English Production Learning Cycle Using Generative AI in Image Description Tasks: Focusing on Lowering Psychological Barriers and Promoting Autonomous Noticing

Harumi Kashiwagi, Kobe University, Japan
Min Kang, Kobe University, Japan

The European Conference on Education 2026
Official Conference Proceedings

Abstract

This study examined how a three-step learning cycle—(1) independent English production, (2) reference to AI-generated samples, and (3) self-correction—affected university students' psychological barriers and "noticing" in AI-generated image description tasks. A total of 129 university students were divided into two groups (n = 65, n = 64) and performed two identical tasks while referring to AI-generated sample descriptions based on two different educational frameworks. Participants rated the AI-generated samples on a 5-point Likert scale across vocabulary, grammar/syntax, and organization, followed by an overall survey. Key findings revealed that: First, over 90% of participants perceived the cycle positively, as the subsequent revision process lowered anxiety about mistakes and fostered psychological security. Second, both groups assigned usefulness ratings of 4.4 or higher on average across all dimensions. While the ratings for vocabulary and grammar/syntax showed no significant differences between the groups, a significant difference was observed in one task, in which structures explicitly describing spatial relationships (e.g., "in the middle," "to the left," or "on the right side") were rated as particularly helpful. Furthermore, over 90% reported a sense of improvement from revising their own work using AI-generated samples as a point of reference rather than merely viewing model answers. Additionally, 60–70% identified AI-generated similar images and texts as effective support resources. These results suggest that sample English sentences generated by generative AI can serve as triggers to promote autonomous awareness in learners, while also providing design principles for psychologically supportive AI-assisted language learning environments.

Keywords: generative AI, learning cycle, psychological barriers, noticing, autonomous learning, L2 production

iafor
The International Academic Forum
www.iafor.org

Introduction

Research Background

In recent years, driven by growing global demands, higher education English language instruction has increasingly emphasized the development of "production skills"—the ability to proactively express one's own ideas and describe situations—moving beyond the passive reception of knowledge. Specifically, enhancing speaking skills, which directly connect to real-world communication, serves as a central challenge in fostering autonomous language learners.

To effectively enhance these production skills, learners' conscious awareness is crucial for internalizing linguistic information as "intake," rather than merely receiving it passively. In Second Language Acquisition (SLA) research, Schmidt (1990) proposed the "Noticing Hypothesis," positing that subjective "noticing" of linguistic features is a necessary condition for converting input into intake. According to this framework, actively attending to linguistic forms and recognizing the gaps between one's current interlanguage and the target language play a central role in helping learners process subsequent input more attentively and internalize new linguistic knowledge. Regardless of the abundance of input exposure, input is unlikely to be processed as data that promotes language acquisition unless the learner's conscious "noticing" occurs. Therefore, designing an environment that effectively directs learners' attention to appropriate linguistic forms and triggers autonomous "noticing" is a critically important task for modern foreign language instruction, as well as for material development and learning system design.

Challenges in Foreign Language Production and Speaking

In foreign language production, particularly during speaking tasks that demand fluency and immediacy, learners face several intrinsic challenges at both psychological and cognitive levels.

First, there is the issue of learners' psychological barriers and the accompanying "sense of incompleteness." Psychological factors such as anxiety and lack of confidence can pose significant barriers that hinder both language processing and production. Particularly among English as a Foreign Language (EFL) learners, such as Japanese university students, an excessive fixation on accuracy—and the pervasive fear of making mistakes—raises high psychological hurdles during sentence generation and speaking. Furthermore, in conventional classroom settings, time constraints often force tasks to conclude before learners can adequately reflect on or self-correct their imperfect output. The accumulation of this "sense of incompleteness" diminishes learners' motivation and amplifies their apprehension toward speaking.

