

Digital Storytelling as a Space for Speaking Preparation among Lower-Motivation EFL Learners

Endang Sulistianingsih ^{1*}

¹Universitas Pancasakti Tegal, Tegal, Central Java – Indonesia

Article Information

Received: June 04, 2026

Revised: October 20, 2026

Accepted: October 28, 2026

Published online: October 28, 2026

Abstract

Reluctant speakers in EFL classrooms often have few opportunities to practice oral communication. This study aimed to describe speaking-score patterns between pre- and post-intervention assessments and to explore how lower-motivation learners experienced digital storytelling (DST) as a context for speaking preparation and participation. First-semester undergraduates at a private university in Central Java, Indonesia, were purposively selected based on a motivation questionnaire and participated in an eight-week DST intervention. Data were collected through speaking assessments, classroom observations, reflective journals, interviews, and student-produced digital stories. The findings showed that post-intervention means were higher across fluency, pronunciation, vocabulary, grammar, and comprehension, with the largest gains in vocabulary and comprehension. Wilcoxon signed-rank tests indicated significant differences across all dimensions. No learner scored lower, although several showed no change in some dimensions. Qualitative evidence portrayed DST as a staged process in which rehearsal, self-monitoring, revision, and peer and teacher support were intertwined with the technical demands of editing and recording. Without a comparison group, these findings describe observed patterns rather than causal effects. Teachers may support reluctant speakers by staging the DST workflow, providing technical scaffolding, and assessing individual oral performance separately from group products.

Keywords: digital storytelling; EFL speaking; lower-motivation learners; reluctant speakers; speaking performance; learner experiences; higher education

Introduction

Speaking in an additional language can be difficult when learners have few opportunities to use English beyond the classroom and are reluctant to speak in front of peers or an instructor. In EFL settings, classroom speaking is often experienced as a high-stakes performance

in which errors become immediately visible to others. Such conditions can restrict oral practice even among learners who have studied English for several years. For this reason, research on speaking instruction has attended not only to linguistic performance but also to the classroom conditions under which learners are willing to

***Corresponding Author:** Endang Sulistianingsih
(endang.sulistia@gmail.com) Jalan Halmahera KM
1, Mintaragen, East Tegal, Tegal, Central Java,
Indonesia, 52121

©2026 by the Authors, published by Vision: Journal for
Language and Foreign Language Learning
<https://journal.walisongo.ac.id/index.php/vision>

participate (Cendra, 2022; Hawanti, 2014; Révész et al., 2016).

Digital storytelling (DST) has received growing attention as one way of creating such conditions. Prior studies have associated DST with gains in speaking, vocabulary learning, self-regulation, learner engagement, and motivation (Bai & Xian, 2024; Isaacs et al., 2024; Mohammadi Khomjani, 2020; Shen et al., 2024; Tatli et al., 2022; Yasmin et al., 2025). Less is known, however, about how learners who are reluctant to speak experience the preparation and sharing of digital stories, and how those experiences relate to their measured speaking performance. The point is not that low-motivation learners or Indonesian EFL settings have been overlooked altogether. Related Indonesian work, for example, has examined pre-service teachers' perceptions of digital storytelling through the Technology Acceptance Model (Sulistianingsih & Taufiqulloh, 2025). Studies of this kind tend to rely on a single type of evidence, and each type has limits. An overall score can show a difference between two occasions. However, it cannot reveal how learners prepared or what they found difficult. Reflective accounts can illuminate experience but cannot establish the size or cause of a performance change. Accounts that bring both kinds of evidence to bear on a lower-motivation subgroup remain scarce.

