



Castledown

 OPEN ACCESS

Technology in Language Teaching & Learning

ISSN 2652-1687

<https://www.castledown.com/journals/tltl/>

Technology in Language Teaching & Learning, 6(3), 1764 (2024)

<https://doi.org/10.29140/tltl.v6n3.1764>

AI-enhanced Collaborative Story Writing in the EFL Classroom



NICOLAS EMERSON 

Kyushu Sangyo University, Japan
emerson@mail.kyusan-u.ac.jp

Abstract

Artificial intelligence (AI) is opening new possibilities for educators to explore innovative solutions to classroom challenges. While the benefits of AI in enhancing individual learning are well established, its potential in collaborative learning environments remains underexplored. This paper reports on the first cycle of a design-based research project that involved the development of Collabowrite, a collaborative story-writing application designed for compulsory English-as-a-foreign-language (EFL) classes at a Japanese university. During a 14-week semester, 52 students engaged with the application in 160 separate usage instances, using AI features such as grammar checks, virtual group members, artwork generation, and story narration. The grammar checks were evaluated for accuracy at 90.4% across 892 instances, with students adhering to elements of AI-provided guidance 91.4% of the time. Students rated their enjoyment of the activities at an average of 4.63 out of 5. Pre- and post-intervention surveys involving 27 participants and using single-item measures indicated significant positive shifts in perceptions toward English writing, collaborative work, and the role of technology in language learning. Reflexive thematic analysis of qualitative data generated three salient themes that capture the diverse experiences of participants. These findings suggest the need to revisit sociocultural learning theories to incorporate AI into classroom dynamics. This initial research cycle led to the formulation of six design principles for integrating AI-enhanced collaborative writing tools into EFL classrooms. Future cycles will further evaluate the efficacy of Collabowrite as an educational tool and its impact on students' foreign language writing abilities.

Keywords: artificial intelligence, AI-enhanced learning, collaborative writing, design-based research, educational technology, EFL, language learning tools, story writing

Copyright: © 2024 Nicolas Emerson. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. **Data Availability Statement:** All relevant data are within this paper.

Introduction

While numerous artificial intelligence (AI) educational tools are geared towards individual learning, there are relatively few designed for collaborative learning in classroom settings. This paper introduces Collabowrite, a web application designed to support collaborative story writing in English-as-a-foreign-language (EFL) classes at Japanese higher education institutions. Currently, the application's instructional elements are in Japanese only, but an English version is being developed to expand its accessibility to a broader user base. Building on insights from an ongoing design-based research project, this study presents design principles for integrating AI-enhanced collaborative writing tools into EFL classrooms. It also explores key design considerations for Collabowrite, its practical applications, and the perceptions of its users.

Although the Japanese government does not mandate EFL programs for higher education institutions, most undergraduates are required to complete approximately 90 hours of EFL courses to graduate (Gurney et al., 2023). Writing instruction is typically integrated with reading or taught as part of a comprehensive four-skills program (Yoneda, 2011). This study documents the development of an intervention aimed at enhancing the writing components of such courses. The intervention is designed with three primary objectives: 1) to increase student engagement in English writing classes, 2) to support the development of students' writing skills through the use of AI technologies, and 3) to broaden students' understanding of AI tools and to develop their proficiency in using them.

The genesis of this project unfolded against a backdrop of notable technological advancements that have reshaped educational methods globally. The COVID-19 pandemic accelerated the integration of educational technology, especially in Japanese universities where adoption had previously been slow (Tejero, 2022). Concurrently, the release of potent commercial AI tools by companies like OpenAI in 2022 enabled the integration of advanced AI capabilities into educational platforms through simple application programming interface (API) calls. Additionally, the near saturation of smartphone usage among young people in Japan (Cao et al., 2022) has opened new avenues for mobile learning. These shifts provided the impetus for reevaluating and exploring new approaches to courses at the university where this study took place, using technology to address educational challenges and improve learning experiences.

Problem Statement

By the time students in Japan enter university, they have undergone eight years of compulsory English education, predominantly characterized by traditional, teacher-centered methods aimed at preparing students for entrance examinations (Hirose & Harwood, 2019). English education in Japan is less effective than English education in other countries, as evidenced by the country's ranking of 92nd out of 116 non-English speaking countries in terms of English proficiency (EF Index, 2024). In response to these challenges, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) has advocated for more active learning and dialogical classroom practices (Nakashima, 2021) while highlighting creativity, collaboration, and independence as essential competencies for a "lifelong learning society" (MEXT, 2013). Despite this advocacy, shifts in classroom practice have been slow to materialize, with teachers reporting a lack of confidence in implementing these changes (Nishino, 2011).

Japanese university students enrolled in compulsory EFL programs often lack a well-defined reason for learning English, which manifests as low engagement in learning activities (Moritoshi, 2009; Fryer et al., 2014; Samifanni, 2024). Integrating technology into course design is one way teachers at higher education institutions can increase engagement, transforming both the content and manner in which

students learn (Jahnke & Liebscher, 2020). New technologies hold the potential to support second language (L2) acquisition, with AI tools capable of providing near-instant, personalized guidance, scaffolding learning, and traversing language barriers. The rise in generative AI usage, with 30% of students now regularly using tools like ChatGPT (NFUCA, 2024), underscores the need for policies that promote the ethical use of AI (Dakakni & Safa, 2023). As AI's capacities expand, it is vital that learners are equipped with the critical thinking skills necessary to assess AI outputs and make informed decisions regarding their use (Gašević et al., 2023). Exploration of these issues should be collaborative, with educators guiding their learners in using AI to maximize this technology's benefits (Carvalho et al., 2022).

This study responds to the need for innovative pedagogies that foster competencies in creativity, collaboration, and critical thinking. It examines how AI-enhanced tools can support collaborative learning within the language classroom, increasing student engagement and providing a platform for exploring responsible AI usage.

Background

Collaborative Writing Within L2 Learning

Collaborative writing is a pedagogical practice where learners co-author texts, collectively owning and shaping a single piece of writing (Storch, 2013, 2019). This method has proven effective in focusing L2 learners on the form of their output, with peers providing corrective feedback and opportunities to modify the output, thereby fostering improved L2 development (Li & Zhang, 2023). Collaborative writing is optimally understood through the sociocultural theory of mind (Vygotsky, 1978), which posits that learning occurs through social interaction. In collaborative writing tasks, participants often assume roles that are active but unequal, characterized by Vygotsky as expert and novice. This dynamic allows the novice to progress from their current understanding to their potential capability within the zone of proximal development (ZPD). The roles of expert and novice are fluid, allowing learners to exchange ideas, share resources, and provide immediate feedback to one another through collective scaffolding (Kowal & Swain, 1994). In classroom settings, teachers provide an additional layer of support, but their capacity to offer continuous guidance to every learner is inherently limited. Educational tools that incorporate AI can augment teacher and peer support by providing rapid and personalized assistance, helping learners remain within their ZPD. To date, most research on the classroom deployment of AI in support of L2 writing has focused on individuals making use of AI tools, such as a study by Song and Song (2023) which found significant improvements in writing skills and increased motivation. While studies such as this highlight AI's benefits for individual writing, its impact in collaborative settings is less well understood.

Collaborative Writing Activity Dynamics

Designing the Collabowrite web application and its learning environment required a thorough understanding of the various factors that affect collaborative writing outcomes. An important consideration is the selection of the mode of writing. One of the most common forms of collaborative writing in L2 classrooms is through narrative and story-writing tasks (Zhang et al., 2021). These tasks often fall into two categories: fictional narratives and personal narratives. Writing fictional stories has the potential to create more motivating learning environments by capitalizing on students' imaginative energies, thereby increasing engagement (Maley & Kiss, 2018). Incorporating creative writing into language instruction encourages playful experimentation with language, enabling learners to explore the breadth of their expressive capabilities (Crystal, 2001). Additionally, creative writing supports language acquisition at various proficiency levels by providing a differentiated learning space where

students can practice and improve their mastery of grammar, discourse, and vocabulary while also promoting higher-order thinking skills (Thorpe, 2022).

Another critical factor in designing collaborative writing activities is the choice of communication language among students. Zhang (2018) explored language use among Chinese university EFL students in collaborative writing tasks and found that learners who communicated in their first language (L1) produced texts with higher syntactical complexity than those who used their L2. While communicative language teaching methodology generally advocates for extensive L2 use within the language classroom (Nation, 2003), the strategic use of L1, particularly for clarifying complex grammar rules, is recognized as a pedagogically sound practice (Shibasaki, 2020). Variables such as L2 proficiency, task type, and pedagogical goals should be weighed when determining how students interact (Koizumi, 2022). If the primary goal of the collaborative writing activity is to elicit linguistic forms in the learners' texts rather than focusing on L2 oral production—as is the case in this intervention—then allowing learners to choose their language of interaction may be most effective (McDonough et al., 2016).

Group configuration also influences collaborative writing outcomes, with sizes of two to four members often preferred. Larger groups offer diverse linguistic resources, but this can lead to decreased ownership of the text and reduced individual accountability (Storch, 2021). Moreover, more proficient learners in larger groups may limit contributions from less proficient peers (Storch, 2013). Despite these challenges, larger groups have been linked to improved text accuracy and enhancements in fluency and syntactic complexity (Fernández Dobao, 2012). In addition to group size, the method of group formation also impacts learning outcomes. Teacher-assigned groups tend to produce significantly more language-related episodes (instances where learners reflect on language use and negotiate meaning) because teachers can strategically match students who complement each other (Mozaffari, 2017). Conversely, self-selected groups often increase enjoyment in collaborative tasks since students are working with friends; however, these groups may more frequently go off-topic (Tsoi & Aubrey, 2024). Given the scant research on effective group configurations for AI-assisted collaborative story writing, especially with smartphones, this study employed an inductive approach to explore and identify optimal group dynamics, ensuring Collabowrite was tailored effectively to classroom needs.

