
31	My teammates are very excited to plan and film the next videos.	Collaboration
28	I like this course and I enjoy it.	Perceptions
28	I really appreciate the opportunities to speak in English more frequently to actually improve my speaking ability.	Perceptions
22	I learned to recognize my groupmate's strengths.	Collaboration
18	Every group's videos got better and improved toward the end of the semester.	Learning achievement

Note. N represent is the number of times the same post was mentioned. A total of 418 statements were posted on Facebook over the 18-week period. (Chen, 2021 p. 339)

Table 6

Positive Comments from the Replication Study

N	Post	Code
84	We had good discussions	Collaboration
44	Everyone's videos were well-made and enjoyable	Perception
15	We are making good progress in producing the video	Collaboration
8	Creating our video was challenging but fun	Perception
4	I learnt a lot from watching the other groups' videos	Learning
1	It was easy to understand what we have to do.	Learning
1	I'm looking forward to seeing the other groups' videos	Perception
1	I thought our video was the best.	Perception
1	I am trying to improve my planning	Learning

Table 7

Negative Comments from the Original Study

N	Facebook posts	Code
23	Consensus among the group was hard to achieve.	Collaboration
19	Group work was time consuming.	Collaboration
16	It was more difficult and challenging than the traditional class.	Learning achievement
15	Two weeks to produce a video seems to be insufficient.	Perceptions
8	Our team had spent so much time editing the videos and felt that we did not receive the score we deserved.	Perceptions
7	It would be better if the instructor gives us clear instructions on how we could get higher scores.	Learning achievement
6	I do not like the evaluation method. The teacher should be the only one to evaluate. Peer evaluation should be eliminated because some students do not give fair scores.	Perceptions

(Chen, 2021 p. 340)

Table 8

Negative Comments from the Replication Study

N	Post	Code
11	It was difficult to come up with ideas for videos	Perception
5	This particular topic is difficult for me	Perception
4	Making videos is too difficult	Perception
4	It was difficult to make progress because of lack of cooperation	Collaboration
4	The videos are getting a bit stale	Perception
3	We have a lot of disagreements	Collaboration
2	Editing was very difficult	Perception
1	I wish we had worked more efficiently in order to be ready for the video deadline	Perception

The final method of qualitative data collection was a questionnaire administered to students at the end of the project. Chen's (2021, pp. 341–343) table summarising her results is too long to reproduce here, but in summary, her students found the project challenging but enjoyable, believed it helped their English-speaking skills, and felt it had developed their collaborative skills, though they suggested more time should have been allocated for each video. A summary of the results from the replication study is provided in Appendix A. Overall, the results were very similar: students felt the project had benefited their speaking skills and was enjoyable. The chief difficulty mentioned was editing the videos, which many students felt they lacked the skills to complete effectively in the time given.

Discussion

The results of the replication study were, overall, quite similar to those of the original and thus serve to strengthen the original findings. Both studies indicated that the project had a large effect on participants' speaking skills and was received positively by the students. However, several key points must be discussed, along with certain limitations present in both studies, before considering pedagogical applications and directions for future research.

The key difference in the quantitative results lies in the magnitude of improvement in speaking proficiency as measured by the pre- and post-tests. Although both studies found large effect sizes, the actual increase in mean scores was more substantial in Chen's (2021) original study. The most obvious explanation can be found in the qualitative teacher observation data. As noted earlier, no groups in the replication received a check mark for "Speaks English more frequently." Efforts to encourage the use of English during planning and editing phases were rarely successful. Chen (2021, p. 335) refers several times to her students "communicating in English" during video production. However, it must be borne in mind that different measures were used (of necessity). It is quite possible that the difference in improvement resulted from variations in the accuracy of the two tests or differing conceptions of speaking proficiency.

A second point of interest is the effect found on reading proficiency in the replication. As noted, this cannot be compared directly with the original study, as it was a new addition resulting from the change of test. Upon closer inspection, the increase in effect size after removing one group of students suggests that the effect was related to the students' approach to the task. Most groups produced videos consisting largely of students reading scripts over still visuals—a practice I actively discouraged. I assumed that a more dramatic approach would be both more enjoyable for the students and more likely to benefit their language skills, particularly speaking, which was the study's main focus. Admittedly, however, this assumption was not based on empirical evidence. In fact, the students who relied heavily on scripts showed no less improvement in speaking skills and appeared to make greater advances in reading.

Chen (2021) acknowledged certain limitations of her own project, particularly the small sample size and the impossibility of isolating participation in the video project as the sole source of improvement. Such limitations are common in action research (Stratton, 2025). In terms of potential interference from other sources of learning, the replication participants were also studying a textbook as part of the same class (an unavoidable limitation given the resources available). However, the lack of an effect on listening and writing provides some support for the conclusion that the observed effect on the targeted speaking skill was genuine. The effect on reading may similarly reflect the extent to which students engaged with reading during the project.

Conclusion

Taken together, the quantitative and qualitative evidence from the two studies suggests a strong case that this type of video project is an effective and enjoyable way to implement speaking practice in a university English class. It could also be adapted to focus more on the reading effect observed here—for example, a project involving students creating scripts with AI assistance and then producing videos based on those scripts could benefit reading skills while ensuring engagement with AI-generated content. For future research, similar studies with larger samples would provide stronger corroboration (or refutation) of the results of these two small-scale studies. The design could also be modified to focus on reading in addition to speaking. Finally, the process of replicating a study, though challenging in many ways, can be highly instructive and rewarding. It is hoped that more novice researchers will undertake such replication studies in the future.