DST combines spoken narration with selected visual and audio resources through a sequence of composing, recording, revising, and sharing. Its potential value for hesitant speakers lies less in the addition of technology than in the opportunity it gives learners to prepare and revise their oral language before an audience hears the final narration. Repeated recording

may help learners notice unclear wording or pronunciation, scripting and story sequencing may support the organization of ideas, and peer work may offer assistance during production (Fu et al., 2021; Liang & Hwang, 2023; Sulistianingsih & Aflahatun, 2026; Wei et al., 2025; Yu & Wang, 2025). These processes are plausible, but they should not be treated as demonstrated mechanisms unless the data collected in a particular study support them.

Two theoretical perspectives inform the interpretation of such processes. Self-Determination Theory identifies autonomy, competence, and relatedness as motivational needs (Ryan & Deci, 2020), while Krashen's Affective Filter Hypothesis offers a lens for examining how anxiety may constrain participation. Neither construct was measured with a validated scale in this study. Both, therefore, serve as interpretive lenses rather than as tested explanatory variables, and the analysis does not claim to test Self-Determination Theory, measure an affective filter, or establish causal mechanisms.

Against this background, the present study focuses on learners classified as part of the lower-motivation group by a questionnaire administered before the intervention. It aims to describe the patterns observed in speaking-assessment scores between the pre- and post-DST assessments among participating learners, and to explore how these learners described and experienced DST as a context for speaking preparation and participation. By reading descriptive score patterns alongside reflective and observational evidence, the study offers a deliberately bounded, context-specific account of how reluctant speakers engaged with DST,

one that distinguishes what was observed from what might explain it.

Method

Research Design

The study used a one-group pre-test/post-test design with a qualitative component. The quantitative strand consisted of speaking assessment scores collected before and after an eight-week (16-session) digital storytelling (DST) intervention, and was used to describe score patterns. The qualitative strand drew on classroom observations, reflective journals, semi-structured interviews, and the digital stories students produced, and it was used to explore learners' experiences of preparation, rehearsal, collaboration, and production. The two strands were combined at the interpretation stage by examining whether recurring accounts matched the direction of the score patterns. This integration was intended to provide a contextualized reading of the results rather than to show that reported experiences caused the changes in scores. Because the design had no comparison group, it could not separate the contribution of DST from the rehearsal, teacher feedback, peer collaboration, and task familiarity that accompanied it.

Instruments

The Learners' motivation was screened with the Motivation for English Speaking Performance Questionnaire (MESPQ). This is a 21-item, closed-ended questionnaire covering four dimensions: speaking activities (9 items), peer and teacher evaluation (4 items), task attitude (5 items), and rewards (3 items). Experts reviewed the questionnaire, and item validity was examined with Pearson correlations.

Cronbach's alpha values were .937 for speaking activities, .886 for peer and teacher evaluation, .933 for task attitude, and .852 for rewards.

Speaking performance was assessed with a rubric informed by Brown and Abeywickrama (2019) and O'Sullivan (2024). The rubric covers five weighted dimensions: comprehension (30%), pronunciation (20%), fluency (20%), grammar (15%), and vocabulary (15%). Each dimension is rated on five levels: poor (1), needs improvement (2), fair (3), good (4), and excellent (5). Expert judgment produced an Aiken's V of .86, and inter-rater agreement produced a Cohen's kappa of .81. The term "comprehension" follows the rubric and is used consistently throughout this article.

The qualitative instruments were reflective-journal prompts, a structured observation sheet, a semi-structured interview guide, and the students' completed videos. The journal prompts asked learners to reflect on their preparation, speaking performance, confidence, anxiety, motivation, and challenges, as well as on teacher support, peer relationships, and their own roles in completing the project. The observation sheet recorded participation, engagement, grouping, task involvement, and classroom interaction. The interview guide asked about learners' experiences of scripting, recording, revision, and presentation, and about how they felt their speaking had developed. The completed videos served as contextual artifacts of the group projects rather than as substitutes for individual speaking scores.

These sources provide different kinds of evidence. A learner's reported confidence is not the same as a fluency rating, and a polished video does not necessarily reflect spontaneous speaking ability. The analysis therefore kept

rated performance separate from participants' perceptions and from the researcher's observations.