AI-assisted Collaborative Writing

Writing is a fundamental mode of output in L2 learning, offering learners opportunities to apply and practice linguistic forms. However, writing in an L2 is a complex activity involving multiple cognitive and linguistic processes (Kormos, 2023), requiring teachers to support learners through differentiated instruction and corrective guidance. Providing timely feedback enables students to make connections between their errors and the correct forms while the writing task is still fresh in their minds (Klimova, 2015). Yet, managing this level of support in large classes is often impractical. Among technological solutions, AI-based feedback for EFL writing stands out, offering benefits such as enhanced speed, personalization, and multilingual capabilities (Rad et al., 2023). These features position AI as a promising solution to enduring educational challenges, making it possible to provide the necessary support to learners effectively. However, reliance on AI for feedback raises questions about the quality of guidance and the potential reduction of teacher-student interaction (Kawashima, 2023). Careful integration is required to ensure that AI tools complement rather than replace essential instructional support.

The deployment of platforms like ChatGPT in classes on persuasive writing illustrates AI's potential as a collaborative tool in the classroom (Su et al., 2023). Research, such as that by Utami et al. (2023), shows that students perceive AI as a valuable intelligent assistant during the drafting phase of academic writing. Further studies by Gayed et al. (2022) and Zhao (2023) reveal AI's ability to mitigate the

cognitive challenges of L2 writing and assist learners in refining their written output. The integration of AI features into word processing software, such as Microsoft Copilot in Word, indicates a future where AI is an integral part of the writing process. Empirical evidence from Nazari et al. (2021) supports the notion that AI-assisted writing can improve performance scores and boost engagement in EFL contexts. AI also has the potential to scaffold reflection during English writing tasks through providing real-time feedback, offering substantial benefits such as enhanced writing performance, increased self-efficacy, and reduced cognitive load (Liu et al., 2023). While these studies demonstrate AI's potential in supporting individual writing, there is limited research on its impact within collaborative writing involving multiple students—a deficiency to which this study aims to contribute.

Beyond providing guidance on written content, generative AI can facilitate the creation of multimodal artifacts that combine media such as images, audio, and video. The advent of Web 2.0 technologies has expanded concepts of audience and readership, making multimodal resources essential components of digital communication and enriching the transmission of meaning (Li & Storch, 2017). Digital storytelling, employed as a teaching strategy, taps into this multimodal potential by enabling learners to use diverse online tools for narrative creation. This approach not only enhances the enjoyment of writing but also fosters creative learning (Alemi et al., 2022). However, the lack of platforms that fully support multimodal writing necessitates the use of multiple tools, which complicates the learning process and results in a loss of control for the teacher. Recognizing the pedagogical potential of multimodal collaborative writing, this project responds to the call for further research on the topic, as emphasized by Li & Zhang (2023). Integrating multimodal elements into collaborative writing can potentially increase engagement but it may also introduce complexity that challenges learners' technical skills (Walsh, 2010). Therefore, user-friendly platforms that streamline multimodal creation are essential.

Application Design Considerations

Cognitive load theory, which posits that the limited capacity of working memory requires instructional design that minimizes extraneous cognitive load to optimize learning, serves as a foundational guideline (Sweller, 1988). This principle is particularly critical for an application which involves users writing in an L2, a task inherently laden with high cognitive demands (Kormos, 2023). Within digital learning environments, the design of the user interface (UI) plays a key role in managing cognitive load. An effectively designed UI helps learners concentrate on essential tasks by minimizing distractions from unnecessary information or cluttered interface elements (Zhang & Adipat, 2005). Along with a user-friendly UI, another important consideration when designing web applications for the classroom is device-agnostic accessibility, especially in contexts where access to diverse technological resources might be limited. Ensuring that educational applications function across multiple devices not only enhances their usability but also broadens their reach (Krull & Duarte, 2017). In the context of Japanese higher education, where smartphone usage is widespread (White, 2019) yet computer literacy may vary (Murray & Blyth, 2011; Cochrane, 2022), prioritizing accessibility via smartphones is essential.

Research Questions

The development and implementation of Collabowrite are guided by the following core research questions (RQs):

- RQ1: How can Collabowrite be designed to effectively support collaborative story writing in the EFL classroom?
- RQ2: Which pedagogical approaches and deployment strategies enhance learning outcomes from Collabowrite activities?
- RQ3: How do students' perceptions of writing, collaboration, and technology supported learning evolve through the use of Collabowrite?

Method

Building on the theoretical foundations outlined in the background section, an embodied conjecture was developed, illustrated in Figure 1. This conjecture map delineates the design environment and its intended mechanisms for achieving the targeted educational outcomes.

Informed by a thorough review of literature, collaboration with an experienced web developer, and consultations with teaching staff and stakeholders, a strategic plan was developed to address the identified challenges in EFL writing classes at the target institution. The design principles for integrating AI-enhanced collaborative writing tools into EFL classrooms that have guided this project are as follows.

Design Principles

Principle 1: Utilize Generative Ai to Support L2 Learning

Employ generative AI tools to scaffold learning for L2 learners by identifying mistakes and providing guidance on how to correct them, along with explanations for these corrections. Ensure these explanations are delivered in either the L2 or the learners' L1, tailored to their specific needs.

Principle 2: Harness Generative AI to Engage Learners

Use generative AI to transform learners' text outputs into engaging multimodal artifacts. This fosters creative expression and enhances engagement through technological interactivity.

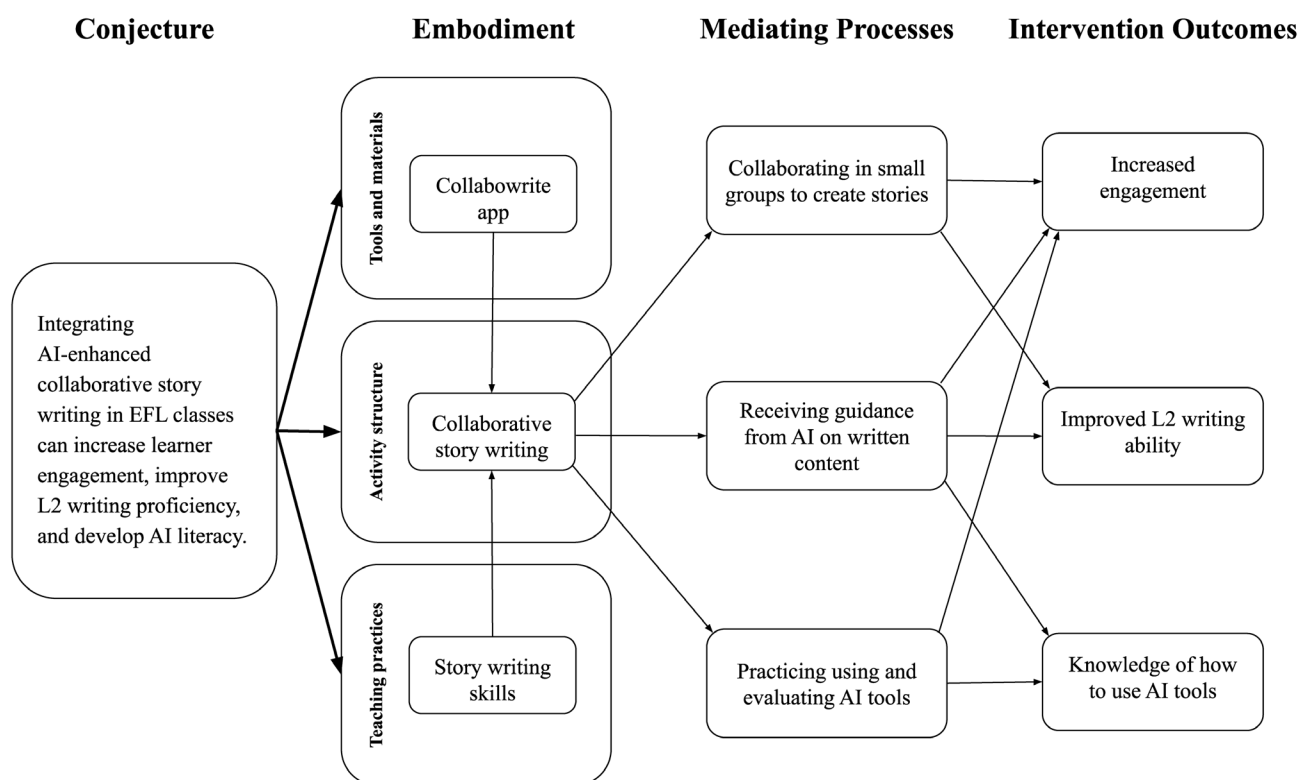


Figure 1 Conjecture Map for AI-enhanced Collaborative Story-Writing in EFL Classes.

Principle 3: Develop Competencies in AI Tool Usage

Integrate AI elements into digital educational tools to provide a platform for learners to develop their AI usage competencies. At the same time, provide instructors with guidance and resources to address ethical considerations and promote safe technology practices.

Principle 4: Ensure Safety and Control Through a Walled-Garden Approach

Position generative AI elements within a walled garden to maintain a safe and controlled environment. This approach restricts external influences, ensures the security of digital interactions, and allows educators to manage and oversee their learners' experiences using AI.

Principle 5: Prioritize Usability, Accessibility, and Comprehensibility

Design educational applications to be easy to use and intuitive, and to be accessible across various devices including smartphones. Employ learners' L1 in instructions, when necessary, to enhance clarity and comprehension.

Principle 6: Establish a Structured Activity Framework and Supporting Pedagogy

Integrate structured activities within a defined learning sequence to develop learners' collaborative writing competencies. Organize activity time to include targeted skill instruction and ensure that tasks are manageable within single class sessions and that the sessions include time for guided reflection.