Second, there is difficulty in connecting "noticing" to "autonomous self-correction." In speaking instruction, presenting immediate, multidimensional feedback—such as vocabulary, grammar/syntax, and organization—for each individual learner is extremely difficult due to structural classroom constraints. Particularly in tasks that involve expressing structured content, it is crucial to direct learners' attention not merely to isolated vocabulary or grammatical corrections, but also to organizational aspects, such as the logical sequence of information and spatial relationships needed to describe a situation objectively. However, simply providing pre-prepared "model answers" often leads learners to perceive them as the

"one and only correct answer." This approach has clear limitations in promoting autonomous noticing, where learners must actively consider how to refine their own original production.

Designing a Learning Cycle Utilizing AI-Generated Sample Answers

To address these psychological and cognitive challenges in speaking outlined above, the educational application of generative AI has garnered significant attention. Generative AI can flexibly produce a wide range of English texts tailored to task contexts and learning goals, making it a powerful tool for supporting autonomous language learning. In language teaching practice, efforts to provide opportunities for language production (in both speaking and writing) are being actively implemented through AI technologies. These include AI-assisted speaking practice using conversational AI tools (Shen et al., 2025) and AI-supported writing instruction (Chen, 2026).

However, researchers have pointed out that relying on AI for direct error correction or immediate provision of correct answers poses a notable challenge: learners tend to accept AI output uncritically, which hinders active revision and deep "noticing." Although these insights stem primarily from second language (L2) writing, the process of monitoring one's language production and underlying knowledge through written form—and subsequently engaging in revision—plays a crucial cognitive role in building the foundation for future speaking proficiency (oral production). Specifically, Chen (2026) highlights that while generative AI can enhance planning and evaluative strategies, easy access to AI validation often leads to a "monitoring displacement" mechanism—where learners outsource real-time self-regulation during drafting, consequently weakening their motivation for internal monitoring. Furthermore, Chen and Liu (2026) caution that when learners use AI as an outcome-oriented tool, they tend to outsource their cognitive evaluation and agency to the system. Over time, this passive reliance diminishes their sense of ownership and can erode their capacity for self-regulation and strategic planning. Consequently, when provided with direct corrections or model answers, learners tend to accept or copy outputs uncritically rather than actively engaging with the feedback.

To address these challenges, Zare et al. (2025) emphasize that providing long-term, scaffolded support to help learners become independent producers—rather than allowing them to rely passively on generative AI—is essential for sustaining motivation and self-regulation. As a concrete approach to task design, Pust (2026) demonstrates that an inductive framework—in which students actively manipulate digital tools to discover language rules and undergo iterative text revision—effectively fosters learner engagement. Furthermore, Michta and Frankenberg-Garcia (2026) demonstrate that, unlike systems that automatically rewrite text, a design that allows users to select suggestions while retaining their agency enables learners to make autonomous choices—such as rejecting unnecessary edits—thereby fostering noticing and vocabulary learning.

Building upon these insights, this study designed a learning cycle using an "image description task" as a concrete means of mitigating the psychological and cognitive challenges learners face in speaking. In this cycle, immediately after learners produce their own English sentence, they are presented with pre-generated "sample answers" by generative AI based on two distinct educational frameworks. These samples serve as a reference for self-correction and re-production.

This approach intentionally adopts a text-based (writing) framework for sentence production, aiming to enhance the quality of "noticing"—the foundation for subsequent speaking—by allowing learners to visually and carefully compare their own English with the AI sample answer. Crucially, the sample answers provided in this cycle are neither passive feedback through direct AI error correction nor fixed, conventional model answers that learners often perceive as the "one and only correct answer." This design acknowledges that AI-generated text is not always optimal, thereby encouraging learners to critically evaluate the provided options rather than passively adopt them.