Participants

The participants were 19 first-semester undergraduates (13 women and 6 men), aged 20–22. They were enrolled in an ICT for English Language Learning course at a private university in Central Java, Indonesia, and the institution classified their English proficiency as CEFR B1. Thirty-four students were screened with the MESPOQ. Of these, 19 were classified as having lower motivation and were purposively selected. In contrast, the remaining 15 were classified as having higher motivation and were not included in this analysis. The study records do not specify the numerical cut-off used for this classification. The label therefore reflects the study's questionnaire-based grouping rather than a diagnosis or a validated threshold.

All participants were informed about the study, gave their consent, and are referred to by pseudonyms. The researcher was also the course instructor and data collector. This dual role may have influenced participation and the accounts participants gave, and it was taken into account when interpreting the qualitative evidence.

Data Collection

Data were collected during the eight-week intervention, which ran across 16 sessions within the regular course. The speaking assessment was administered before the intervention began and again after it ended. The intervention opened with an introduction to DST, illustrated with examples of multimodal stories. Students then selected their topics, wrote scripts, and prepared storyboards. They went on to rehearse

and record their narration, choose visual and audio materials, and edit their videos. The sequence closed with feedback, revision, and the sharing of completed stories. Students worked together on story production and could rehearse and record their narration several times. The instructor gave feedback on pronunciation, clarity, grammar, vocabulary, and organization, and peer consultation and reflection were built into the process. The records do not consistently state how long each session lasted or how students divided production tasks, so these details are not assumed.

Qualitative data were collected during the intervention. Observation notes were written during each session to capture students' participation and engagement. Reflective journals were completed at the beginning, middle, and end of the intervention. Semi-structured interviews explored learners' perceptions of DST and their engagement in speaking, and the completed videos were collected at the end of the project.

Data Analysis

The quantitative analysis summarized the pre- and post-intervention means and their differences for each of the five speaking dimensions. Paired differences in each dimension were tested with the Wilcoxon signed-rank test, which is appropriate for small sample sizes and ordinal rating scales. Rank summaries showed how many learners obtained higher, lower, or tied scores. The overall mean was calculated as the unweighted average of the five dimension means and is reported only descriptively. SPSS values displayed as .000 are reported as $p < .001$. Effect sizes and confidence intervals are not reported,

and statistical significance is not taken as evidence that DST alone caused the observed differences.

The qualitative data were analyzed thematically following Braun and Clarke (2012). The analysis began with repeated reading of the data, followed by initial coding. Related codes were then grouped into candidate themes, and these themes were reviewed against the data before the recurring patterns across journals, interviews, and observations were interpreted. The themes are reported as interpretive patterns rather than prevalence estimates, because the unit of counting was not defined consistently. Contrasting experiences, including technical difficulty and the limits of group work, were retained so that DST was not presented as uniformly supportive.

Trustworthiness was addressed by reading learners' reported experiences alongside classroom observations and speaking assessment results. Because these sources serve complementary purposes, they were combined for interpretive analysis. Points of agreement and difference were considered without assuming that reported experiences caused changes in scores. The researcher's dual role, the small purposive sample, and the absence of a comparison group were all considered when judging the strength and transferability of the findings.

Findings and Discussion

Findings

The findings are organized around the two research objectives. The first section describes patterns in speaking assessment scores between the pre- and post-DST assessments among the

19 learners classified in the lower-motivation group. The second section examines how these learners described and experienced DST as a context for speaking preparation and participation, drawing on their reflective journals and on classroom observations.

Patterns in Speaking-Assessment Scores Between the Pre- and Post-DST Assessments

Table 1 presents the pre- and post-assessment mean scores of the 19 learners across the five dimensions of the speaking rubric, together with the results of dimension-specific Wilcoxon signed-rank tests. The mean differences are group-level summaries and should not be read as evidence that every learner improved.