Collabowrite Application Design

Collabowrite's UI adopts a chat-style format, which mirrors the familiar communication modes prevalent on smartphones. This design choice lowers the initial learning barrier, enabling students to interact with the application more fluidly. Additionally, the application incorporates Japanese for all instructional elements, allowing users to focus more effectively on their creative and linguistic tasks. Security is prioritized with single sign-on access through institutional accounts, and robust data protection protocols are in place, including server-side encryption and strict personal data collection limits.

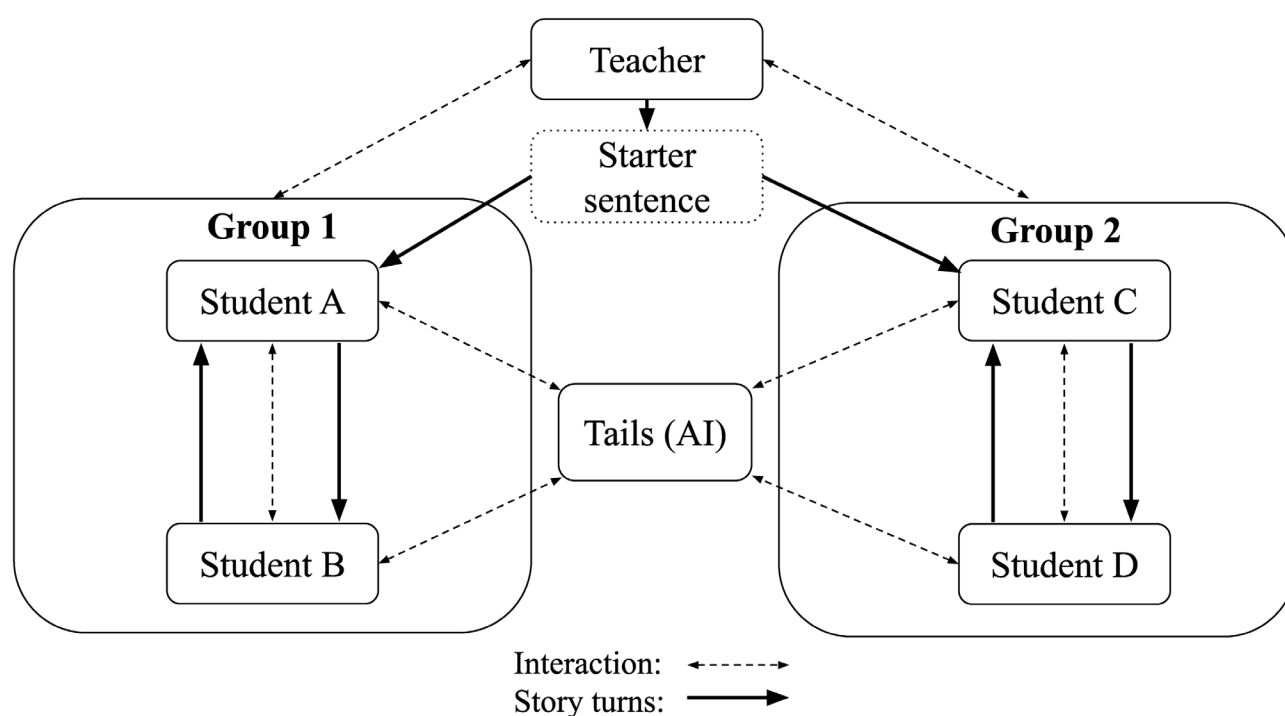
The activity flow of a typical class using Collabowrite is detailed in Table 1. A critical consideration in designing the application was the duration of activities. Opting for a single-class format primarily addresses the potential disruptions caused by absent group members in subsequent classes. Additionally, completing a story within one class session facilitates the immediate presentation of collaborative works, and enables reflection on points learned within the same session.

Given that Collabowrite activities typically occupy most of a class of 90 to 100 minutes, effective time management is crucial. While the application design assists in part with pacing, it is essential for teachers to monitor time constraints actively and ensure students have sufficient time to complete their stories. Group sizes are recommended to be between two and three members. Grouping is performed by the teacher at the beginning of each activity, after confirming activity settings and inputting a starter sentence that is uniform across all groups. Figure 2 illustrates the interaction patterns and sequence of turns during the writing phase of a standard Collabowrite activity.

Following the starter sentence, students take turns to input text, typically on their mobile devices. Throughout the writing phase, group members engage in narrative and language-related discussions,

Table 1 Activity Flow of a Typical 90-Minute Collabowrite Class with 15 Students

Phase	Activities	Duration Approx.
Prepare	1. Teacher introduces the learning focus.	10 mins
	2. Students get into groups and discuss plot, etc.	5 mins
	3. Students take turns entering text to write a story.	40 mins
Create	4. Students work together to write the ending and title.	10 mins
	5. Each student generates an image from a prompt.	5 mins
	6. Students vote to select the story's cover image.	2 mins
	7. Students rate and provide feedback on activity.	3 mins
	8. Teacher selects an AI narrator for each story.	3 mins
Reflect	9. Stories are presented to the class.	10 mins
	10. Stories are published to a shared library.	2 mins

**Figure 2** Interaction Patterns and Sequence of Turns During the Writing Phase.

with each member responsible for entering text on their device during their turn. They can request AI grammar checks limited per turn by the activity settings. These checks result in a response from the Collabowrite avatar, Tails, powered through an API call to OpenAI's GPT-4 model, returning results within seconds. All content is vetted through OpenAI's moderation API to ensure compliance with established guidelines. Figure 3 illustrates the grammar check process in action. On the left, Tails provides corrective feedback to improve grammatical accuracy. On the right, where no significant issues are detected, Tails delivers personalized praise based on content that was input.

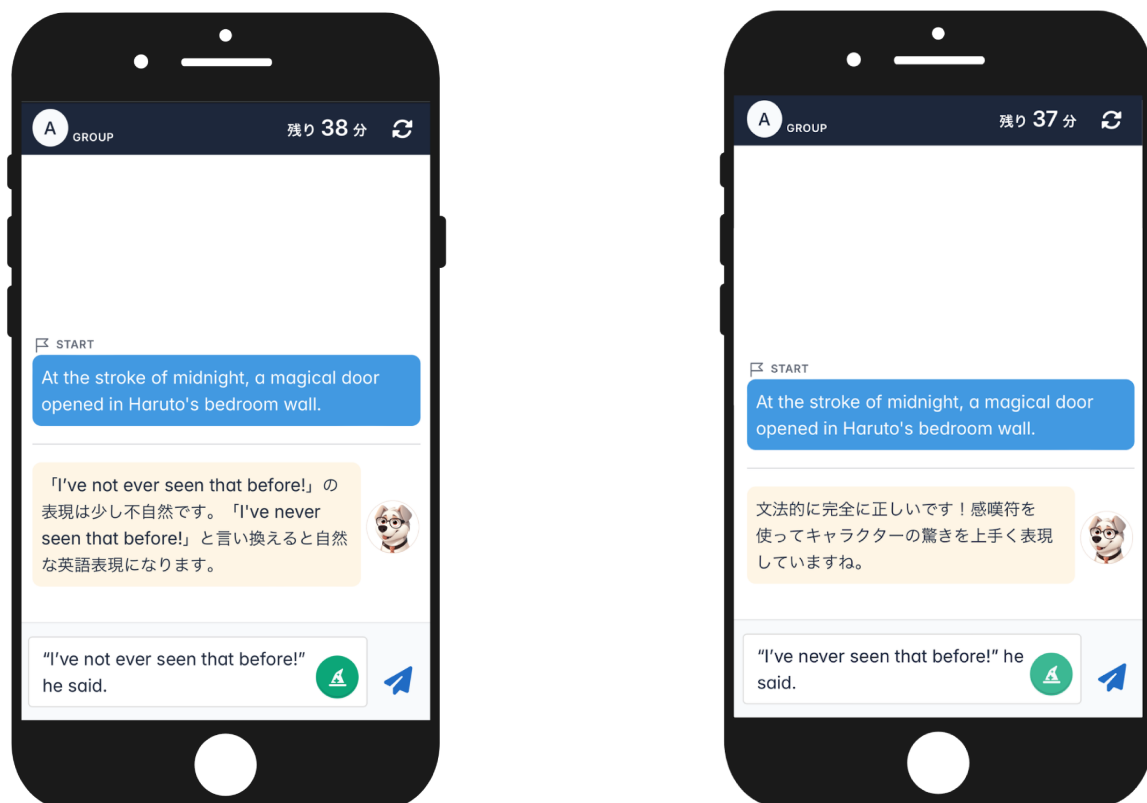


Figure 3 Grammar Check Responses: Corrective Feedback (Left) and Personalized Praise (Right).

After the main writing period ends, the next student composes the final sentence, followed by another student who gives the story a title. In-application guidance assists students through these stages, fostering collaboration and shared decision-making.

In an alternate writing mode, students collaborate with an AI group member named Ducky. During each turn, the AI group member presents two options, and students discuss and vote on their preferred choice. The AI group member steps back for the final sentence and title, which are completed solely by the students. Figure 4 shows students being introduced to Ducky at the beginning of the activity (left) and the voting mechanism in action (right).

Once the stories are finalized, an API call to GPT-4 generates an image prompt intended to visually complement the story (Figure 5, left). Students can generate an image using this prompt, but they are generally encouraged to edit it so that the image better matches their own vision. Students select an image medium and style from a curated list of visual examples. If the AI group member mode is used, then Ducky makes an image too. Images are generated through API calls, utilizing either OpenAI's DALL-E 3 model or Stability AI's Stable Diffusion models. This creative process culminates in a student vote to select the image, which will serve as the cover.

