


Effects of three levels of AI integration on second language academic writing: Evaluating restricted, guided, and free use of ChatGPT

Seungjin Hong^a, Yu Kyoung Shin^{b,*} 

^a Institute for Language and Information, Sogang University, Seoul, Republic of Korea

^b School of Global Studies, Hallym University, Chuncheon, Gangwon-do, Republic of Korea

ARTICLE INFO

Keywords:

Instructed SLA
Generative AI integration
English learner
L2 writing instruction
Instructional scaffolding

ABSTRACT

This study evaluates the impact of different ChatGPT usage protocols as instructional tools for academic writing among L2-English learners by exploring how three levels of AI integration affect learning outcomes. Three groups of Korean college students ($n = 103$) were taught by the same instructor in English essay writing courses: The Restricted group did not use ChatGPT, the Guided group accessed it only during supervised in-class activities, and the Free group had unrestricted access to ChatGPT throughout the semester. The results indicate that both the Guided and Free groups consistently outperformed the Restricted group, demonstrating the educational benefits of AI engagement as an innovative approach to language learning. A close examination of essay scores finds that the Guided group excelled in content quality, although there were no significant differences in organization and language use between the Guided and Free groups. This suggests that while structured guidance is crucial for content development, more autonomy may be advantageous in other areas of writing. An interview with the instructor provides further insights into how different levels of autonomy in engagement with AI affected student engagement, and suggests the benefits of strategically integrating AI to improve L2 academic writing skills.

1. Introduction

Artificial intelligence (AI) has emerged as a transformative force in education, attracting growing interest for its ability to reshape conventional language learning practices. In second language (L2) classrooms, AI-driven tools are being adopted at a rapid pace, yet robust empirical evidence supporting their effectiveness remains limited (e.g., Barnes & Tour, 2025; Won et al., 2025). Much of the existing research focuses on speculative arguments or untested suggestions rather than data-driven analysis (e.g., Jin & Fan, 2023; Pack & Maloney, 2023; Praphan & Praphan, 2023; Warschauer et al., 2023).

The limited body of empirical research available often relies on basic binary comparisons between students using AI tools and those engaged in traditional, human-centered instruction (e.g., Banihashem et al., 2024; Escalante et al., 2023; Ghafouri, 2024; Kim & Su, 2024; Steiss et al., 2024; Teng, 2024). Other studies focus exclusively on AI-assisted groups, offering no comparative framework (e.g., Koltovskaia et al., 2024). Such studies typically assume uniform AI exposure across participants, neglecting individual differences and specific learning needs. Although this approach has provided some foundational insights, it cannot fully capture the diverse outcomes

* Corresponding author. School of Global Studies, Hallym University, Chuncheon-si, Gangwon-do 24252, Republic of Korea.
E-mail address: yshin@hallym.ac.kr (Y.K. Shin).

that AI integration might achieve in a variety of learning environments.

Addressing these gaps, this study explores the effects of varying degrees of generative AI integration in L2 classroom settings by comparing the outcomes of three groups of students allowed restricted, guided, or free use of the latest version of ChatGPT for academic writing tasks. By adjusting the extent of this cutting-edge AI tool usage in L2 writing classes, this research is expected to provide a more comprehensive understanding of how different levels of ChatGPT access can help educators tailor strategies to enhance L2 learning processes. The findings should offer insights that go beyond simple AI versus non-AI comparisons, potentially revealing how the strategic application of advanced AI can be optimized to significantly improve learning outcomes.

2. Literature review

2.1. L2 writing as a social and contextual process

Writing in a second language is widely recognized as a process oriented and socially influenced activity, involving not only individual composition skills but also the tools, goals, and collaborative interactions that support learning. This view underscores the importance of multiple factors that shape writing development, such as peer feedback, classroom norms, and access to instructional support.

This perspective aligns with the Writer(s) Within Community (WWC) model proposed by [Graham \(2018\)](#), which conceptualizes writing as a socially situated activity shaped by the goals, tools, and interactions of the community in which the writer participates. Rather than treating writing as an isolated mental task, the model presents it as a dynamic exchange between the individual and the surrounding social and material environment. The model emphasizes how tools such as genre knowledge, instructional feedback, and collaborative dialogue mediate writing development, and how writers' goals and identities evolve depending on the opportunities and constraints of their writing environment. This interpretation is supported by earlier studies that highlight the role of context and interaction in writing instruction (e.g., [Langer, 2002](#); [Newell et al., 2015](#); [Prior, 2006](#)). Classroom practices aligned with the WWC framework have been observed in both first and second language settings, particularly in studies where students participate in shared writing practices, engage in dialogue, and reflect on audience and purpose (e.g., [McCarthy & Mkhize, 2013](#); [Newell et al., 2011](#); [Pu & Xu, 2024](#); [Wang, 2017](#)). It offers a valuable lens for examining how writing communities, learning tools, and instructional practices contribute to the development of writing proficiency.

Building on this understanding of writing as a socially situated process, L2 writing instruction has evolved to prioritize not only the accuracy and appropriateness of written communication but also the interactive processes that enhance learning. Modern pedagogical approaches emphasize the value of interaction and collaboration, both with peers and through digital platforms, as essential components for developing strong writing skills. Many studies on writing pedagogy have demonstrated the benefits of process-oriented and collaborative instruction (e.g., [Banihashem et al., 2024](#); [Hu, 2005](#); [Kang, 2022](#); [Lee et al., 2016](#); [McCarthy & McMahon, 1992](#); [Mortensen, 2018](#); [Pérez-Llantada, 2024](#); [Ruiz-Funes, 2001](#); [Wambsganss et al., 2022](#)). Writing is widely recognized as an iterative process that includes stages such as brainstorming, drafting, peer review, revision, and editing, with peer interaction serving as a key component throughout this cycle.