Post-assessment means were higher than pre-assessment means in all five dimensions. Vocabulary showed the largest difference, rising from 3.10 to 4.00 (+0.90). Comprehension came next, rising from 3.52 to 4.36 (+0.84), followed by grammar, which rose from 3.15 to 3.94 (+0.79). Fluency and pronunciation showed slightly smaller differences, rising from 3.15 to 3.84 (+0.69) and from 3.11 to 3.79 (+0.68), respectively. Averaged across the five dimensions, the overall mean rose from 3.21 to 3.99 (+0.78).

The rank summaries show how these group-level differences were spread across individual learners. No dimension recorded a negative rank, so no learner received a lower post-assessment score on any dimension. Vocabulary had the most learners with higher scores (17, with 2 ties), followed by grammar (15, with 4 ties). Pronunciation, fluency, and comprehension each had 13 learners with

higher scores and 6 with tied scores. Nearly a third of the group therefore showed no change

in these three dimensions, which the mean differences alone would not reveal.

Table 1

Pre- and Post-Assessment Speaking Scores and Wilcoxon Signed-Rank Results

Speaking dimension	Pre mean	Post mean	Difference	Wilcoxon Z	P
Fluency	3.15	3.84	+0.69	-3.606	< .001
Pronunciation	3.11	3.79	+0.68	-3.606	< .001
Vocabulary	3.10	4.00	+0.90	-4.123	< .001
Grammar	3.15	3.94	+0.79	-3.873	< .001
Comprehension	3.52	4.36	+0.84	-3.358	< .001
Overall Mean*	3.21	3.99	+0.78	—	—

Note. "Comprehension" follows the term used in the speaking rubric. The overall mean is the unweighted average of the five dimension means and is included as a descriptive summary only; no inferential test was conducted on this aggregate.

The Wilcoxon signed-rank tests showed statistically significant pre/post differences in all five dimensions: pronunciation ($Z = -3.606$, $p < .001$), grammar ($Z = -3.873$, $p < .001$), vocabulary ($Z = -4.123$, $p < .001$), fluency ($Z = -3.606$, $p < .001$), and comprehension ($Z = -3.358$, $p = .001$). These results agree with the direction of the mean differences, but they do not show that DST alone produced the change. Repeated practice, teacher feedback, peer collaboration, and growing familiarity with the task all took place alongside digital production, and the one-group design has no comparison condition. Because effect sizes and confidence intervals are not reported, the size and precision of these results should also be read with caution. The score data therefore show what changed between the two assessments. They do not show how learners prepared or what they found difficult, which the second objective addresses.

Learners' Experiences of DST as a Context for Speaking Preparation and Participation

Learners' experiences were explored through their reflective journals and through classroom observations. Five recurring themes were identified in these sources: the first concerns speaking confidence and communicative pressure, and the second concerns practice and self-monitoring. The third and fourth relate to the social side of the activity: peer and teacher support, and participation and classroom interaction. The fifth covers the difficulties learners met in digital production. These themes are interpretive patterns, so they are neither mutually exclusive nor ranked by frequency of occurrence.

The first theme concerns how learners experienced the pressure of speaking. The reflective journals invited learners to describe their anxiety, confidence, and motivation after the intervention, and some of their accounts presented preparation before sharing as manageable. These learners linked this sense of manageability to the staged nature of DST, which let them plan what to say and prepare before sharing a finished story. This finding

reflects reported experience. It is not a direct measurement of anxiety, and it does not show that anxiety disappeared for all participants.

The staged structure that some learners found manageable also gave them repeated chances to practice. Script preparation, rehearsal, recording, and revision gave learners repeated opportunities to prepare their narration and to notice the parts that needed revising. The journals invited learners to describe their speaking progress and challenges, and the observations recorded how they engaged with the task. Read together, these sources suggest that repeated preparation and the chance to revise were prominent features of the learning process. They do not, however, confirm the accuracy of the final narration, nor do they single out rehearsal as the cause of the score differences reported in Table 1.