Positioning controlled AI-generated text as merely "one example" and encouraging learners to reference it autonomously—as a guide to incorporate more appropriate expressions and structures while using their own English as a baseline—is expected to elicit autonomous noticing not only of vocabulary and grammar/syntax, but also of discourse structure and an active revision process. We consider this process of contrasting one's own English with a sample that may contain imperfections, rather than receiving individualized direct correction, to hold significant educational value for autonomous learning. Furthermore, establishing a framework premised on revision rather than expecting flawless output from the start is anticipated to alleviate learners' sense of incompleteness, mitigate anxiety over failure, and foster psychological safety. Task design that accounts for such psychological safety and autonomy may lead learners to positively embrace generative AI as a partner in self-directed learning and cultivate a sustained attitude toward utilizing AI.

Research Objectives

Based on the issues outlined above, the purpose of this study is to implement and evaluate a "sentence production learning cycle" among Japanese university students to foster future speaking proficiency. This cycle incorporates self-correction through autonomous reference to pre-generated AI sample answers in an image description task. Specifically, this study addresses the following two Research Questions (RQs):

RQ1: How does a learning cycle premised on revision, in which learners autonomously reference and compare AI-generated "sample answers (as one of many possibilities)," affect the alleviation of learners' sense of incompleteness and the cultivation of psychological safety toward sentence production?

RQ2: How do learners perceive the utility of AI sample answers across linguistic and structural dimensions (vocabulary, grammar/syntax, and organization), and how does this perception contribute to their linguistic noticing and expectations for further generative AI-assisted learning?

Connecting linguistic utility with broader learning outcomes in RQ2 is essential because evaluating the perceived utility across specific linguistic dimensions serves as the immediate cognitive trigger that drives learners' linguistic noticing and shapes their subsequent expectations for AI-assisted learning. Through these investigations, this study seeks to explore the potential of leveraging generative AI to support autonomous learning and offer concrete implications for building a foundation for future speaking proficiency.

Methods

Participants

The participants in this study were 129 first-year Japanese university students. Based on the specific types of AI-generated sample sentences provided, they were assigned to two independent groups under different experimental conditions: Group A (n = 65) and Group B (n = 64).

Tasks and Procedures

In this study, an "English picture description task," which required participants to describe AI-generated illustrations in English, was conducted across two separate sessions. The illustrations depicted scenes of a "park" and a "supermarket," respectively. Both the task illustrations and the AI-generated sample sentences were generated using Gemini 3.6 Flash. To address the purpose of this study, the tasks were carried out according to the three-step "sentence production learning cycle" outlined below:

1. **Independent English Production**: After viewing the presented illustration, participants first produced (wrote) their initial descriptive sentences entirely on their own.
2. **Reference to Sample Sentences**: Next, pre-controlled sample sentences generated by Gemini 3.6 Flash were presented, and participants referred to and compared them with their own sentences. Through specific prompt instructions, these sample sentences were developed based on two different educational frameworks: the CEFR A2 level and the first-year high school level. Specifically, one framework was based on the Common European Framework of Reference for Languages (CEFR) A2 level, an international language proficiency standard, while the other was aligned with the "first-year high school level" from the domestic national curriculum guidelines. The assignment for each group and session is presented in Table 1.
3. **Self-Revision and Reproduction**: Based on the insights gained from referencing the sample sentences, participants revised their initial sentences and produced the final version of their English text.

Table 1

The Assignment of AI-Generated Sample Sentences for Each Group and Session

Session (Topic)	Group A (n = 65)	Group B (n = 64)
Session 1 (Park)	1st-year High School level	CEFR A2 level
Session 2 (Supermarket)	CEFR A2 level	1st-year High School level

Note. n = number of participants.

Within this learning cycle, the sentences provided by the generative AI were not positioned as a "model answer" that learners were expected to imitate word for word. Instead, they were treated as one possible "sample answer" generated by AI for the same illustration. While using their initial draft as the primary foundation, participants were instructed to actively and autonomously reference this sample text as a comparative resource to incorporate more appropriate expressions and structures.