There is extensive support for integrating peer collaboration into the writing process, as it allows learners to engage with authentic audiences, receive diverse feedback, and build rhetorical awareness. For example, [Zaccaron and Xhafaj \(2020\)](#) compared anonymous and face-to-face peer feedback among L2-English writers and found that while both types were effective, students preferred live discussion for its clarity and collaborative nature. Nevertheless, peer interaction must be carefully structured. In a study of university students, [Shulgina et al. \(2024\)](#) observed that overly abundant feedback led to diminished uptake, suggesting that the quality and manageability of peer comments are key to their effectiveness.

In recent years, computer-mediated tools have significantly expanded the opportunities for interaction in L2 writing instruction. These tools, from online forums to advanced AI systems, offer dynamic spaces for engagement and personalized feedback (e.g., [Chew et al., 2019](#); [Hadwin et al., 2018](#); [Kang, 2022](#); [Kessler & Bikowski, 2010](#); [Lee et al., 2016](#); [Molenaar et al., 2012](#); [Pérez-Llantada, 2024](#); [Wambsganss et al., 2022](#)). Research shows that optimal learning often results from the thoughtful integration of digital platforms with human-centered practices. For instance, [Wambsganss et al. \(2022\)](#) found that automated feedback was more impactful when paired with comparative data on peer performance, reinforcing the influence of social context. In particular, students responded more positively when the feedback included a "social comparison nudge that indicated how other students performed on the same assignment" (p. 1), suggesting that awareness of community standards can motivate reflection and revision.

Taken together, these developments reflect a growing consensus that effective L2 writing instruction requires a balance of structured interaction, contextual support, and access to meaningful tools. Grounded in models such as WWC, this approach emphasizes the importance of both individual agency and community engagement in the development of writing proficiency.

2.2. Advancing AI in language education

Research on the integration of AI into language education suggests its significant potential to transform traditional educational models, but remains in the emergent stage, having provided little empirical evidence of specific benefits from specific applications (e.g., [Barnes & Tour, 2025](#)). Existing studies comparing student performance with and without AI tools have shown mixed results. In some instances, AI has outperformed traditional methods, whereas in others, it has not.

AI tools like ChatGPT, which can provide individually tailored feedback to students, have been reported to lead to improved language learning, especially when combined with teacher and peer feedback. For instance, [Kim and Su \(2024\)](#) conducted a study with

65 L1-Chinese university students learning L2-Korean. One group interacted with AI chatbots, and the other group did not. Kim and Su observed significantly better willingness to communicate, lower anxiety, and greater communication confidence in the chatbot-interaction group than in the control group. In contrast, however, a study by Escalante et al. (2023) found no significant differences between two L2-English learner groups, one of which received feedback from ChatGPT and the other from human tutors. Interestingly, the researchers noted that the learners attributed different specific benefits to each kind of feedback, leading them to support the idea that the two approaches are complementary. In yet another study comparing peer and ChatGPT feedback on learner writing (i.e., argumentative essays), Banihashem et al. (2024) found no association between feedback type and essay quality, but they reported that ChatGPT gave more descriptive feedback, while student peers were better at giving detailed advice to address specific writing problems.

Although these studies offer valuable insights, they commonly employ straightforward binary comparisons, with one group receiving AI intervention and another using traditional methods. Consequently, the scope of current research is narrow, with few investigations into the impact of varying degrees of AI integration on student engagement and learning outcomes. This situation calls for more comprehensive studies that not only assess the immediate benefits of AI tools but also explore the optimal conditions and levels of AI integration to improve educational effectiveness.

A further need is more critical research that addresses the many concerns raised by the use of AI, not just in education generally but in language pedagogy specifically. A major concern among L2 educators is the risk that language learners can easily become over-dependent on technological tools (e.g., Abbas et al., 2024; O'Sullivan, 2023; Voss et al., 2023; Wang et al., 2023; Warschauer et al., 2023). An illuminating study by Abbas et al. (2024) argued that university students who were experiencing high levels of academic pressure relied more heavily on ChatGPT, and that increased reliance on the tool was associated with higher levels of procrastination, forgetfulness, and poor academic performance. In addition, preliminary findings from (Hong & Shin, 2025). (under review) involving L2-English learners who had used ChatGPT underscore several drawbacks, particularly the ethical concerns raised by students themselves. The language learners in that study were worried about how they could balance their desire to produce authentic and original work with the draw to use AI tools to receive, for instance, immediate feedback and targeted suggestions for fixing grammatical and vocabulary errors.

These observations suggest the necessity of tailoring AI integration to meet the specific needs and preferences of individuals so that it enhances rather than replaces student effort. Implementing a personalized approach may help maintain the balance between technological assistance and learner engagement, which has been noted as important in effective language education (Won et al., 2025). Rather than viewing AI as a standalone tool, it may be useful to consider its role within the broader social and instructional context in which writing occurs. When thoughtfully embedded in learning environments that foster interaction, reflection, and shared purpose, such tools may offer meaningful support for writing development.