Preparation did not take place in isolation. The journal prompts and classroom observations addressed teacher encouragement, peer relationships, learners' roles in the project, and collaboration. Learners were asked to reflect on the teacher's role, their own project roles, and their relationships with classmates. At the same time, the observations attended to grouping and collaborative work. These sources place social support within the activity itself. The records do not, however, show that every learner received equal assistance or that collaboration was always positive.

The observation notes extended this picture to participation in class, recording grouping, task engagement, participation, and learners' readiness to take turns during the intervention. These records describe participation as it was observed during the DST sessions. They do not

measure each learner's willingness to speak, and they cannot show whether the observed participation would carry over to speaking activities beyond DST.

Digital-production difficulties qualify the account given so far. The reflective materials asked learners to identify challenges and suggest ways of dealing with them. Editing and recording were among the practical challenges they identified, which points to the need for technical scaffolding. These difficulties matter because a digital project requires more than speaking: learners also had to manage scripts, files, recording, and editing. Such demands may create opportunities for collaboration. However, they may also spread the work unevenly. The available evidence does not measure how much speaking practice each learner did individually, or how technical skill affected the quality of the final group product.

Taken together, the two sets of findings describe the same intervention from two angles. Post-assessment means were higher in all five dimensions, and no learner scored lower on any dimension. However, many learners showed no change in pronunciation, fluency, and comprehension. The reflective and observational evidence portrays DST as a structured setting in which preparation, revision, peer support, and technical demands were closely intertwined. The strongest defensible reading is that these features were salient in some learners' accounts and in the observed

classroom processes. The qualitative evidence places the score patterns in context. It does not show that any particular experience caused a change in speaking scores, or that all participants experienced the activity in the same way.

Discussion

Patterns in Speaking-Assessment Scores Between the Pre- and Post-DST Assessments

Post-assessment means were higher than pre-assessment means in all five speaking dimensions. This pattern is consistent with previous work linking DST to speaking and vocabulary (Bai & Xian, 2024; Tatlı et al., 2022). The largest differences were in vocabulary and comprehension, which is plausible given how the task was built. Preparing a story gives learners a communicative reason to choose words and organize a message. Scripting and narrating then require them to attend repeatedly to wording and meaning. Earlier DST research has likewise examined speaking competence and vocabulary in task-based digital storytelling (Fu et al., 2021; Tatlı et al., 2022). A review of qualitative DST research identifies learner engagement, multimodal resources and the social organization of storytelling as recurring areas of interest (Isaacs et al., 2024). The present score patterns fit these process-oriented accounts. Here, gains in vocabulary and comprehension came from a task in which learners shaped their words before an audience heard them.

The present findings also differ from studies that report DST outcomes mainly as group-level gains. The rank summaries show that no learner scored lower on any dimension. However, six of

the nineteen learners showed no change in pronunciation, fluency, or comprehension. The improvement was therefore not uniform, and group means alone would have hidden this variation. The design also limits interpretation in ways that controlled comparisons do not. Without a comparison group or documentation of equivalent tasks, the study cannot single out DST as the cause of the differences. Repeated practice, growing familiarity with the task, teacher feedback, peer collaboration, and the chance to revise recorded speech may all have contributed, and their relative weight cannot be separated.

For the same reason, the findings cannot show whether the vocabulary and comprehension differences came from collaboration rather than rehearsal, feedback or familiarity. Effect sizes are not reported, and no follow-up assessment was conducted. As a result, it is unknown how large the differences were in practical terms, whether they would persist beyond the course, or whether they would transfer to unrehearsed conversation. The score patterns are therefore best read as describing the direction and relative size of change rather than as a full evaluation of the intervention's effectiveness.