Data Collection and Analysis

To investigate the research questions, two distinct types of questionnaire surveys were administered: one immediately after each task, and a post-activity questionnaire after completing both sessions. First, immediately following each description task, participants evaluated the usefulness of the presented sample answer. Specifically, they rated the extent to which the sample answer served as a helpful reference for their self-revision across three distinct domains—lexical, grammatical/syntactic, and organizational—using a 5-point Likert scale (see Table 2).

Furthermore, after completing both learning sessions, a post-activity questionnaire was administered to comprehensively measure the overall influence of this learning cycle on the learners. This questionnaire consisted of two primary domains—psychological/attitudinal aspects and learning/linguistic aspects—and similarly utilized a 5-point Likert scale to assess participants' perceptions (see Table 3).

For the statistical analysis, the Wilcoxon rank-sum test was used to investigate whether statistically significant differences existed between the ratings of the two groups.

Table 2

Questionnaire Items Administered Immediately After Each Task

Usefulness of Sample Sentences for Self-Revision
Q1. Domain: Lexical When revising your own English sentences, did you find the use of words and vocabulary expressions in the sample answer helpful? 1. Not helpful 2. Not very helpful 3. Neutral 4. Somewhat helpful 5. Helpful
Q2. Domain: Grammatical/Syntactic When revising your own English sentences, did you find the use of grammar and syntactic structures in the sample answer helpful? 1. Not helpful 2. Not very helpful 3. Neutral 4. Somewhat helpful 5. Helpful
Q3. Domain: Organizational When revising your own English sentences, did you find the overall structure and text organization of the sample answer helpful? 1. Not helpful 2. Not very helpful 3. Neutral 4. Somewhat helpful 5. Helpful

Table 3
Post-activity Questionnaire Items

Learners' Perceptions of the Sentence Production Learning Cycle
Domain: Psychological /Attitudinal
Q1. The method of revising and resubmitting my initial drafts based on the sample answers was preferable to simply creating and submitting them.
1. Disagree 2. Somewhat disagree 3. Neither agree nor disagree 4. Somewhat agree 5. Agree
Q2. It was beneficial that the process of revision and resubmission prevented the learning task from ending with the submission of incomplete sentences.
1. Disagree 2. Somewhat disagree 3. Neither agree nor disagree 4. Somewhat agree 5. Agree
Domain: Learning /Linguistic
Q3. Revising my drafts based on the sample answers helped me notice vocabulary and expressions that I could not initially produce on my own.
1. Disagree 2. Somewhat disagree 3. Neither agree nor disagree 4. Somewhat agree 5. Agree
Q4. Using my own drafts as a foundation for revision, rather than merely reading the sample answers, will contribute to the improvement of my English skills.
1. Disagree 2. Somewhat disagree 3. Neither agree nor disagree 4. Somewhat agree 5. Agree
Q5. It would have been better if similar illustrations and sample answers explaining them had been provided before I created my initial drafts.
1. Disagree 2. Somewhat disagree 3. Neither agree nor disagree 4. Somewhat agree 5. Agree

Results

Effects on Psychological Aspects (RQ1)

Regarding RQ1, the post-activity questionnaire results (Table 4) revealed that participants in both groups responded favorably to the proposed learning cycle. Specifically, over 90% of the learners in both groups evaluated the revision-and-resubmission cycle positively.

For Q1, which evaluated the method of revising and resubmitting initial drafts based on sample answers rather than simply submitting them, the vast majority of participants in both groups agreed. Specifically, 25% of the participants in Group A selected "Somewhat agree," and 69% selected "Agree," totaling 94%. In Group B, 52% selected "Somewhat agree," and 48% selected "Agree," representing a combined total of 100% (Group A: $M = 4.6$, $SD = 0.8$; Group B: $M = 4.5$, $SD = 0.5$).