3. Present study

This study evaluates the impact of varying degrees of ChatGPT usage on English essay writing among Korean college students in naturalistic classroom settings. All participants were taught by the same instructor in three classes of the same course, which focused on academic English essay writing. Different levels of AI interaction resulted in three groups: Restricted AI use, Guided AI use, and Free AI use. The instructional design of each group was informed by the WWC framework (Graham, 2018), which conceptualizes writing as a socially situated and tool-mediated process. Based on this model, the Restricted group received traditional instruction without access to ChatGPT, serving as a baseline. The Guided group engaged in teacher-scaffolded writing, peer collaboration, and structured ChatGPT use, closely aligning with WWC principles of supported writing within a learning community. The Free group, in contrast, used ChatGPT autonomously with minimal guidance, reflecting a more self-directed approach to tool-mediated learning.

The research investigates how different levels of ChatGPT integration affected students' writing performance across three real classroom settings. This study is guided by the following research questions.

1. Are there significant differences in immediate and delayed posttest essay scores among the Restricted, Guided, and Free groups?
2. How does the instructor, who taught all three groups, perceive the integration of ChatGPT in L2 writing instruction?

4. Methodology

4.1. Participants

The study involved 103 participants, all of whom were students enrolled in three general English courses at a university near Seoul, South Korea. Each course represented one of the study's three experimental groups: the Restricted group (32 students; 23 males, 9 females), which had no access to ChatGPT; the Guided group (36 students; 15 males, 21 females), which used ChatGPT only during designated in-class writing activities under strict guidelines; and the Free group (35 students; 25 males, 10 females), which had unlimited access to ChatGPT throughout the course. To ensure instructional consistency, all three courses were taught by the same instructor and held twice weekly in 90-min sessions. All participants were informed of the study's purpose and procedures prior to data collection, and written informed consent was obtained, with participation being entirely voluntary.

4.2. Instructional conditions

The three groups of students participated in a structured learning process, where they first learned how to write paragraphs, then progressed to essays, and finally tackled various types of essays such as descriptive, cause and effect, argumentative, and problem-solving. A key preparatory process shared by all groups included discussions on essay topics to improve understanding of the subject matter, organize ideas effectively, and encourage critical engagement with fellow students or ChatGPT, depending on the group. The instructor introduced each topic and facilitated discussions that allowed students to explore both supporting and opposing viewpoints. As part of this instructional design, the courses were designed for students with low to mid-intermediate English proficiency, focused on L2 essay writing, and were conducted entirely in English. As all instruction, materials, and assignments were in English, students primarily interacted with ChatGPT in the target language, even when prompts were occasionally entered in Korean.

During these pre-writing sessions, the Restricted group engaged exclusively with their classmates, focusing on face-to-face interaction. The Guided group used ChatGPT within structured parameters, receiving specific guidelines and a detailed task handout from the instructor to ensure relevance to their assignments (see [Appendix A](#) for a sample in-class handout). Their ChatGPT use was confined to supervised in-class tasks such as generating counterarguments, with no access permitted outside class. The Free group collaborated with their classmates while also having unrestricted access to ChatGPT, which they used freely in their writing activities. To facilitate this, students accessed ChatGPT using their own devices, primarily laptops or tablets, with a few using smartphones. While no formal logging system was employed, the instructor closely monitored class activities to confirm that students in this group actively engaged with ChatGPT during all writing phases, including brainstorming, outlining, and composing. In all groups, regardless of AI access, following the discussions, the instructor summarized the key points raised and organized them visually on the blackboard to help learners outline their essay structures. An overview of these instructional conditions is provided in [Table 1](#), which clarifies how the level of AI integration and instructional mediation differed across groups.

The course followed a planned timeline, as in [Table 2](#), beginning with a pretest in week 4 to establish a baseline for students' writing abilities. Next, in Weeks 9 and 13, following targeted instruction that included the pre-writing activities mentioned above, they took immediate posttests, where they were given 50 min to complete essays on topics such as "What are the key factors that contribute to an individual's success?"; "Identify a common challenge faced by university students and propose two practical solutions to address it"; and "Agree or disagree: It is better to be a member of a group than to be the leader of a group." For these posttests, the use of external resources, including ChatGPT, was strictly prohibited for the Restricted and Guided groups but allowed for the Free group. The Free group was encouraged to fully leverage this access as preparation for the delayed posttest (Week 15), receiving virtually no instruction on how to use ChatGPT and being challenged to integrate AI tools into their writing independently.

Post-writing sessions included an editing phase where students used editing handouts. These handouts facilitated the evaluation of key essay aspects, such as the clarity of the thesis, the effectiveness of the introduction, topic sentence consistency, and the flow of the conclusion (see [Appendix B](#) for an editing handout). During these sessions, the Restricted group worked alongside classmates and the instructor, creating a traditional learning environment. The Guided group received feedback from either classmates or ChatGPT, customized according to the specific guidelines for each activity. The Free group participated in peer discussions and again was encouraged to make full use of ChatGPT for feedback while receiving no specific guidance on how to do so, to facilitate their autonomous use of AI throughout the editing process.

The semester concluded with the delayed posttest in week 15, during which no group was allowed to use any external resources, including ChatGPT. This standardized testing condition guaranteed a fair comparison of writing skills across all groups in order to reveal the extent to which each group benefited from their respective levels of AI integration throughout the semester.

4.3. Essay scoring procedures

The essay scoring rubric developed by [Kim et al. \(2023\)](#) was employed to assess the participants' essays across three categories: content, organization, and language use. Each category was scored on a scale from 0 (no credit) to 5 (excellent), with a maximum potential total of 15 points per essay. Content scores measure how well an essay addresses the topic and the clarity of the explanations. Organization scores are based on the essay's structural clarity and the coherence of ideas. Language use is evaluated based on grammatical accuracy, syntactic variety, and vocabulary appropriateness.