Learners' Experiences of DST as a Context for Speaking Preparation and Participation

The qualitative findings show that the way the speaking activity was organized was a salient part of learners' experience. Learners did not have to produce a single immediate response. Instead, they could draft, record, listen, revise, and then share a more prepared narration. Some described this staged preparation as manageable. This account can be read alongside

work on digital storytelling and willingness to communicate (Shen et al., 2024) and research on speaking anxiety among tertiary-level English learners (Yasmin et al., 2025). Both bodies of work point to the affective conditions under which learners are willing to speak. It also supports the view of digital storytelling as a multimodal literacy practice and a form of learner-created material (Mizusawa & Kiss, 2024). In this sense, a “supportive environment” did not mean that the audience disappeared. The audience remained, but learners met it later, after they had had time to prepare. The same pattern fits Self-Determination Theory. Repeated editing may have given some learners a sense of control over their oral production, and peer assistance may have supported participation and perceived competence (Ryan & Deci, 2020).

The present study departs from that research in the kind of evidence it offers. Anxiety, willingness to communicate, autonomy, competence, and relatedness were not measured with validated scales. The findings therefore rest on how learners described the activity and how their participation was observed. They do not show a measured reduction in anxiety, nor do they show that specific psychological needs drove changes in participation. Nor do they suggest that every learner experienced the same degree of comfort. Self-Determination Theory and the affective perspective serve here as interpretive resources, not as confirmed mechanisms.

Peer and teacher support formed another part of the reported experience. This is consistent with research on digital storytelling as a site for collaborative language learning and digital-skill development (Yu & Wang, 2025). The

digital production difficulties learners reported, especially with editing and recording, also echo findings from work on mobile-assisted video-making. That work identifies proficiency, narrative ability, and digital literacy as factors in task outcomes (Kang & Kim, 2024). Broader research on digital storytelling has considered affective outcomes such as motivation and anxiety (Mohammadi Khomjani, 2020). Work on educational technology in English language teaching stresses that digital tools need to be aligned with instructional objectives (Shruthi et al., 2025). The present findings point the same way. When a digital project is used for speaking, learners face demands that go beyond oral language, particularly if they have little experience with video tools.

Collaboration, however, complicates interpretation in ways that studies focused on group outcomes may not face. Research on AI-supported story creation, for example, has examined team creativity (Wei et al., 2025), which is a different outcome from individual speaking performance. A group can offer vocabulary suggestions, technical help, and reassurance. Even so, its final video may reflect shared work rather than each learner’s independent speaking ability. If scripting, editing and narration were divided unevenly, the quality of the product cannot be credited equally to every member. The present data describe peer support as part of the process. They do not measure individual contributions or show that peer support meant the same thing for every learner, and this matters most when digital competence varies within a group. Collaboration is therefore both a potential resource and a design issue that needs deliberate management.

These findings have practical implications for teachers who want to use DST with reluctant speakers. Teachers can introduce the workflow in manageable stages, demonstrate a short recording-and-revision cycle, and provide a simple storyboard or script template. They can also let learners practise the editing functions the assignment requires and allow repeated attempts at recording. If the goal is oral communication, the task does not need advanced effects or expensive software. Assessment should prioritise intelligibility, organisation, vocabulary use and reflective revision over production polish. It should also separate individual oral performance from the quality of the group video, for example through brief role records or individual narration samples, so that editing skill does not stand in for speaking evidence. These implications are drawn from the documented process and were not tested as intervention components. The study's contribution is contextual and process-focused, and its findings cannot be generalized to all learners who report low motivation. Nor do they show that DST is better than other structured speaking activities. Future studies should document their screening and instrument procedures and use clearly described, preferably equivalent, speaking tasks. They should report score distributions and effect sizes and include a comparison condition where feasible. They should measure anxiety and motivational needs using validated scales, record individual contributions to group products, and maintain the records needed to audit both quantitative and qualitative analyses. Such work would more precisely show whether rehearsal, self-listening, revision, peer assistance, and the demands of digital

production shape the speaking development of reluctant learners, and how.