Similarly, for Q2, which addressed the benefit of preventing the learning task from ending with the submission of incomplete sentences, a substantial majority in both groups viewed the cycle favorably. Specifically, 25% of the participants in Group A selected "Somewhat agree," and 74% selected "Agree," totaling 99%. In Group B, 36% selected "Somewhat agree," and 64% selected "Agree," representing a combined total of 100% (Group A: $M = 4.7$, $SD = 0.6$; Group B: $M = 4.6$, $SD = 0.5$).

Table 4
Participants' Responses to Post-activity Questionnaire Items

Group A	1. Disagree	2. Somewhat disagree	3. Neither agree nor disagree	4. Somewhat agree	5. Agree
Q1	1.5%	3%	1.5%	25%	69%
Q2	1%	0%	0%	25%	74%
Q3	2%	0%	0%	18%	80%
Q4	2%	0%	3%	37%	58%
Q5	0%	20%	8%	38%	34%

Group B	1. Disagree	2. Somewhat disagree	3. Neither agree nor disagree	4. Somewhat agree	5. Agree
Q1	0%	0%	0%	52%	48%
Q2	0%	0%	0%	36%	64%
Q3	0%	0%	0%	19%	81%
Q4	0%	0%	8%	42%	50%
Q5	0%	16%	20%	34%	30%

Linguistic and Organizational Utility (RQ2)

Regarding RQ2, data obtained from both the questionnaire items administered immediately after each task and the post-activity questionnaire demonstrated that the participants deeply recognized the linguistic and organizational benefits of the proposed learning cycle.

Regarding the utility of the sample answers, the immediate questionnaire results indicated a high evaluation across all three targeted aspects—lexical (vocabulary), grammatical/syntactic, and organizational (discourse structure). The mean scores exceeded 4.4 on a 5-point Likert scale across both groups and tasks, ranging from 4.6 to 4.8 in Group A and from 4.4 to 4.7 in Group B (Table 5). While no statistically significant differences were found between the two groups regarding the lexical and grammatical/syntactic aspects at the 0.05 significance level, a significant difference between Groups A and B emerged in the organizational aspect for the specific "park" task (Wilcoxon rank-sum test, $p = .019$).

Table 5
Participants' Responses to the Questionnaires Administered Immediately After Each Task

Park		1. Not helpful	2. Not very helpful	3. Neutral	4. Somewhat helpful	5. Helpful
lexical	Group A	3%	0%	0%	12%	85%
	Group B	5%	0%	2%	23%	70%
grammatical/	Group A	5%	0%	2%	18%	75%
syntactic	Group B	4.7%	0%	7.8%	25%	62.5%
organizational	Group A	3%	0%	0%	15%	82%
	Group B	6%	0%	8%	22%	64%

Supermarket		1. Not helpful	2. Not very helpful	3. Neutral	4. Somewhat helpful	5. Helpful
lexical	Group A	0%	0%	0%	18%	82%
	Group B	0%	2%	0%	28%	70%
grammatical/ syntactic	Group A	0%	0%	2%	29%	69%
	Group B	0%	1.6%	4.7%	31.2%	62.5%
organizational	Group A	0%	0%	3%	31%	66%
	Group B	0%	1.6%	1.6%	20%	76.8%

A comparative analysis of the two sample English texts describing the park scene (Table 6) identified clear structural differences between the conditions. The Group A sample utilized a systematic spatial progression—guiding the reader's focus seamlessly from the overall scene to the middle, left, and right areas using explicit locative markers. In contrast, while the Group B sample also adopted a standard top-down organization and attempted spatial framing using "In the middle," its spatial signposting remained limited. Rather than guiding the reader across the entire scene, it concentrated various elements—such as the playground, path, benches, trash can, and flowers—within a single central zone. Consequently, the description of this central area relied primarily on listing specific amenities through repetitive structures (e.g., "There is/are..."), ultimately resulting in a feature-centric listing rather than a systematic spatial progression. Specifically, in Group A, the inclusion of explicit spatial markers—such as "in the middle," "to the left," or "on the right side"—was found to be highly instrumental in guiding the learners' paragraph organization. This suggests that textual structures explicitly highlighting spatial arrangements are perceived as more informative and useful by learners when organizing their descriptive writing. Explaining these structural differences is pedagogically meaningful because distinct framework designs can direct learners' attention to specific linguistic features during self-correction before examining their overall impact on learners.