Each essay test was administered within a 50-min time limit. While no explicit word count was required, students were encouraged to write approximately one page to ensure sufficient development of their ideas. Essay scoring was conducted independently by two

Table 1
Instructional design.

Group	ChatGPT Access	Instructional Focus
Restricted	No access	Traditional instruction; peer review and teacher feedback
Guided	Guided access	Teacher-guided ChatGPT use during structured tasks only (e.g., pre-writing; see Appendix A); other activities mirrored those of the Restricted group
Free	Unrestricted access	Autonomous ChatGPT use across all writing stages; self-directed learning

Note. The sequence, content, and writing tasks were consistent across all groups. However, the type of mediation varied, with AI support provided in the Guided and Free groups and human-only interaction in the Restricted group.

Table 2
Overview of the study procedure.

Week	Session details
Week 4	Pretest (50 min)
Week 9	Pre-writing discussion on a topic Posttest 1 (50 min)
Week 13	Post-writing editing task Pre-writing discussion on a topic Posttest 2 (50 min)
Week 15	Post-writing editing task Delayed posttest (50 min)

researchers (including the instructor) and a native English-speaking undergraduate majoring in English Education, all of whom had relevant training and experience. Inter-rater reliability was calculated using the holistic scores for the three categories of content, organization, and language use based on the standardized rubric described above. Agreement was determined by the percentage of identical scores assigned across raters, with an overall agreement rate exceeding 90 %. Any discrepancies were resolved through discussion.

4.4. Statistical analyses

To examine the effects of varying levels of ChatGPT access on students' English writing performance, a series of analyses of covariance (ANCOVAs) were conducted. This analytic approach was selected to statistically control for students' baseline writing proficiency, using their pretest scores as a covariate in all subsequent comparisons. By adjusting for these initial differences, the ANCOVAs provided a more accurate estimation of the effect of the treatment conditions.

Prior to conducting the ANCOVAs, assumptions of normality, homogeneity of variances, and homogeneity of regression slopes were tested. The normality of residuals was assessed using the Shapiro-Wilk test, which was non-significant ($W = 0.986$, $p = .3545$), indicating that the residuals followed a normal distribution. Thus, the assumption of normality was met. To assess the reliability of all the tests, Cronbach's alpha was computed. The obtained alpha value of 0.77 indicates an acceptable level of internal consistency, confirming that the tests were consistently reliable and comparable in difficulty.

Separate ANCOVAs were conducted for total scores on both the immediate and delayed posttests. Further ANCOVAs were performed for each of the three rubric categories: content, organization, and language use. In all cases, the pretest writing scores were included as a covariate. When a statistically significant main effect of treatment condition was found, pairwise comparisons were conducted using Bonferroni-adjusted post hoc tests to identify specific group differences.

4.5. Follow-up interview with the instructor

The instructor, who taught all three groups, was interviewed to gather her insights into the effectiveness of AI integration into this writing course. The instructor had years of experience teaching L2-English writing to Korean college students in a traditional manner focused on human interaction, and had recently begun integrating AI tools like ChatGPT into her instruction through a blended approach. The semi-structured interview was centered around a set of predefined questions focusing on the instructor's observations across the different levels of AI interaction among the groups. Key questions inquired into differences in student engagement and writing quality across the groups, specific instances where AI integration significantly impacted a student's performance for better or worse, and the instructor's perspective on the primary benefits and drawbacks of using AI tools like ChatGPT in academic writing courses. This interview was designed to provide qualitative insights that would complement the quantitative data from the study, offering a more comprehensive understanding of the practical impacts of AI support on L2 writing instructions.

5. Results

5.1. Essay scores

This section presents the study's results by examining how varying levels of ChatGPT access affected students' essay performance. Initial pretest results revealed no significant differences among the three groups ($F = 0.79$, $df = 2$, $p = .46$) as shown in Table 3, indicating that all students started with similar baseline English writing abilities.

Table 3
One-way ANOVA results of pretest writing score.

Predictor	SS	Df	MS	F	p	η^2
Treatment	5.32	2	2.66	0.79	0.46	0.02
Residuals	337.28	100	3.37			
Total	342.6	102				

To examine group differences in immediate posttest scores while controlling for baseline writing ability, an ANCOVA was conducted using pretest scores as a covariate. The results revealed a significant main effect of treatment on total immediate posttest scores ($F(2, 99) = 34.02, p < .001$, partial $\eta^2 = 0.407$), and a significant effect of the covariate ($F(1, 99) = 21.75, p < .001$, partial $\eta^2 = 0.180$). The results demonstrate that ChatGPT access significantly enhanced students' immediate posttest writing performance, independent of their initial proficiency levels. Although baseline writing ability emerged as a significant predictor of posttest outcomes, the treatment condition continued to explain a considerable portion of the variance. This suggests that the intervention exerted a meaningful and robust influence on students' writing development, even after controlling for prior differences in ability.

Post hoc comparisons using Bonferroni-adjusted pairwise tests revealed significant differences among all three conditions. Participants in the Free group scored significantly higher than those in both the Guided and Restricted groups ($p < .001$), and the Guided group also outperformed the Restricted group ($p < .001$) as shown in Table 4. This outcome was anticipated, given their open use of AI assistance throughout the writing process, in contrast to the other two groups. Thus, particularly noteworthy is the performance of the Guided group. Despite having only limited and structured access to ChatGPT, this group significantly outperformed the Restricted group, which had no access to ChatGPT at all. This finding illustrates that even a controlled application of AI can greatly benefit students.