Conclusion

After the DST intervention, the learners in the lower-motivation group achieved higher mean scores in all five dimensions of the speaking rubric. Vocabulary showed the largest difference (3.10 to 4.00), followed by comprehension (3.52 to 4.36), grammar (3.15 to 3.94), fluency (3.15 to 3.84), and pronunciation (3.11 to 3.79). The overall mean rose from 3.21 to 3.99. Wilcoxon signed-rank tests indicated statistically significant differences in every dimension, and no learner scored lower on any of them. The gains were not uniform, however: six of the nineteen learners showed no change in pronunciation, fluency, or comprehension.

The reflective journals and classroom observations help explain these scores. Some learners described the staged nature of DST as making preparation more manageable, because they could plan what to say before sharing a finished story. Scripting, rehearsal, recording, and revision gave them repeated chances to prepare their narration and to notice what needed improvement. This preparation took place in a social setting shaped by teacher encouragement, peer relationships, project roles, and collaboration. Observation notes recorded how learners were grouped, how they engaged with the task, and how ready they were to take turns. Learners also reported that editing and recording placed demands on them beyond speaking. Taken together, these findings show DST as a structured setting in which preparation, revision, social support, and technical demands were closely bound up with one another.

These findings have practical implications for teachers who use DST with reluctant speakers. Teachers can introduce the workflow in manageable stages, model a short recording-and-revision cycle, offer a simple storyboard or script template, and give learners time to practice the editing functions the task requires. Assessment should value intelligibility, organization, vocabulary use, and reflective revision over production polish. It should also judge each learner's oral performance separately from the group's video, so that editing skill does not stand in for evidence of speaking.

Future research could add a comparison condition or a carefully designed repeated-measures study to separate the contribution of DST from repeated practice, feedback, collaboration, and task familiarity. Such studies should report the motivation-screening threshold, the full scoring protocol, rater training, and score distributions, along with effect sizes and confidence intervals. They could also measure anxiety, willingness to communicate, and motivational needs with validated scales, and record each learner's contribution to group products. Larger samples drawn from more than one institution would show whether the patterns found in this course hold for other groups of reluctant speakers.

Conflict of Interest

The author declare that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Generative AI Statement

The author used Grammarly solely for language editing and proofreading to improve the clarity and readability of the manuscript. All AI-assisted output was critically reviewed and revised by the author, who takes full responsibility for the accuracy, integrity, and final content of the manuscript.

References

- Bai, Y., & Xian, H. (2024). Exploring the interplay of digital storytelling, L2 speaking skills, self-regulation, and anxiety in an IELTS preparation course. *Humanities and Social Sciences Communications*, 11(1), 1–15. <https://doi.org/10.1057/s41599-024-04109-8>
- Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology: Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57–71). American Psychological Association. <https://doi.org/10.1037/13620-004>
- Cendra, A. N., & Sulindra, E. (2022). Speaking accuracy, fluency, and beyond: Indonesian vocational students' voices. *LLT Journal: A Journal on Language and Language Teaching*, 25(2), 379–394. <https://doi.org/10.24071/llt.v25i2.4579>
- Fu, J. S., Yang, S.-H., & Yeh, H.-C. (2022). Exploring the impacts of digital storytelling on English as a foreign language learners' speaking competence. *Journal of Research on Technology in Education*, 54(5), 679–694. <https://doi.org/10.1080/15391523.2021.1911008>