After selecting their cover images, students rate the activity on a scale of 1 to 5 stars and provide feedback on their experiences and suggestions for improvements. Concurrently, the teacher selects an AI narrator to create an audio version of each story via an API call to the speech synthesis software company ElevenLabs. Once these steps are completed, the stories are presented to the class and subsequently published to a shared library, shown in Figure 5 (right).



Figure 4 Introduction of AI Group Member (Left) and the Voting Mechanism in Use (Right).

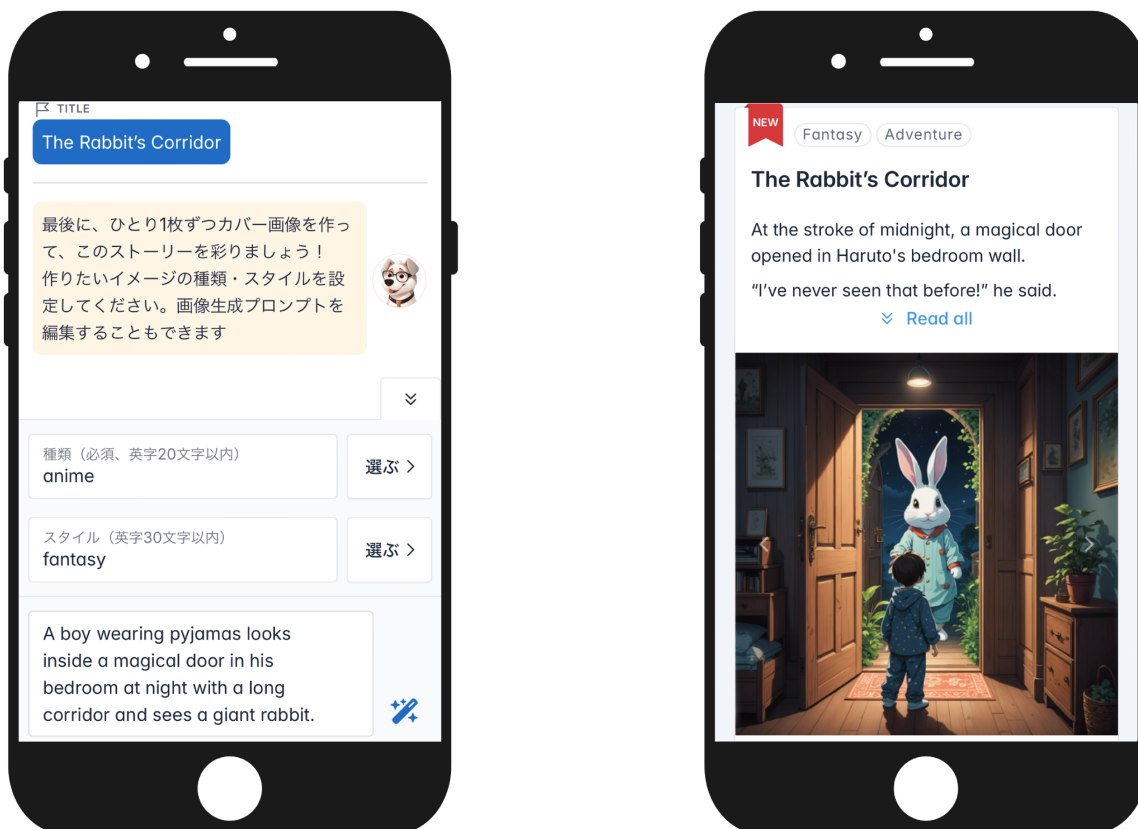


Figure 5 Image Generation Prompt (Left) and Story Library Interface (Right).

Research Design and Implementation

This project employed a design-based research approach, following the educational design research procedures outlined by Bakker (2018). The initial planning stage culminated in the formulation of a conjecture map, based on Sandoval (2014), which captures how the design embodies theoretical insights to achieve intended outcomes through mediating processes. This conjecture map was developed alongside design principles that have guided the project.

Design-based research is known for its pragmatic nature, facilitating continual adjustments through iterative cycles and seeking solutions for complex real-world educational challenges. A defining characteristic of design-based research is its collaborative nature, incorporating multiple perspectives to understand multifaceted problems thoroughly. Core personnel involved in this project included the author, serving as a teacher/researcher at the university where the study took place, and an experienced web developer who led the technical development. Stakeholders, including teachers and students, provided valuable insights that informed design choices and revisions.

The project commenced in December 2022 with an exploration of compulsory writing courses and the identification of existing issues. Challenges identified included low student engagement, insufficient collaborative activities, scant opportunities for creative expression, and excessive dependence on machine translation for writing assignments. Through consultation, planning, and development, Collabowrite and its supporting pedagogy were constructed. The first phase of research took place between September and December 2023, with the application being piloted by three teachers across four classes. This paper discusses the findings from this first phase of research. Additional phases are in progress, positioning the results discussed here as preliminary.

Participants and Ethics

The study was conducted at a four-year private university in western Japan, where no English majors are offered, and all students must earn credits in compulsory English courses to graduate. A total of 52 first- and second-year undergraduate students from four classes participated. These students provided consent for their anonymized application usage data and user feedback to be utilized for research purposes. They were invited to participate via the learning management system used for their EFL courses, where informed consent was obtained.

All participants were enrolled in compulsory EFL courses with writing components. Classes 1 ($n = 13$) and 2 ($n = 17$), taught by the researcher (T1), followed a comprehensive study protocol designed to evaluate the educational impact of Collabowrite. The students in these classes consented to the use of their anonymized survey responses for research, both prior to and following the intervention. This protocol included an introductory lesson on AI, a standalone lesson on image prompting, and short writing skills development activities prior to using Collabowrite. Consequently, only students in T1's classes participated in the pre- and post-intervention surveys. In contrast, T2 taught Class 3 ($n = 16$) and T3 taught Class 4 ($n = 6$); their classes used the application twice without following the dedicated program. Both T2 and T3 participated in interviews and responded to post-intervention email surveys after providing informed consent.

Ethical clearance for the study was secured in two stages. Initially, the project received funding through an institutional grant, which required a detailed outline of ethical procedures and demonstrated benefits for students. Subsequently, the Language Education and Research Center within the university reviewed and granted ethical approval for conducting the study.

Positionality

In the design and interpretation of this study, it is important to acknowledge my dual role as both a classroom practitioner and researcher. My educational philosophy is rooted in the pragmatist tradition, as advocated by Dewey (1938), which views learning as an active and constructive process shaped by experiences and interactions. As a practitioner, I trained as a creative media teacher and have consistently aimed to foster creativity within the classroom. Adhering to task-based learning principles, I believe that students learn most effectively when engaged in producing meaningful artifacts. This commitment to practical, experience-based learning underpins my approach to integrating technology and pedagogy, aiming to transform educational environments and outcomes.

Instrumentation and Analyses

A wide range of quantitative and qualitative data were collected as part of this project, reflecting the multifaceted nature of design research in real-world settings. This diverse dataset helped provide answers to research questions, informing the refinement of design principles and guiding changes to Collabowrite and its pedagogical approach.

Quantitative Data Collection and Analyses

Quantitative data were collected from various sources to provide a comprehensive understanding of Collabowrite's usage and effectiveness.

Application Metrics: Data on mean group size, writing time, words per use, and grammar checks per user were collected directly from Collabowrite. These metrics helped inform adjustments to the application and its classroom use throughout the study period.

AI Grammar Checks: Responses were evaluated by the researcher based on the criterion that any guidance containing incorrect or false information was deemed an inaccurate check, even if some aspects of the advice were correct.

Uptake Rates: Student outputs following AI grammar check guidance were evaluated based on the criterion that if any part of the guidance was incorporated into the output text after advice was provided, then this constitutes successful guidance uptake.

Student Pre- and Post-intervention Survey: Surveys were conducted immediately before the start and at the conclusion of the intervention period with students ($n = 27$) in classes taught by T1. These surveys included quantitative ratings of four statements using a 5-point Likert scale to capture shifts in student perceptions throughout the intervention (see the Appendix). The measures, adapted from established scales for emotional engagement (Skinner et al., 2009), self-efficacy (Schwarzer & Jerusalem, 1995), and technology acceptance (Davis, 1989), were presented as single-item measures to minimize participant burden. This approach is supported by Matthews et al. (2022), who affirm the practicality and reliability of single-item measures in applied research settings. Responses ranged from 1 ("strongly disagree") to 5 ("strongly agree") and were analyzed using JASP statistical software. Due to the non-normal distribution of data, non-parametric Wilcoxon signed-rank tests were used to assess significant changes in perceptions, with Rank-Biserial Correlation employed to measure effect sizes.

Qualitative Data Collection and Analyses

Qualitative data for this study were collected from various sources and analyzed using reflexive thematic analysis to understand participant experiences. This multilayered approach helped generate three

salient themes that encapsulate the diverse experiences of participants throughout the study. These themes are detailed in the results section of this paper.

Development Records: Comprehensive records included initial concept mapping, task logs, change logs, and a document on end-of-cycle issues and approaches.

Field Notes: Collected by T1 within the notes section of Collabowrite after every application usage, recording observations and providing insights into classroom dynamics.

Interviews: Conducted with T3 before the intervention for situational context, and with T2 after their first session using Collabowrite to capture initial reactions.

Email Surveys: Sent to T2 and T3 after the intervention period to collect further feedback on their experiences using the application.

In-application Feedback: Students responded to open-ended feedback requests after each use of the application, offering insights into their immediate reactions and identifying any issues encountered.

Student Surveys: Alongside Likert scale ratings, the surveys requested students to elaborate on their numeric responses to questions one through four. Additionally, the post-intervention survey incorporated six open-ended questions to gauge students' experiences with Collabowrite (see the Appendix).

All Japanese language content was initially translated via DeepL, followed by verification and adjustments by a bilingual Japanese native to ensure translation accuracy.