Further ANCOVAs were conducted separately for each rubric category (content, organization, and language use), while controlling for pretest scores. As shown in Table 5, the results revealed significant main effects of treatment across all three dimensions ($ps < 0.001$). Specifically, the effect of treatment was significant for content ($F(2, 99) = 19.12, p < .001$, partial $\eta^2 = 0.278$), organization ($F(2, 99) = 22.35, p < .001$, partial $\eta^2 = 0.311$), and language use ($F(2, 99) = 20.75, p < .001$, partial $\eta^2 = 0.296$). These findings indicate that access to ChatGPT, whether guided or unrestricted, substantially influenced students' writing performance across all rubric components, even after accounting for initial writing ability. Adjusted mean scores and corresponding post hoc comparisons are also summarized in Table 6.

However, post hoc comparisons revealed more differentiated patterns of effect. In the content, students in the Guided group ($M = 4.10, SE = 0.13$) significantly outperformed both the Free ($M = 3.60, SE = 0.15$) and Restricted ($M = 2.95, SE = 0.16$) groups, with $p < .001$ for both comparisons (see Table 7). Notably, there was no significant difference in content scores between the Free and Restricted groups ($p = .09$), suggesting that the use of ChatGPT without structured support may not be sufficient to enhance content development. By contrast, students in the Guided condition, who received such support, showed significantly better performance in this domain.

Meanwhile, for both organization and language use, participants in the Free and Guided groups performed significantly better than those in the Restricted group ($p < .001$), with no significant differences observed between the Free and Guided groups ($p > .05$). These results indicate that autonomous access to ChatGPT was sufficient to yield improvements in students' organizational and linguistic abilities.

Delayed posttest scores were analyzed to determine the lasting effects of AI-assisted writing, controlling for pretest performance. The analysis revealed a significant main effect of treatment on total delayed posttest scores, $F(2, 99) = 21.08, p < .001$, partial $\eta^2 = 0.299$, as well as a significant covariate effect, $F(1, 99) = 37.17, p < .001$, partial $\eta^2 = 0.273$. These findings confirm that, even after statistically controlling for students' initial writing proficiency (which did not differ significantly across groups at the pretest stage), the treatment condition remained a strong predictor of delayed posttest performance. Interestingly, despite having the lowest pretest average ($M = 7.33$), the Guided group achieved the highest posttest scores, further underscoring the effectiveness of structured AI integration. The significance of the covariate reflects the expected trend: students with stronger initial skills tended to perform better, regardless of group assignment. Nonetheless, the slightly stronger effect of treatment compared to prior ability highlights the influence of AI use, whether structured or unstructured, on students' writing development over time.

As shown in Table 8, the Guided group achieved the highest adjusted mean on the delayed posttest ($M = 10.87$), despite entering the study with the lowest average pretest score. This suggests that instructional guidance – combining ChatGPT with structured teacher-led materials – facilitated more sustained writing development. Post hoc pairwise comparisons further revealed that both the Guided and Free groups significantly outperformed the Restricted group ($ps < 0.001$), but no significant difference was found between the Free and Guided groups ($p = .60$). This pattern implies that while structured support yielded the highest scores, unstructured use of ChatGPT still led to measurable long-term benefits.

Category-specific ANCOVAs provided further insight into these effects. In terms of content, the Guided group again significantly outperformed both the Free and Restricted groups ($ps < 0.001$), while the Free and Restricted groups did not differ significantly ($p = .09$), as summarized in Table 9. This finding suggests that the structured integration of ChatGPT was particularly effective in enhancing and sustaining students' content development over time. In contrast, unstructured AI access in the Free group did not lead to significantly better content understanding compared to the Restricted group, which had no access to AI support.

The superior performance of the Guided group could be attributed to the interactive approach that combined the use of ChatGPT with meticulously designed handouts, which included specific examples and contrasting scenarios relevant to the essay topic. All three

Table 4
ANCOVA results for immediate posttest total scores (controlling pretest scores).

Predictor	SS	df	MS	F	p	η^2
Treatment	415.92	2	207.96	34.02	<0.001	0.407
Pretest	265.50	1	265.50	21.75	<0.001	0.180
Residuals	608.12	99	6.14			
Total	1289.54	102				

Table 5
ANCOVA results for immediate posttest scores by rubric categories.

Rubric	Predictor	SS	df	MS	F	<i>p</i>	η^2
Content	Treatment	82.34	2	41.17	19.12	<0.001	0.278
	Pretest	35.42	1	35.42	16.44	<0.001	0.143
	Residuals	213.15	99	2.15			
Organization	Treatment	78.65	2	39.33	22.35	<0.001	0.311
	Pretest	29.83	1	29.83	16.96	<0.001	0.146
	Residuals	174.24	99	1.76			
Language Use	Treatment	71.05	2	35.53	20.75	<0.001	0.296
	Pretest	26.37	1	26.37	15.41	<0.001	0.135
	Residuals	169.40	99	1.71			

Table 6
Adjusted mean scores and post hoc pairwise comparisons for immediate posttest total scores.

Group	Adjusted Mean (SE)	Free vs. Guided	Free vs. Restricted	Guided vs. Restricted
Free	13.05 (0.28)	–	$p < .001$	–
Guided	10.88 (0.25)	$p < .001$	–	$p < .001$
Restricted	8.54 (0.30)	–	$p < .001$	–

Note. Each category has a maximum score of 5 points, for a total score of 15.