Furthermore, the post-activity questionnaire provided deeper insights into the learners' awareness of their improvement and the task design. For Q3 in Table 4, which examined whether revising their drafts based on the sample answers helped learners notice vocabulary and expressions they could not initially produce on their own, a substantial majority in both groups responded positively. Specifically, 18% of the participants in Group A selected "Somewhat agree," and 80% selected "Agree," totaling 98%. In Group B, 19% selected "Somewhat agree," and 81% selected "Agree," representing a combined total of 100% (Group A: $M = 4.8$, $SD = 0.6$; Group B: $M = 4.8$, $SD = 0.4$). Similarly, for Q4 (Table 4), which evaluated the effect of using their own sentences as a foundation for revision, rather than merely reading the sample answers, the participants reported strong perceived benefits, recognizing it as an effective way to improve long-term English skills. Specifically, 37% of the participants in Group A selected "Somewhat agree," and 58% selected "Agree," totaling 95%. In Group B, 42% selected "Somewhat agree," and 50% selected "Agree," representing a combined total of 92% (Group A: $M = 4.5$, $SD = 0.7$; Group B: $M = 4.4$, $SD = 0.6$).

Interestingly, the responses to Q5 (Table 4) revealed a clear learner preference regarding the task sequence. Approximately 60% to 70% of the participants in both groups (72% in Group A and 64% in Group B) selected either "Somewhat agree" or "Agree," indicating a distinct preference for having similar illustrations accompanied by sample answers provided as

references prior to creating their initial drafts (Group A: $M = 3.9$, $SD = 1.1$; Group B: $M = 3.8$, $SD = 1.0$).

Table 6

Comparison of AI-Generated Sample Texts for the Park Scene Task

Group A: 1st-Year High School Level	Group B: CEFR A2 Level
This is a beautiful and sunny park scene. The image shows a playground and a nice, green area with lots of trees. **In the middle**, there is a winding dirt path. On the path's edge, you can see a wooden bench, and another bench is further back. **To the left**, there's a big tree and a trash can. **On the right side**, there is a play structure with an orange slide and a set of swings. The sky is blue with some fluffy white clouds. Overall, the picture is colorful and peaceful, like a perfect day outside.	This is a picture of a nice park on a sunny day. I can see a lot of green trees and grass. The sky is light blue and there are a few white clouds. **In the middle**, there is a playground. It has a slide that is orange and a swing set that is blue. There is a brown path that goes through the grass. I also see two wooden benches where people can sit and a black trash can. There are some yellow flowers near the front bench. It looks like a good place to play and relax.

Discussion

Psychological Comfort and Safety (RQ1)

Regarding RQ1, the post-activity questionnaire results suggested that over 90% of the participants evaluated the proposed learning cycle very positively. This highly positive self-reported response may not be solely attributable to the novelty or surface-level appeal of using generative AI. Instead, it implies a potential underlying psychological factor associated with the specific task design.

Furthermore, the explicit structure of repeated revision may have helped mitigate learners' anxiety regarding failure. In conventional production tasks, learners often experience a sense of incompleteness or frustration when they cannot express their ideas perfectly on their initial attempt. However, by positioning the sample answers not as an absolute standard of correctness but as a supportive resource for secondary revision, the current learning cycle may have helped reduce this psychological burden. This design could have fostered a sense of psychological safety, potentially transforming hesitation into a constructive noticing experience and enabling learners to engage with the tasks with less apprehension about making mistakes.