Validity

The validity of this study was ensured through a comprehensive approach that utilized multiple data sources. Quantitative data underwent statistical analysis, while qualitative data were examined to uncover deeper insights, employing Braun & Clarke's (2021) method of reflexive thematic analysis. This method leveraged the author's dual role as both a practitioner and investigator, providing an embedded perspective that enriched the analysis. Triangulation across these diverse datasets reinforced the robustness of the findings, allowing for a nuanced understanding of the observed phenomena. However, it is important to recognize that the specific context of this study may affect the generalizability of the findings.

Results

The results are presented in two sections, looking first at the quantitative results and then at the qualitative findings.

Quantitative Results

Application Usage Data

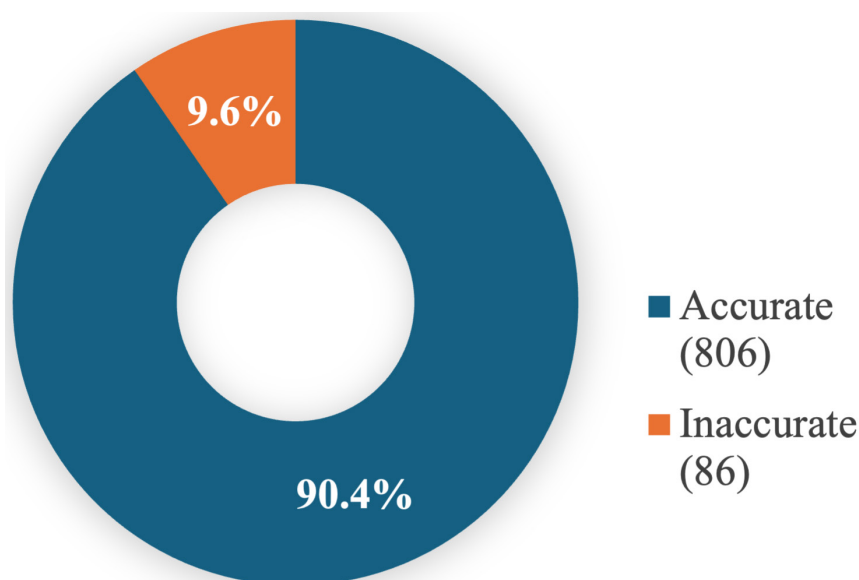
Over a 14-week semester, Collabowrite was used a total of 14 times across four classes, by 52 students. The data showing the average usage metrics for each class are shown in Table 2.

Students performed 892 grammar checks during this period, averaging 6.17 checks per student, per activity. The accuracy of grammar checks was a key concern for the research team. Grammar check

Table 2 *Average Collabowrite Usage Metrics*

Class (Total Activities)	Avg. Writing Time	Avg. Group Size	Avg. Turns per User	Avg. Words per User	Avg. GC per User
1 (6)	41.40	2.23	5.65	81.11	7.13
2 (4)	41.15	2.75	4.64	46	3.42
3 (2)	40	2.07	6.01	70.34	5.21
4 (2)	45	2	6.47	81.08	9.75
Average	41.47	2.32	5.61	69.54	6.17

Note: “GC” stands for “Grammar Check”. The values in this table are mean averages for each metric, calculated from the respective number of uses of the application by each class.

**Figure 6** *Grammar Check Response Accuracy Evaluation (Model Used: GPT-4).*

accuracy was evaluated by the researcher based on the previously outlined criterion, with the results displayed in Figure 6.

Another area of focus was analyzing the uptake rate of advice provided. Of the 806 grammar checks deemed accurate, 399 resulted in student output that did not include errors. This left 407 accurately rated grammar checks that levied advice for students to amend their story content. The researcher, following the previously outlined criterion for evaluation, assessed how often students followed some aspect of the advice provided. The results are displayed in Figure 7.

In-application Ratings

Students evaluated their enjoyment of each session using a scale of 1 to 5 stars, responding to the prompt, “Did you enjoy today’s activity?” after each of the 160 uses of the application. The average rating derived from these responses was 4.63. Figure 8 illustrates the distribution of these ratings.

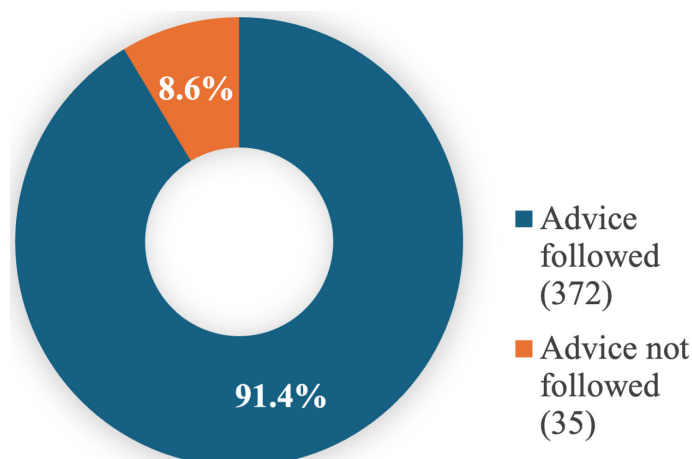


Figure 7 Uptake Rates After the AI Provided Advice Based on a Grammar Check.

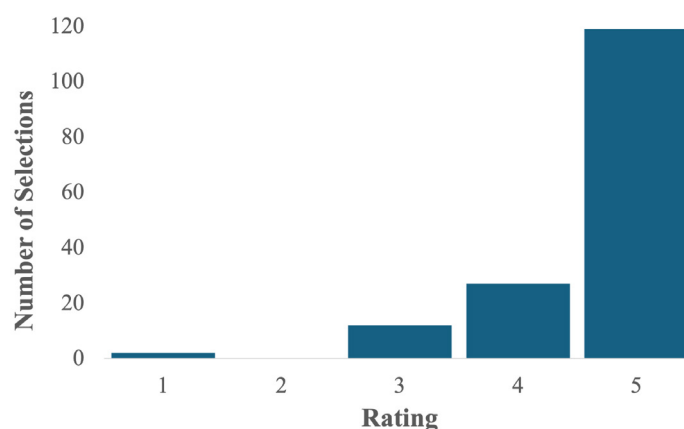


Figure 8 Distribution of Students' Ratings for Collabowrite Activities.

Student Pre- and Post-intervention Survey

A survey of 27 students assessed changes in their perceptions across four areas related to their experiences in English classes, both before and after the intervention. Due to deviations from normality as indicated by Shapiro-Wilk tests, non-parametric Wilcoxon signed-rank tests were utilized to determine the significance of changes in perceptions. The results of these analyses are summarized in Table 3.

The results indicate significant improvements with large effect sizes in students' enjoyment of writing and group work in English classes, as well as their perceptions of the usefulness of technology in understanding English. Although there was a positive trend in confidence toward English writing skills, the change was not significant.

Qualitative Findings

A reflexive thematic analysis of the large qualitative dataset generated three salient themes. These themes are discussed in detail as follows, supported by illustrative quotations from participants. For clarity and conciseness, student comments given in Japanese have been translated into English.

Table 3 *Survey Responses Before and After the Intervention that Indicate Changes in Student Perceptions*

Survey Item	Pre <i>M</i> (<i>SD</i>)	Post <i>M</i> (<i>SD</i>)	<i>p</i> value	Effect Size
1. I enjoy writing in English classes.	3.8 (0.8)	4.3 (0.6)	.007	−0.69
2. I enjoy working in groups during English class.	4.3 (0.9)	4.7 (0.6)	.034	−0.71
3. I am confident in my English writing skills.	2.3 (0.9)	2.5 (0.9)	.16	−0.49
4. Technology (e.g., AI) helps my understanding of English.	3.9 (1.2)	4.4 (0.7)	.017	−0.71

Note: Effect sizes were calculated using rank-biserial correlation. Values from −0.1 to −0.3 indicate a small effect, −0.3 to −0.5 indicate a medium effect, and below −0.5 indicate a large effect. Negative values indicate an increase from pre- to post-intervention.

Theme 1: Enhancement of Learning Outcomes via Effective Activity Structuring

Tasking EFL students with collaboratively writing a short story in English in one class session is demanding. To ensure success, it is important that the activity dynamics are structured to minimize difficulty and enhance learning outcomes.

Initially, groups of three or four students were expected to be optimal, but the use of smartphones for text input was observed in initial sessions to lead to disengagement among non-active members. To improve collaboration, pairs and groups of three were tested. These smaller configurations allowed for increased writing time per student, diminished the pressure of waiting among multiple group members, and fostered enhanced collaboration. The preference for pairs was evident, as indicated by P1: “It’s hard to participate in conversations in groups larger than two.” (P1, post-intervention survey).

The choice of group members was important, as certain combinations resulted in greater engagement and enjoyment. While random groupings sometimes introduced beneficial new interactions, with some students reflecting that they enjoyed working with new people, selecting students who chose to sit together often seemed to lead to improved group cohesion, possibly due to their established communicative patterns.

The social dynamics of group work posed challenges. In the post-intervention survey, some students reported issues like misaligned ideas and varying motivation and English proficiency levels. Conversely, positive feedback highlighted the enjoyment of collaborative English learning and imaginative engagement. Environmental factors also influenced collaboration effectiveness. Field notes indicated that students cooperated better when seated side by side, allowing easier viewing of each other’s smartphones and more private conversations. Additionally, light background music seemed to enhance communication between group members.