Table 7
Adjusted mean scores and post hoc pairwise comparisons for immediate posttest content scores.

Group	Adjusted Mean (SE)	Free vs. Guided	Free vs. Restricted	Guided vs. Restricted
Free	3.60 (0.15)	$p < .001$	0.09	–
Guided	4.10 (0.13)	–	$p < .001$	–
Restricted	2.95 (0.16)	–	–	–

Note. Each category has a maximum score of 5 points, for a total score of 15.

Table 8
Adjusted mean scores and post hoc pairwise comparisons for delayed posttest total scores.

Group	Adjusted Mean (SE)	Free vs. Guided	Free vs. Restricted	Guided vs. Restricted
Free	10.55 (0.23)	–	$p < .001$	–
Guided	10.87 (0.20)	0.60	–	$p < .001$
Restricted	8.50 (0.24)	–	–	–

Table 9
Adjusted mean scores and post hoc pairwise comparisons for delayed posttest content scores.

Group	Adjusted Mean (SE)	Free vs. Guided	Free vs. Restricted	Guided vs. Restricted
Free	3.33 (0.14)	$p < .001$	0.09	–
Guided	3.90 (0.12)	–	$p < .001$	–
Restricted	2.95 (0.15)	–	–	–

groups used the same handouts (see [Appendix A](#) for a sample), but the way they interacted with them differed by instructional condition. For instance, Guided group students were required to engage with ChatGPT to explore counterexamples and then independently develop responses that supported their thesis. This method, blending AI assistance with student initiative, likely fostered

Table 10
Adjusted mean scores and post hoc pairwise comparisons for delayed posttest organization and language use scores.

Rubric	Group	Adjusted Mean (SE)	Free vs. Guided	Free vs. Restricted	Guided vs. Restricted
Organization	Free	3.70 (0.12)	–	$p < .001$	–
	Guided	3.60 (0.11)	0.76	–	–
	Restricted	2.85 (0.13)	–	–	$p < .001$
Language Use	Free	3.50 (0.11)	–	$p < .001$	–
	Guided	3.45 (0.10)	0.98	–	$p < .001$
	Restricted	2.75 (0.12)	–	–	–

deeper engagement with the content, encouraging students to develop their own ideas and arguments. In contrast, the Free group used ChatGPT autonomously with minimal instructor guidance, which may have resulted in a greater reliance on AI-generated content. The Restricted group, on the other hand, had no access to ChatGPT and engaged with the handouts solely through peer discussion and instructor-led support, without technological mediation. Their lower content scores may suggest a lack of deep engagement in activities such as rebutting, elaborating, or considering alternative perspectives, unlike their counterparts in the Guided group. This difference in approach likely accounts for the disparities in content scores observed between the groups, with the Guided group benefiting from both structured support and purposeful AI engagement.

For organization and language use, the results echoed the patterns observed in the immediate posttests. As shown in Table 10, both the Free and Guided groups significantly outperformed the Restricted group across both domains ($ps < 0.001$), and no significant differences were found between the Free and Guided groups ($ps > 0.05$). On one hand, although the Guided group benefited from instructor support in content development, such structured guidance may not be as essential for improving organization and language use. On the other hand, in the Free group, the lack of a performance gap may also reflect the limited effectiveness of students' unstructured ChatGPT use, suggesting that autonomous AI use alone may not lead to substantial gains. Given that neither group outperformed the other in these domains, organization and language use may be areas where a lighter instructional touch, offering students more freedom and less direct oversight, could be reasonably considered, particularly when learners already possess basic writing competence and motivation. This approach could also help reduce instructional demands without compromising student outcomes.

Taken together, these delayed posttest results demonstrate that while structured guidance appears especially important for the development of content, access to AI writing assistance, whether guided or unguided, can support sustained improvements in organization and language use. Importantly, the treatment effects remained significant even after controlling for pretest scores, reinforcing the role of AI tools in facilitating L2 writing development over time.

5.2. Instructor interview

The interview with the instructor provided significant insights into the integration of ChatGPT across the three distinct groups: Guided, Restricted, and Free. These groups were sequentially taught by the instructor over the semesters, from Spring 2023 to Spring 2024, offering a progressive view of how AI tools were adopted and adapted in teaching strategies throughout the academic year. At the beginning of the interview, the instructor highlighted the challenges she faced in preventing the use of ChatGPT in the Restricted group. She candidly expressed her reservations about this approach:

It was very difficult to prevent the use of ChatGPT [...] I had to constantly move around the class, catching those who used it and deducting [points from] their scores if [I] caught [them]. It's unfair that those who strictly adhered to the rules might have scored lower than those who didn't follow my instructions.

She also noted that students with lower proficiency in the Restricted group might have benefited more from using ChatGPT. Completely barring AI tools could leave these students struggling with writing, resulting in considerable frustration. However, access to ChatGPT would have at least allowed them to produce some writing, which could have been beneficial. "This doesn't necessarily mean their writing skills would have improved dramatically [if they had had access to ChatGPT], but the ability to write without pressure certainly might have aided their learning," she explained.

For the Guided group, the instructor observed that students generally found structured activities engaging. She commented, "When conducting in-class activities, most students seem to enjoy them [...] They appreciate the opportunity to exchange opinions and gain insights." This observation correlates with the group's high scores on essay tests in the study, underscoring the benefits of interactive and collaborative learning environments. However, the instructor also recognized challenges, noting that students in the Guided group struggled with the series of in-class writings that contributed to their grades:

The grading of in-class essays introduces tension. There's frequent feedback that they have to write too much, which perpetuates this tension during writing [...] Students often sit with serious expressions, approaching these assignments as if they were taking a formal exam.