- Hawanti, S. (2014). Implementing Indonesia's English language teaching policy in primary schools: The role of teachers' knowledge and beliefs. *International Journal of Pedagogies and Learning*, 9(2), 162–170. <https://doi.org/10.1080/18334105.2014.11082029>
- Isaacs, M.-A., Tondeur, J., Howard, S., Claro, M., & van Braak, J. (2024). Digital storytelling as a strategy for developing 21st-century skills: A systematic review of qualitative evidence. *Technology, Pedagogy and Education*, 33(5), 573–593. <https://doi.org/10.1080/1475939X.2024.2343929>
- Kang, S., & Kim, Y. (2024). Examining the quality of mobile-assisted, video-making task outcomes: The role of proficiency, narrative ability, digital literacy, and motivation. *Language Teaching Research*, 28(6), 2326–2353. <https://doi.org/10.1177/13621688211047984>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Liang, J.-C., & Hwang, G.-J. (2023). A robot-based digital storytelling approach to enhancing EFL learners' multimodal storytelling ability and narrative engagement. *Computers & Education*, 201, Article 104827. <https://doi.org/10.1016/j.compedu.2023.104827>
- Mizusawa, K., & Kiss, T. (2024). Multiliteracies and student-created materials design: Learner engagement and agency in the EFL classroom. In V. Tavares (Ed.), *Social justice through pedagogies of multiliteracies: Developing and strengthening L2 learner agency and identity* (pp. 41–58). Routledge. <https://doi.org/10.4324/9781003438847-4>
- Mohammadi Khomjani, M. (2020). Effect of digital storytelling on learning motivation and reducing anxiety of learners to talk to others in a foreign language [Master's thesis, Concordia University]. Spectrum: Concordia University Research Repository. <https://spectrum.library.concordia.ca/id/eprint/987911/>
- Révész, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, 37(6), 828–848. <https://doi.org/10.1093/applin/amu069>
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, Article 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Shen, X., Shen, X., & Peng, J.-E. (2024). Digital storytelling: A literacy-building tool to promote willingness to communicate in a second language. *RELC Journal*, 55(3), 819–826. <https://doi.org/10.1177/00336882231157461>
- Shruthi, H. L., Radhakrishnan, A., Veigas, A. D., Railis, D. J., & Dinesh, R. S. (2025). Analyzing pedagogy and education in English language teaching using information and communication technology. *Education and Information Technologies*, 30(12), 16551–16573. <https://doi.org/10.1007/s10639-025-13439-2>

- Sulistianingsih, E., & Taufiqulloh, T. (2025). Exploring pre-service teachers' perceptions of digital storytelling in EFL classes using the TAM. *Indonesian EFL Journal*, 11(1), 79–90. <https://doi.org/10.25134/iefj.v11i1.11431>
- Sulistianingsih, E., & Aflahatun, N. (2026). Learner agency and student engagement as co-emergent processes in ICT- supported student-generated video production. *Journal of English Education Program(JEEP)*, 7(2), 205–220. <https://doi.org/10.26418/jeep.v7i2.107723>
- Tatlı, Z., Saylan, E., & Kokoç, M. (2022). Digital storytelling in an online EFL course: Influences on speaking, vocabulary, and cognitive load. *Participatory Educational Research*, 9(6), 89–112. <https://doi.org/10.17275/per.22.130.9.6>
- Wei, X., Wang, L., Lee, L-K., & Liu, R. (2025). The effects of generative AI on collaborative problem-solving and team creativity performance in digital story creation: An experimental study. *International Journal of Educational Technology in Higher Education*, 22(1), Article 23. <https://doi.org/10.1186/s41239-025-00526-0>
- Yasmin, T., Islam, M. M., Hoque, M. S., Othman, W. M. B., Reza, M., & Al Murshidi, G. (2025). Speaking anxiety of tertiary-level learners of English in Bangladesh: An investigation into the predicaments and prospects. *English Teaching & Learning*, 49(1), 87–108. <https://doi.org/10.1007/s42321-024-00165-0>
- Yu, B., & Wang, W. (2025). Using digital storytelling to promote language learning, digital skills and digital collaboration among English pre-service teachers. *System*, 129, Article 103577. <https://doi.org/10.1016/j.system.2024.103577>