Understanding the appropriate amount of time to allocate to each Collaborowrite activity was essential. Initially, in-application feedback indicated that some students found the allotted time insufficient to fully develop their narratives, leading to frustration over unfinished stories. To address this, the author (T1) introduced a pre-writing discussion phase for brainstorming, enhancing the narrative development process. This revision in approach seemed to help, as indicated by P6: “It was good to think about the story more than before.” (P6, in-application feedback). Consequently, a “ready” stage was added

to the application, prompting students to discuss plot and character details. Despite time constraints, dedicating time at the end of the class for review allowed for praise and reflection, enriching the learning experience. Playing the audio stories enabled students to appreciate how narrative elements, like adverbs of manner, influence the AI narrators' delivery.

In reviewing the semester's activities, T1 and the software developer identified two issues: obvious errors in students' published stories and students not using grammar checks effectively. Although the grammar check provided useful guidance, some errors were still overlooked, with students often disregarding its advice. As P10 observed, "Pointing out each other's mistakes is educational, but often we don't notice mistakes just among students." (P10 post-intervention survey). While the teacher's primary role in a Collabowrite session is to facilitate, they also support learning. However, without students having the ability to edit, pointing out errors after the fact could diminish their efforts. Introducing an editing phase may extend the time required for the activity, but it also may offer educational benefits by allowing students to review and refine their inputs, while empowering teachers to better support learning.

The activity dynamics of Collabowrite were explored through the perceptions of its users. This allowed the author to refine its use in individual classes by adjusting factors such as group configuration and activity timings. To achieve the pedagogical aims of the intervention, it was important to understand how these sessions could be effectively sequenced.

Theme 2: The Importance of Progressive Learning Paths

The initial approach with Collabowrite was to develop learning points organically based on an analysis of the activities of the previous class. This reflexive process allowed T1 to introduce focused learning activities, such as board examples and practice worksheets, tailored to students' needs. Reflections on the chosen teaching and learning strategies at the conclusion of the research cycle guided the development of curricular materials that defined learning outcomes, preparing for the inclusion of additional teachers in the program.

While T1's hands-on involvement allowed for the needs-responsive creation of learning materials, this approach is unfeasible for broader faculty implementation. Clearly defined goals and learning pathways are more likely to encourage adoption by other teachers, helping them understand the application's purpose and the learning outcomes it targets. T3 emphasized the importance of clear learning objectives for each class, stating, "If I were to integrate it into a semester curriculum, I would require them to use specific grammatical structures in each session." (T3, email survey).

The implementation of Collabowrite varied among teachers. T2 used the application twice, due to curriculum constraints, with these two sessions six weeks apart. The instructor expressed concerns about a lack of student collaboration, "I was a little surprised at how many pairs did not seem to interact with each other while creating their story." (T2, email survey). T3 used the application twice in a single day during end-of-semester make-up classes, observing, "It was just a one-time activity, so they mostly found it to be an enjoyable departure from the usual writing activities." (T3, email survey).

In contrast, T1's more concerted approach appeared to lead to clearer learning outcomes through a more structured use of Collabowrite, combined with progressive skills development. Initially facing collaboration issues similar to those faced by T2, this instructor addressed them by modeling interaction patterns and carefully managing group dynamics. After the sixth session with Class 1, T1 noted improved incorporation of writing skills: "They are using the writing skill elements within their

stories and crafting more complete narratives.” (T1, field notes). Similarly, Class 2, which used the application four times, showed advancements in narrative skills and creativity, as P21 reflected. “It develops both my imagination and English usage.” (P21, post-intervention survey).

Structuring Collabowrite sessions within a clearly defined learning sequence provides learners with sufficient time to develop their story writing skills. Developing comprehensive curricula and teaching materials not only enhances learning outcomes but also facilitates Collabowrite’s wider adoption among educators.

Theme 3: Generative AI as a Potent Educational Tool

The perceptions of students and teachers regarding the AI features in Collabowrite established a feedback loop essential for the application’s iterative development. According to responses to the post-intervention survey question, “What did you think about the AI grammar check?” many students appreciated the feature for its clarity, immediacy, and educational value in enhancing their understanding of English grammar. P1 found it instructional, responding, “Very meaningful. It helped me understand correct grammar and the reasons for it.” P5 noted its autonomy-enhancing aspect, “Helps find grammar mistakes without constantly asking the teacher.” However, P10 pointed out occasional inaccuracies, “Very convenient. It helped me understand my weak areas. Sometimes, though, it incorrectly points out things I’ve written correctly.” (P1, P5, and P10, post-intervention survey). Despite some imperfections in the grammar check, students generally viewed it as a valuable tool for improving their English proficiency and confidence in writing, acknowledging and accepting its limitations as part of the learning process.

A significant drawback of the grammar check was its reliance on students first producing sentences, which some found challenging due to difficulty in generating ideas. Compounding this issue, initial sentence composition was also problematic, as students often lacked the vocabulary or grammatical structures needed to articulate their thoughts effectively. As P12 noted, “I like writing, but it’s hard to put English into words.” (P12, post-intervention survey).

While the grammar check could offer guidance based on students’ output, they also needed help with input. In the post-intervention survey, over half the respondents mentioned using translation tools like Google Translate alongside Collabowrite. While T1 regarded students using machine translation to look up words and phrases as logical, constructing whole sentences with it raised concerns. To address this, T1 advised students suspected of over-using machine translation, “Write shorter sentences in-application and consult Tails for guidance.” (T1, field notes).

During T1’s final sessions using Collabowrite, students engaged with the AI group member to write stories, with varied experiences. Some students appreciated Ducky’s unique contributions, finding the unexpected ideas stimulating. P24 shared, “It was fun because the AI changed the story in an interesting way with ideas we didn’t have.” (P24, in-application feedback). Conversely, some students felt frustrated by a perceived loss of creative control, though they acknowledged benefits, “There were times when the story didn’t go the way I wanted it to. But it was good to be presented with expressions I don’t normally use.” (P7, in-application feedback).

The image generation phase of Collabowrite activities provided a refreshing shift after intensive writing sessions. Initially, some students observed mismatches between their envisioned images and those produced by the AI. In response, T1 implemented a standalone image-prompting class to refine students’ visual description skills. The class involved exercises where students verbally described an image card, featuring diverse scenes and characters, for their partner to draw. They then crafted

30-word prompts for a new image card, which their partner used as the basis for another drawing. The images and drawings were displayed side-by-side, sparking discussions about the effectiveness of the prompts. This practical exercise improved satisfaction with AI-generated images, as students reported fewer frustrations and took ownership of their prompting techniques. In the post-intervention survey, students reflected on what they'd learned. For example, one student reported, "I realized it's better to give a clear, simple description for the AI to generate images." (P5, post-intervention survey).

In the Collabowrite activities, T1 facilitated the audio generation feature by having students in each group select their preferred narrator from available narrator samples. T1 then input the final selection, allowing students to see how lexical choices influenced the delivery of the narration. P8 noted, "I realized that adding details like 'said happily' makes the narration more expressive." (P8, post-intervention survey). Students also observed how different narrators' voices altered the mood of their stories, adding emotional depth to their writing.

A primary goal of the intervention was to deepen students' understanding of AI technologies. T1 crafted classroom activities both to enhance familiarity with AI tools and to cultivate critical awareness of the risks of relying on them overly. The intervention commenced with a session dedicated to introducing AI, followed by a joint exploration of its advantages, limitations, and best practices for use. After the final session using Collabowrite, students were assigned a writing task to reflect on what they had learned about AI. Responses in the post-intervention survey illustrated the students' learning outcomes. For example, P26 pointed out AI's potential for inaccuracy, stating, "It can help with understanding but it's sometimes incorrect," while P18 described the importance of moderate use, noting, "It's not good to use it too much." (P18 and P26, post-intervention survey).

Integrating AI into classroom practice constitutes a paradigm shift for teachers. T2 expressed concerns about not being able to monitor student errors directly, stating, "I wanted to see what mistakes they were making." (T2 interview). Making AI-generated content visible to teachers can help foster trust and add an extra layer of human checks. Such transparency in application design is crucial for building confidence in AI-enhanced learning tools.

Generative AI is transforming the way people learn, making it important for educators to reevaluate their teaching methods. Collabowrite serves both as a language learning tool and as a medium for cultivating students' understanding of AI.

Discussion

The present study reported findings from the first cycle of a design-based research project involving the development of Collabowrite, an AI-enhanced collaborative story writing application. A pedagogical intervention using the application was implemented as part of the writing component of compulsory EFL courses at a Japanese university. This intervention aimed to increase student engagement, bolster L2 writing skills, and expand knowledge of AI tool usage. The following discussion addresses the insights derived from the data in relation to the three guiding research questions.

Research Question 1: How can Collabowrite be designed to effectively support collaborative story writing in the EFL classroom?

Collabowrite integrates commercial large language models (LLMs) via API calls for all its AI components. Dependence on external providers introduces challenges, including unpredictability of service and limited control over the content of responses. To mitigate potential outages, the application employs alternative providers as a contingency. Achieving more consistent responses requires

strategies such as training LLMs on fine-tuned datasets, refining prompts, or utilizing different models. Matching the intent of the walled-garden approach suggested in this paper, hosting open-source AI models on dedicated servers would provide optimal control. However, this approach introduces new issues, such as the high costs associated with running advanced models on virtual servers and the need for in-house moderation. Gayed et al. (2022) highlighted challenges with open-source AI models, observing that using OpenAI's GPT-2 LLM for the AI KAKU tool required close instructor oversight due to unmoderated responses. Developers using commercial APIs to integrate AI must balance user data control issues with the benefits of improved performance and moderation.

The evaluation of Collabowrite's design highlighted several areas for improvement. Students often struggled to generate ideas, so integrating creativity scaffolding alongside language support would be helpful. Additionally, providing in-application vocabulary support could reduce reliance on external translation tools.

A critical enhancement needed is the capability for students to edit their texts. The current inability to revise outputs places unnecessary pressure on students to submit perfect drafts. Introducing an editing phase would not only allow students to refine their writing but would also enable teachers to act as facilitators by highlighting errors with the assurance that students can correct them. Another area for improvement involves better utilization of application usage data to deliver detailed analytics to both teachers and students.