She added that similar concerns were raised by the Restricted group regarding the impact of in-class essays on their final scores in the delayed posttest. These concerns were evident in her end-of-semester course evaluations, with comments such as "In-class essays are too burdensome" from the Guided group and "I wish there were more opportunities to write in English without the pressure of grades" from the Restricted group.

Drawing on the feedback, the instructor realized the need to make the writing process less stressful and more engaging for students. She then devised the Free Group approach, where traditional in-class grades were eliminated, and unrestricted use of ChatGPT was permitted. Students were encouraged to independently manage their use of the AI tool, aiming to excel in their delayed posttest essay, where ChatGPT was not allowed. This new teaching strategy, directly shaped by her initiative and the insights gained from earlier groups, was adopted for the Free group in Spring 2024.

Unlike the Guided group, no specific guidance on using ChatGPT was provided in the Free group, promoting autonomy. She stated, "I told them to figure it out on their own. They were to find and handle it themselves," emphasizing the independence given to students. When asked about the potential misuse of ChatGPT responses, or reliance solely on ChatGPT from the start, she elaborated on the students' methodical approach: "Initially, they discussed topics in groups, forming their essays through these collaborative efforts. During these discussions, they used ChatGPT to tackle difficult points."

Building on this autonomous use, the instructor further noted that students wrote their in-class essays freely, actively engaging with ChatGPT. Importantly, she observed that they used the tool strategically rather than simply copying and pasting its responses. To support this kind of engagement, the instructor incorporated specific reflective tasks into the post-writing handouts. For instance, students were asked to document both the original sentences corrected by ChatGPT and the revised versions they included in their essays. They were also prompted to analyze the types of errors identified by ChatGPT and to apply that feedback during revision. Because these in-class essays were not graded, students viewed them as low-pressure opportunities to improve their writing, which encouraged independent learning. In turn, this process played a meaningful role in preparing them for the delayed posttest (i.e., the graded end-of-semester essay). According to the instructor, it significantly motivated students to prepare more diligently for the final assessment: “Surprisingly, they displayed the most enthusiasm for in-class writing, despite it not being graded, compared to other groups whose essays were graded.”

The relaxed atmosphere in the Free group was noticeable during classes, fostering a comfortable environment for both the students and the instructor. She observed, “Students in the Free group felt more at ease participating in class, freely using ChatGPT.” Likewise, the instructor found this teaching style more comfortable, noting, “I also felt more at ease. It was far less stressful than monitoring the Restricted group for unauthorized use of ChatGPT, which was incredibly taxing.” The ease in teaching was greatly supported by adopting a hands-off approach, especially in the area of language use:

I didn’t need to micromanage every aspect of the writing process, particularly when it came to language use. With the Free group, I allowed students to manage their language with ChatGPT independently [...] Minimal intervention (on my part) seemed better. Students were encouraged to consult ChatGPT on their own [...] I believe that learning language use through imitation and repetition, where students internalize phrases for future use, is an effective learning process.

Reflecting on these varied experiences, the instructor contemplated future instructional strategies, aiming to blend the structured support of the Guided group (specifically for essay content development) with the autonomy of open access to ChatGPT (for the remaining aspects of the writing process) enjoyed by the Free group. This hybrid approach seeks to harness the benefits of structured guidance along with the creative and independent learning opportunities that proved effective in the Free group. The instructor also stressed the growing need to reassess the broader viability of attempting to constrain students’ use of AI, noting, “There are doubts about the effectiveness of restricting AI usage in the modern era.” This assertion points to the importance of aligning educational practices with the evolving tools and demands of the digital age, suggesting that conscientious integration of AI tools into L2 writing courses could be crucial for both educational outcomes and student engagement.

6. Discussion and conclusion

This study combined quantitative analysis of student essay scores and qualitative insights from an instructor interview to assess the impact of varying levels of ChatGPT access on L2 academic performance. The findings indicate that the two AI-assisted groups, Free and Guided, outperformed the Restricted group, which had no AI assistance. Notably, while the Free group, with unrestricted ChatGPT access and minimal instructor guidance, nearly matched the Guided group’s high scores overall, their lower scores in content may suggest that their use of the AI tool reduced motivation to engage in the deeper thinking necessary for idea development. Conversely, the high overall and content scores of the Guided group, with structured AI access, indicate the advantages of balanced AI integration that supports, rather than replaces, student effort. This aligns with previous findings that emphasize the pedagogical value of guided AI use in fostering communicative confidence and reducing anxiety (e.g., Kim & Su, 2024), as well as studies that highlight the complementary roles of human and AI feedback (Banihashem et al., 2024; Escalante et al., 2023).

These findings reflect the view of writing as a socially situated and tool-mediated practice, where learning is shaped by how students engage with available resources in their classroom community. The Guided group’s success demonstrates how structured support can scaffold students’ use of AI tools to foster deeper engagement and understanding. Meanwhile, the Free group’s strong overall performance highlights that greater autonomy can also be effective, especially when learners have opportunities to explore and adapt tools in personally meaningful ways. This supports the notion, widely discussed in recent L2 writing research (e.g., Koltovskaia et al., 2024), that students’ behavioral and cognitive engagement with AI is shaped by both individual agency and instructional context. Together, these results underscore that effective writing instruction depends not just on tool access, but on how learners interact with those tools within supportive and purposeful learning environments.