Textual Scaffolding and Learners' Insights (RQ2)

Regarding RQ2, the quantitative analysis of the perceived usefulness of the AI-generated sample answers indicated consistently high ratings. Mean scores exceeded 4.4 on a 5-point Likert scale across all criteria ("vocabulary," "grammar/syntax," and "organization") for both

groups and tasks. However, while no statistically significant differences were observed between the groups in terms of "vocabulary" and "grammar/syntax," a statistically significant difference was found solely in "organization" within a specific task (the "park" task).

A qualitative inspection of the specific text characteristics provided in that task offers a plausible explanation for this variance. The target task involved a descriptive scene of a park, where providing precise spatial descriptions could render the explanation clearer and more coherent. Notably, the sample provided to Group A for the park scene prominently incorporated explicit spatial markers—such as "in the middle," "to the left," and "on the right side"—guiding the reader through a systematic spatial progression. In contrast, while the sample for Group B attempted spatial framing using "In the middle," its spatial signposting remained limited, concentrating various elements within a single zone and ultimately resulting in a more feature-centric listing. The explicit presentation of spatial arrangements in Group A may have functioned as a more effective textual scaffolding. Consequently, this comprehensive visual-spatial support likely provided learners with a clearer structural framework, which may explain why a statistically significant difference was observed specifically in their evaluation of discourse organization.

Furthermore, the survey findings provide valuable insights for future task designs incorporating generative AI. Over 90% of the participants reported that they did not treat the AI samples as mere "model answers" to copy blindly. Instead, they used them as a reference to refine their original writing, which may have enhanced their sense of learning ownership and self-improvement. Additionally, approximately 60% to 70% of the participants in both groups expressed that having access to similar illustrations and corresponding AI-generated sample texts as a pre-task reference—which was not provided in the current study—would have provided useful support. This preference for pre-task materials suggests that integrating complementary visual and textual resources beforehand could potentially enhance the scaffolding effects of AI-assisted language learning environments.

Conclusion

Summary of Findings

This study investigated the potential psychological and pedagogical scaffolding effects of integrating AI-generated resources, specifically descriptive illustrations and corresponding sample texts, into university-level English production tasks. Regarding RQ1, the findings suggest that the pedagogical design incorporating repeated revision appears to have helped mitigate learners' anxiety regarding failure. By positioning the AI samples not as an absolute model but as a supportive resource for secondary revision, this learning cycle likely fostered a sense of psychological safety. Consequently, it may have transformed potential anxiety into a constructive noticing experience, thereby enabling learners to engage with the production tasks with reduced apprehension about making mistakes.

For RQ2, the quantitative evaluation revealed that both groups perceived the AI samples as highly useful across all domains—including vocabulary, grammar, and organization—yielding mean scores exceeding 4.4 out of 5 in both tasks. Notably, while no significant differences were observed between the groups in vocabulary and grammar, a statistically significant difference emerged in the domain of organization within a specific descriptive task. Textual analysis suggested that the inclusion of explicit spatial markers (e.g., "in the

middle," "to the left," or "on the right side") in the AI texts may have served as effective textual scaffolding, which may explain this specific variance.

Furthermore, the survey findings suggested a high degree of learner autonomy, as over 90% of the participants reported that they did not treat the AI samples as a mere template for blind copying. Instead, they perceived them as a reference to refine their original writing, potentially enhancing their sense of learning ownership.

Pedagogical Implications

The findings of this study offer important pedagogical insights for integrating generative AI into foreign language education, particularly regarding the instructional design of production tasks. First, the study suggests the potential benefits of shifting the instructional focus from product-oriented AI evaluation to process-oriented psychological scaffolding. Rather than utilizing AI primarily as an automated grader or a source of instant correction, educators might consider adopting a structured revision cycle that incorporates AI-generated samples as supportive resources. By framing AI-generated texts as a flexible reference rather than an absolute model, this approach may help lower learners' affective filters. Consequently, such a design could foster a supportive environment in which students feel safer to engage with the target language, potentially transforming errors into opportunities for constructive noticing and linguistic growth.