Research Question 2: Which pedagogical approaches and deployment strategies enhance learning outcomes from Collabowrite activities?

Fostering collaborative learning when smartphones serve as the primary input devices necessitates careful planning. While studies such as Gelsomini et al. (2019) have shown that L2 students are generally positive about using their own mobile devices for classroom-based collaborative writing tasks, findings from this study reveal that the personal nature of these devices can sometimes impede collaboration. To address this issue, teachers can promote enhanced collaboration by modeling appropriate interactions, strategically grouping students, and employing effective classroom management techniques.

Collabowrite activities offer opportunities for targeted learning, and equipping students with specific skills to focus on during the task enhances the educational value of the exercise. To maximize L2 learning potential, establishing a clearly defined learning sequence helps students develop and build a repertoire of skills incrementally. Pre-teaching relevant grammatical and narrative structures provides students with a specific aim within the writing process and sets a precise target for feedback. Concluding sessions by showcasing the students' stories highlights the importance of an audience and holds students accountable for creating meaningful work.

Utilizing generative AI for tasks such as image creation provides students with the opportunity to develop prompting skills, giving them a greater sense of ownership over their creations. While stories written in an L2 may pose comprehension challenges, images generally offer a more immediate connection for many students, particularly resonating with young people who often value visual content highly. Furthermore, incorporating audio elements can complement textual and visual information, adding another dimension to their stories. Giving students more control over these multimodal elements not only enriches their storytelling experience but also actively engages them in reflecting on the learning process.

Introducing generative AI through a controlled environment like Collabowrite provides a monitored space for safe and educational AI interactions. Utilizing the application to familiarize students with

AI helps cultivate critical thinking skills, especially in evaluating AI outputs. As students increasingly turn to these tools for academic support, fostering discussions on ethical use equips them to navigate these technologies responsibly.

Research Question 3: How do students' perceptions of writing, collaboration, and technology-supported learning evolve through the use of Collabowrite?

Incorporating feedback mechanisms within Collabowrite has yielded invaluable data on student interactions. Insights from activity ratings and comments have facilitated ongoing monitoring and identification of issues, guiding iterative improvements. This approach ensures that modifications are informed by user experiences, enhancing the usability of the application and consequently its educational effectiveness.

Survey results indicated that students' enjoyment of writing in English and working in groups significantly improved after the intervention. This positive shift is reflected in favorable enjoyment ratings and qualitative feedback from both teachers and students. These results are consistent with those of Hemati and Farahian (2024), who found that collaborative writing via Google Docs enhanced classroom engagement among Iranian EFL students. The present study extends the concept of collaboration by incorporating AI that functions both as a peer and an expert. Although students found the Collabowrite tasks challenging, their consistent enjoyment suggests that the difficulty was within their capabilities. The collective scaffolding provided by peers, teachers, and in-application guidance enabled them to operate effectively within their ZPD. This finding is further supported by Wei (2023), who demonstrated that combining collaborative activities with adaptive AI support significantly enhances students' language achievement.

Findings indicate a significant positive shift in students' perceptions of AI technology's role in enhancing their understanding of English. Analysis of uptake rates reveals that students followed at least some element of the AI-generated grammar guidance 91.4% of the time. Additionally, qualitative data suggest that students value this guidance, despite occasional errors. This demonstrates that the perceived value of the AI technology utilized in this study is high enough for mistakes to fall within an acceptable margin, underscoring its effectiveness in supporting English language learning. However, studies such as Maruzuki et al. (2023) highlight educator concerns about potential negative effects of AI usage, such as diminished critical thinking skills resulting from overreliance on AI tools. As Gašević et al. (2023) note, developing students' digital literacies alongside the use of this technology is vital to ensure they critically engage with AI-generated content.

Theoretical Implications

Grounded in the experiences of practitioners and learners, findings from the initial cycle of this project underscore the potential of AI to transform collaborative learning dynamics in language classrooms. Engeström's activity system model (2014), which extends sociocultural theory by examining the interactions between individuals, tools, and the community, offers a viable framework for interpreting how AI-integrated artifacts, such as Collabowrite, alter interaction patterns within complex sociocultural settings. Further research in this domain can inform the strategic deployment of AI tools in supporting collaborative learning in classroom contexts.

Limitations

The data collected in this study primarily aimed to inform design modifications rather than to validate the efficacy of this pedagogical approach. The use of simplified survey items and basic evaluations

of AI responses lacked the scientific rigor required for generalization to broader populations. This approach is characteristic of the preliminary stages of design-based research, where the focus is more on refining the design and its underlying principles. The design principles outlined in this paper will continue to be refined throughout subsequent design cycles.

Future Directions

Insights from the initial research cycle have prompted enhancements in the application, including new features and UI refinements. A more detailed set of criteria for evaluating AI responses is currently under development. Once completed, these criteria will guide the construction of datasets for fine-tuning. The application is now utilized by a broader user base within the institution. A progressively structured curriculum, accompanied by supporting materials, has been distributed. Comprehensive measures of L2 writing abilities and user perceptions are being collected to guide continuous improvements.

Conclusion

This study presented design principles for integrating AI-enhanced collaborative writing tools into EFL classrooms. By examining usage patterns and user perceptions of the Collabowrite application, the findings demonstrate its potential to engage learners, fostering creativity, collaboration, and critical AI usage practices. Smartphones proved practical for co-authoring multimodal stories with AI guidance in classroom settings, although their personal nature can sometimes impede collaboration. AI can serve as an expert, scaffolding learning and providing fast, accurate guidance that students find valuable. It can also engage in the creative process, acting as a virtual group member and generating ideas. As AI tools evolve, it is essential to understand their integration into collaborative classroom practices and their impact on learning dynamics. EFL educators cannot ignore the inroads AI is cleaving toward better educational outcomes for their students but should not forget the importance of maintaining oversight and control over its implementation.

References

- Alemi, M., Givi, S. S., & Rezanejad, A. (2022). The role of digital storytelling in EFL students' writing skill and motivation. *Language Teaching Research Quarterly*, 32, 16–35. <https://doi.org/10.32038/ltrq.2022.32.02>
- Bakker, A. (2018). *Design Research in Education*. Routledge. <https://doi.org/10.4324/9780203701010>
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage.
- Cao, J., Kurata, K., Lim, Y., Sengoku, S., & Kodama, K. (2022). Social acceptance of mobile health among young adults in Japan: An extension of the UTAUT model. *International Journal of Environmental Research and Public Health*, 19(22). <https://doi.org/10.3390/ijerph192215156>
- Carvalho, L., Martinez-Maldonado, R., Tsai, Y. S., Markauskaite, L., & De Laat, M. (2022). How can we design for learning in an AI world? *Computers and Education: Artificial Intelligence*, 3, 100053. <https://doi.org/10.1016/j.caeai.2022.100053>
- Cochrane, R. (2022). *A case study examining Japanese university students' digital literacy and perceptions of digital tools for academic English learning* (Doctoral thesis). Lancaster University. <https://doi.org/10.17635/lancaster/thesis/1557>
- Crystal, D. (2001). *Language play*. The University of Chicago Press.
- Dakakni, D., & Safa, N. (2023). Artificial intelligence in the L2 classroom: Implications and challenges on ethics and equity in higher education: A 21st century Pandora's box. *Computers and Education: Artificial Intelligence*, 5, 100179. <https://doi.org/10.1016/j.caeai.2023.100179>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>

- Dewey, J. (1938). *Logic: The theory of inquiry*. Henry Holt and Company.
- Education First. (2024). *EF English proficiency index*. <https://www.ef.com/epi>
- Engeström, Y. (2014). *Learning by expanding* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139814744>
- Fernández Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing*, 21(1), 40–58. <https://doi.org/10.1016/j.jslw.2011.12.002>
- Fryer, L. K., Ozono, S., Carter, P., Nakao, K., & Anderson, C. J. (2014). Instrumental reasons for studying in compulsory English courses: ‘I didn’t come to university to study English, so why should I?’ *Innovation in Language Learning and Teaching*, 8(3), 239–256. <https://doi.org/10.1080/17501229.2013.835314>
- Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 4, 100130. <https://doi.org/10.1016/j.caeai.2023.100130>
- Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant’s impact on English language learners. *Computers and Education: Artificial Intelligence*, 3, 100055. <https://doi.org/10.1016/j.caeai.2022.100055>
- Gelsomini, F., Kanev, K., Barneva, R., Bottoni, P., Tatsuki, D. H., Roccaforte, M. (2019). BYOD collaborative storytelling in tangible technology-enhanced language learning settings. In: Auer, M., Tsiatsos, T. (Eds.) *Mobile technologies and applications for the internet of things*. (Advances in Intelligent Systems and Computing, Vol. 909, pp. 55–67). Springer. https://doi.org/10.1007/978-3-030-11434-3_5
- Gurney, P., Michaud, M. & Richardson, J. (2023). A policy analysis of compulsory English classes in Japanese universities. *Journal of Educational Leadership and Policy Studies*, 7(1).
- Hemati, Z., & Farahian, M. (2024). Fostering EFL learners’ classroom engagement: Does collaborative writing through Google Docs make a difference? *Technology Assisted Language Education*, 2(1), 1–20. <https://doi.org/10.22126/tale.2024.10225.1030>
- Hirose, K., & Harwood, C. (2019). Factors influencing English as a foreign language (EFL) writing instruction in Japan from a teacher education perspective. In L. Seloni & S. Henderson Lee (Eds.), *Second language writing instruction in global contexts* (pp. 71–90). Multilingual Matters. <https://doi.org/10.21832/9781788925877-008>
- Jahnke, I., & Liebscher, J. (2020). Three types of integrated course designs for using mobile technologies to support creativity in higher education. *Computers and Education*, 146, 103782. <https://doi.org/10.1016/j.compedu.2019.103782>
- Kawashima, T. (2023). Student perceptions of Grammarly, teacher’s indirect and direct feedback: Possibility of machine feedback. *The JALT CALL Journal*, 19(1), 113–139. <https://doi.org/10.29140/jaltcall.v19n1.1017>
- Klimova, B. (2015). The role of feedback in EFL classes. *Procedia – Social and Behavioral Sciences*, 199, 172–177. <https://doi.org/10.1016/j.sbspro.2015.07.502>
- Koizumi, Y. (2022). Use of L1 in L2 collaborative writing: Does it make any difference to the product or process? *JACET International Convention Selected Papers*, 8, 113–142. https://doi.org/10.50943/jacetselectedpapers.8.0_113
- Kormos, J. (2023). The role of cognitive factors in second language writing and writing to learn a second language. *Studies in Second Language Acquisition*, 45(3), 622–646. <https://doi.org/10.1017/S0272263122000481>
- Kowal, M., & Swain, M. (1994). Using collaborative language production tasks to promote students’ language awareness. *Language Awareness*, 3(2), 73–93. <https://doi.org/10.1080/09658416.1994.9959845>
- Krull, G., & Duarte, J. M. (2017). Research trends in mobile learning in higher education: A systematic review of articles (2011–2015). *The International Review of Research in Open and Distributed Learning*, 18(7). <https://doi.org/10.19173/irrodl.v18i7.2893>