Funding

This study was funded by a JALT Research Grant.

Bio Data

Originally from the UK, **Martyn McGettigan** has an MA in Applied Linguistics and ELT and has been teaching English in Hiroshima since 2017. He currently teaches at four different universities in the city. His research interests include task-based language teaching and corrective feedback.

References

- Bhandari, L. P., Dahal, N., Awasthi, J. R., & Dhungana, S. (2025). Technology-mediated task-based language teaching: A systematic review. *Cogent Education*, 12(1), Article 2560051. <https://doi.org/10.1080/2331186X.2025.2560051>
- Bryfonski, L., Ziegler, N., & Montee, M. (2025). Technology-mediated task-based language teaching: A meta-analysis. *CALICO Journal*, 42(3), 472–501. <https://doi.org/10.3138/calico-2024-0029>
- Chen, K. T. (2021). The effects of technology-mediated TBLT on enhancing the speaking abilities of university students in a collaborative EFL learning environment. *Applied Linguistics Review*, 12(2), 331–352. <https://doi.org/10.1515/applirev-2018-0126>
- Ellis, R., & Shintani, N. (2013). *Exploring language pedagogy through second language acquisition research* (1st ed.). Routledge. <https://doi.org/10.4324/9780203796580>

- González-Lloret, M. (2003). Designing task-based CALL to promote interaction: En busca de Esmeraldas. *Language Learning & Technology*, 7(1), 86–104. <https://doi.org/10.64152/10125/25189>
- González-Lloret, M., & Ortega, L. (2014). *Technology-mediated TBLT: Researching technology and tasks*. John Benjamins Publishing Company.
- Hsu, H. (2024). Evaluating the impact of video-making projects on English for specific purposes learning: A case study. *International Journal of Language and Education Research*, 6(1), 1–23. <https://doi.org/10.29329/ijler.2024.661.1>
- Hwang, W. Y., Luthfi, M. I., & Liu, Y. F. (2026). AI-enhanced video drama-making for improving writing and speaking skills of students learning English as a foreign language. *Innovation in Language Learning and Teaching*, 20(3), 462–479. <https://doi.org/10.1080/17501229.2024.2437655>
- Khawa, D., & Rizkiyah, F. (2022). The implementation of task-based learning strategy by making video introduction to improve students' speaking skills. *Darussalam English Journal*, 2(2), 199–220. <https://doi.org/10.30739/dej.v2i2.1778>
- Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. SAGE Publications.
- Polio, C. (2012). Replication in published applied linguistics research: An historical perspective. In G. Porte (Ed.), *Replication in applied linguistics: A practical guide* (pp. 47–91). Cambridge University Press.
- Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.
- Reinders, H., & White, C. (2010). The theory and practice of technology in materials development and task design. In N. Harwood (Ed.), *Materials in ELT: Theory and practice* (pp. 58–80). Cambridge University Press.)
- Salaberry, M. R. (2000). Pedagogical design of computer mediated communication tasks: Learning objectives and technological capabilities. *The Modern Language Journal*, 84(1), 28–37. <https://doi.org/10.1111/0026-7902.00050>
- Stratton, M. (2025). Small samples in instructed second language acquisition research. *The Modern Language Journal*, 105(1), 187–193. <https://doi.org/10.1111/modl.12700>
- Thomas, M., & Reinders, H. (2013). *Task-based language learning and teaching with technology*. Bloomsbury Publishing.
- Tran, T. T. T., Nguyen, T. N. V., & Nguyen, T. T. H. (2023). Student-produced videos: Why not? In *Proceedings of the 19th International Conference of the Asia Association of Computer-Assisted Language Learning (AsiaCALL 2022)* (pp. 93–102). https://doi.org/10.2991/978-2-38476-042-8_9

Appendix A

Results of the End of Course Student Survey

Open-ended survey questions

Q1 What was your favourite topic?

AI

(Similar: 1)

Movies

(Similar: 2)

Story Telling

(Similar: 1)

Horror

Country

Q2 What was the biggest difficulty you encountered?

Editing was difficult because it was my first time. It was difficult to edit to make it interesting.

(Similar: 4)

Making a script in English

Not having enough time

(Similar: 1)

Coming up with an idea as a group

(Similar: 1)

English pronunciation

Q3 To what extent did the tasks help your English speaking?

It helped me a lot because I spoke a lot and had to make the video interesting in English.

(Similar: 3)

I gained confidence

It helped a lot because I had to use unfamiliar words

Open-ended survey questions

It helped my pronunciation

It helped a little

(Similar: 2)

Q4 What was your perception of these tasks?

It was fun, especially watching other groups' videos.

(Similar: 2)

It was fun and I learned both English and editing skills.

I felt I wanted to improve my English

It was a bit hard but interesting

It was quite difficult

It was a good experience because I don't usually speak English

Making videos was fun but editing late at night was difficult.

I think it was very interesting and might help improve my English.

I think it was very effective for understanding daily English.

Q5 Is there anything you would suggest improving in this course?

Nothing in particular.

(Similar: 2)

The topics should be changed a bit. Some were difficult.

(Similar: 1)

More reflection time during class

Reduce the number of videos and focus on quality.

How to score.

The fact that not everyone has editing skills should be taken into account. Also, there should be a designated place for filming where people feel comfortable.

It would be good to change group members sometimes.

(Similar: 1)