Second, the results suggest that the efficacy of AI scaffolding may be closely linked to the specific structural and linguistic design of the AI-generated sample texts. Rather than merely presenting generic AI-generated models, educators could deliberately construct or prompt AI samples that contain explicit text-structural and discourse-level cues (e.g., spatial markers) to guide learners. Furthermore, an instructional design that encourages students to reflect on and refine their own production based on these AI references may further promote learner autonomy. When structured in this manner, tasks have the potential to enable AI samples to serve as a supportive scaffold for self-directed language learning.

Limitations and Recommendations for Future Research

Although these findings offer meaningful insights into AI-supported production tasks, several limitations should be acknowledged. First, the participants in this study were limited to Japanese university freshmen from a single institution, representing a relatively narrow range of proficiency. Therefore, the generalizability of these findings to learners with different cultural backgrounds, age groups, or proficiency levels remains to be further explored. Future research incorporating a more diverse sample of participants would be beneficial to examine the extent to which the proposed task design can be applied across varied educational contexts.

Second, this study focused primarily on immediate psychological and behavioral responses within a set of structured tasks. Longitudinal investigations are needed to clarify whether the observed reductions in affective filters and proactive revision behaviors can be sustained over an extended period.

Third, the evaluation in this study relied primarily on learners' subjective perceptions and self-reported data. While these insights are invaluable for understanding affective variables, future research could benefit from incorporating objective measures—such as evaluating

actual pre- and post-task linguistic accuracy, syntactic complexity, and lexical range—to investigate the potential impact of this AI-scaffolded approach on learners' actual language development.

Finally, while the quantitative findings suggested that students tended to actively revise their drafts rather than blindly copy the AI references, qualitative investigations—such as analyzing process log data or conducting retrospective interviews—could provide more nuanced insights into the underlying cognitive processes through which learners interact with and internalize the AI-generated structural and linguistic cues.

Acknowledgments

This work was supported by JSPS Grant-in-Aid for Scientific Research (C) (Grant No. JP25K06595).

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the authors used Gemini 3.6 Flash to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. Additionally, generative AI technology (Gemini 3.6 Flash) was utilized as part of the research methodology to generate the sample English texts and images analyzed in this study; specific details are described in the Tasks and Procedures subsection of the Methods section.

References

Chen, M.-R. A. (2026). Generative AI, visual scaffolds, and metacognition in EFL writing. *Language Learning & Technology*, *30*(1), 1–25. https://doi.org/10.64152/10125/73695

Chen, P.-H. P., & Liu, Y. (2026). Distributed agency in AI-assisted L2 writing. Language Learning & Technology, *30*(1), 1–20. https://doi.org/10.64152/10125/73674

Michta, T., & Frankenberg-Garcia, A. (2026). The flow must go on: Evaluating the impact of a collocation writing assistant on self-revision. *ReCALL*, *38*(1), 57–74. https://doi.org/10.1017/S0958344025100293

Pust, D. (2026). Learning grammar the explorative way: Integrating interactive grammar animations into CALL. *Language Learning & Technology*, *30*(1), 1–21. https://doi.org/10.64152/10125/73667

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Shen, J., Qian, B., Ma, Q., & Deris, F. D. (2025). Exploring the impact of AI-assisted speaking practice on EFL learners' speaking proficiency and motivation: A mixed-method study. *International Journal of Academic Research in Progressive Education and Development*, *14*(4), 43–52. http://dx.doi.org/10.6007/IJARPED/v14-i4/26672

Zare, J., Al-Issa, A., & Ranjbaran Madiseh, F. (2025). Interacting with ChatGPT in essay writing: A study of L2 learners' task motivation. *ReCALL*, *37*(3), 385–402. https://doi.org/10.1017/S0958344025000035

Contact email: kasiwagi@kobe-u.ac.jp