- Li, M., & Storch, N. (2017). Second language writing in the age of CMC: Affordances, multimodality, and collaboration. *Journal of Second Language Writing*, 36, 1–5. <https://doi.org/10.1016/j.jslw.2017.05.012>
- Li, M., & Zhang, M. (2023). Collaborative writing in L2 classrooms: A research agenda. *Language Teaching*, 56(1), 94–112. <https://doi.org/10.1017/S0261444821000318>
- Liu, C., Hou, J., Tu, Y.-F., Wang, Y., & Hwang, G.-J. (2023). Incorporating a reflective thinking promoting mechanism into artificial intelligence-supported English writing environments. *Interactive Learning Environments*, 31(9), 5614–5632. <https://doi.org/10.1080/10494820.2021.2012812>
- Maley, A., & Kiss, T. (2018). *Creativity and English language teaching*. Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-46729-4>
- Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2). <https://doi.org/10.1080/2331186x.2023.2236469>
- Matthews, R. A., Pineault, L. & Hong, Y. H. (2022). Normalizing the use of single-item measures: Validation of the single-item compendium for organizational psychology. *Journal of Business and Psychology*, 37, 639–673. <https://doi.org/10.1007/s10869-022-09813-3>
- Ministry of Education, Culture, Sports, Science and Technology. (2013). *The second basic plan for the promotion of education*. <https://www.mext.go.jp/en/policy/education/lawandplan/title01/detail01/1373796.htm>
- McDonough, K., Crawford, W., & De Vleeschauwer, J. (2016). Thai EFL learners' interaction during collaborative writing tasks and its relationship to text quality. In M. Sato, & S. Ballinger (Eds.), *Peer interaction and second language learning: Pedagogical potential and research agenda* (pp. 185–207). John Benjamins.
- Moritoshi, P. (2009). Motivating non-English majors in English classes. In A. M. Stoke (Ed.), *JALT2008 conference proceedings*. (pp. 283–291). JALT.
- Mozaffari, S. H. (2017). Comparing student-selected and teacher-assigned pairs on collaborative writing. *Language Teaching Research*, 21(4), 496–516. <https://doi.org/10.1177/1362168816641703>
- Murray, A., & Blyth, A. (2011). A survey of Japanese university students' computer literacy levels. *The JALT CALL Journal*, 7(3), 307–318.
- Nakashima, H. (2021). *Key points of the revised foreign language national curriculum standards (course of study) at elementary school and lower secondary school levels*. National Institute for Educational Policy Research. <https://www.nier.go.jp/English/educationjapan/pdf/20210623-01.pdf>
- Nation, P. (2003). The role of the first language in foreign language learning. *Asian EFL Journal*, 5(2), 1–8.
- National Federation of University Co-operative Associations. (2024). *59th student life survey summary report*.
- Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, 7(5), e07014. <https://doi.org/10.1016/j.heliyon.2021.e07014>
- Nishino, T. (2011). Japanese high school teachers' beliefs and practices regarding communicative language teaching. *JALT Journal*, 33(2), 131–155. <https://doi.org/https://doi.org/10.37546/JALTJJ33.2-2>
- Rad, H. S., Alipour, R., & Jafarpour, A. (2023). Using artificial intelligence to foster students' writing feedback literacy, engagement, and outcome: A case of Wordtune application. *Interactive Learning Environments*, 32(9), 5020–5040. <https://doi.org/10.1080/10494820.2023.2208170>
- Samifanni, F. (2024). Understanding the Japanese university students' motivation and demotivation in attending EFL classes. *International Academic Journal of Social Sciences*, 11(1), 11–25. <https://doi.org/10.9756/iajss/v11i1/iajss1103>

- Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences*, 23(1), 18–36. <https://doi.org/10.1080/10508406.2013.778204>
- Schwarzer R. & Jerusalem M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). NFER-NELSON.
- Shibasaki, T. (2020). Appropriate L1 and L2 use in EFL classrooms: Promoting greater second language use. *The Language and Media Learning Research Center Annual Report*, 1, 49–56.
- Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, 69(3), 493–525. <https://doi.org/10.1177/0013164408323233>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14. Article 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Storch, N. (2013). *Collaborative writing in L2 classrooms*. Multilingual Matters. <https://doi.org/10.21832/9781847699954>
- Storch, N. (2019). Collaborative writing as peer feedback. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 143–162). Cambridge University Press. <https://doi.org/10.1017/9781108635547.010>
- Storch, N. (2021). Collaborative writing: Promoting languaging among language learners. In A. W. Feng & N. Storch (Eds.), *Working collaboratively in second/foreign language learning* (pp. 13–34). De Gruyter. <https://doi.org/10.1515/9781501511318-002>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Tejero, O. M. C. (2022). Residual post-pandemic ICT literacy in higher education: The case of foreign language teachers and students in Japan. *Human Review: International Humanities Review*, 11(Monográfico), 1–10. <https://doi.org/10.37467/revhuman.v11.3997>
- Thorpe, R. (2022). *Teaching creative writing to second language learners: A guidebook*. Routledge.
- Tsoi, S. C & Aubrey, S. (2024) Impact of student-selected pairing on collaborative task engagement, *ELT Journal*, 78(1), 42–53. <https://doi.org/10.1093/elt/ccad024>
- Utami, S. P. T., Andayani, A., Winarni, R., & Sumarwati, S. (2023). Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive? *Contemporary Educational Technology*, 15(4), Article ep450. <https://doi.org/10.30935/cedtech/13419>
- Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes* (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
- Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *The Australian Journal of Language and Literacy*, 33(3), 211–239. <https://doi.org/10.1007/bf03651836>
- Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, 14. Article 1261955. <https://doi.org/10.3389/fpsyg.2023.1261955>
- White, J. (2019). Digital stories: Improving the process using smartphone technology. In F. Meunier, J. Van de Vyver, L. Bradley, & S. Thouësny (Eds.), *CALL and complexity – short papers from EUROCALL 2019* (pp. 402–406). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.38.1044>
- Yoneda, M. (2011). Integrating reading and writing tasks in an EFL class: A case study. *Ochanomizu University Academic Journal*, 7, 153–163.

- Zhang, D., & Adipat, B. (2005). Challenges, methodologies, and issues in the usability testing of mobile applications. *International Journal of Human-Computer Interaction*, 18(3), 293–308. https://doi.org/10.1207/s15327590ijhc1803_3
- Zhang, M. (2018). Collaborative writing in the EFL classroom: The effects of L1 and L2 use. *System*, 76, 1–12. <https://doi.org/10.1016/j.system.2018.04.009>
- Zhang, M., Gibbons, J., & Li, M. (2021). Computer-mediated collaborative writing in L2 classrooms: A systematic review. *Journal of Second Language Writing*, 54, 100854. <https://doi.org/10.1016/j.jslw.2021.100854>
- Zhao, X. (2023). Leveraging artificial intelligence (AI) technology for English writing: Introducing Wordtune as a digital writing assistant for EFL writers. *RELC Journal*, 54(3), 890–894. <https://doi.org/10.1177/00336882221094089>

Appendix

Pre- and Post-intervention Survey Questions for Students

1. I enjoy writing in English classes. (英語のライティングの授業は楽しいです)
2. I enjoy working in groups during English class. (英語の授業でのグループワークは楽しいです)
3. I am confident in my English writing skills. (英語のライティングスキルに自信があります)
4. Technology (e.g., AI) helps my understanding of English. (テクノロジー (AIなど) は英語の理解を助けます)
5. What are the good and bad points about writing stories in a group? (グループで物語を書く際の良い点と悪い点は何ですか?)
6. What did you think about the AI grammar check? (AIでの文法チェックについてどう思いますか?)
7. What other apps did you use whilst using Collabowrite? Please explain why you used these apps. (Collabowriteでのアクティビティ中、他にどのようなアプリを使用しましたか? また、よろしければ、そのアプリを使用した理由を教えてください)
8. What did you learn about AI generated images? (AIでの画像生成について学んだことがあれば教えてください)
9. What did you learn about AI generated narration? (AIでのナレーション生成について学んだことがあれば教えてください)
10. Please provide any additional comments you have about Collabowrite or ideas for changes to the app or how it's used in class. (Collabowriteについて、操作方法の改善点や、授業での使用方法に関するアイデアがあれば、ぜひ教えてください)