Pedagogically, the study provides significant insights. It revealed no substantial differences in the categories of organization and language use between the Free and Guided groups. The autonomy leveraged by the Free group suggests that not all aspects of L2 learning require stringent supervision. This level of independence allowed students to develop individualized AI usage strategies, particularly in language and organization – areas that generally necessitate less direct intervention and critical thinking. These observations point to the potential of less supervised environments to effectively nurture specific academic skills. Furthermore, based on the instructor’s insights during the interview, classes that employ selectively focused structured instruction could be more time- and effort-efficient for instructors. This may indicate that a shift toward more adaptive teaching methods could enhance the efficacy of educational programs involving AI tools.

Furthermore, the Guided group’s high scores on content emphasize the importance of structured interventions to maximize AI’s potential in education. L2 writing classes that incorporate AI should prioritize content development. For example, in classes allowing unrestricted use of AI, conducting training sessions can prepare students to engage critically with the tool. Such sessions could encourage students to broaden their ideas on essay topics, weave in personal perspectives, and creatively utilize prompts. This aligns with the WWC model’s assertion that writing is shaped not only by access to tools but by how learners interact with those tools within

meaningful communities of practice. This approach would not only enhance their ability to critically assess AI-generated content but also support them in incorporating their unique insights into their work, potentially fostering their ability to leverage both human and technological resources more effectively.

This study is not without its limitations. The high scores of the Free group on immediate post-tests, which were likely due to their open AI access, could skew the understanding of the impact of AI integration on learning outcomes. Although allowing free AI use for this group was deliberate in the study, with the aim of understanding their use of AI as a practice tool, this approach might limit broader insights. To better capture the true educational impact of AI and its effects on in-class essay writing performance, further research with different instructional designs on the impact of free uses of AI is required. In addition, the use of a group that did not have access to ChatGPT, while intended as a baseline for comparison, introduces a common limitation found in studies with no-treatment control groups. The absence of intervention in the Restricted group may have contributed to their lower performance, making it difficult to fully disentangle the specific effects of ChatGPT integration from the general benefits of using any educational tool or support mechanism.

An additional factor to consider is the informal and unmonitored nature of ChatGPT use outside of class. In the Guided group, ChatGPT use was structured and confined to teacher-led activities with specific prompts and guidance. However, no measures were implemented to monitor or restrict students' independent access to the tool beyond class hours. While the Restricted group was instructed not to use ChatGPT, it is possible that some students accessed it on their own. Similarly, the Free group was permitted to use ChatGPT freely both during and outside of class, and the extent and manner of this out-of-class usage were not tracked. This variation in out-of-class exposure may have influenced the outcomes, making it more challenging to isolate the effects of in-class instructional design. Nonetheless, this approach reflects realistic classroom conditions, where student use of AI tools is often difficult to regulate. Future studies may benefit from incorporating usage tracking mechanisms or self-reported logs to better understand the relationship between AI engagement patterns and writing performance.

Another important consideration in the study lies in the sequential grouping of participants from Spring 2023 to Spring 2024. While this design allowed us to control for instructor-related variables by having the same instructor teach all three groups, it also resulted in each group interacting with different versions of ChatGPT. Given the rapid development of the tool, later groups may have benefited from more advanced versions, potentially influencing the outcomes. This temporal variation in AI capability should be carefully considered when designing instruction that integrates AI tools. Future studies focused on institutionalizing AI in education should account for the evolving nature of AI systems in their pedagogical planning and analysis.

Furthermore, it is worth considering whether the findings of the present study, which involved students with low intermediate English proficiencies, would apply to different populations. For advanced students who are already proficient in utilizing ChatGPT for their writing process, structured guidance on AI use could potentially limit their ability to independently explore complex ideas. Conversely, the effects could be the opposite; however, this has yet to be determined. Future studies should therefore consider varying levels of student proficiency and familiarity with AI tools like ChatGPT. Tailoring guidance levels and course objectives to specific student groups, as well as determining the extent to which flexible AI use may enhance their learning experience, is crucial.

Relatedly, the instructor's experience with shaping AI use in the Restricted group underscores how instructional approaches must reflect the practical realities of the classroom. This issue highlights a tension between the rigid controls sometimes imposed on the use of technology in academic settings and the actual needs or expectations of the students. It further suggests that L2 learners might benefit from structured, yet adaptable, AI guidelines that align more closely with student learning preferences and technological trends.

Lastly, while this study highlighted the instructor's perceptions, it offers a partial view of how students themselves experienced the intervention. While existing research has begun to examine how students perceive and interact with AI tools like ChatGPT in academic settings (e.g., Albayati, 2024; Ding et al., 2023; Haq et al., 2025), few studies have investigated how instructors adapt to and implement these tools pedagogically. By addressing this gap, the current study adds to a more comprehensive understanding of instructional practice. Still, perspectives from students, particularly those in the Guided and Free groups, could have provided richer insights into how AI-supported writing environments are navigated and perceived from the learner's standpoint. Future research should include both instructor and student viewpoints to more fully capture the dynamics of AI integration in classroom settings.

Taken together, the results of this study highlight that writing development is influenced by the ways in which digital tools are integrated into classroom contexts. When thoughtfully embedded in instructional design and classroom participation, AI tools like ChatGPT can support rather than replace meaningful learning. As educational settings continue to adopt generative AI, it is essential to promote student engagement, autonomy, and development rather than diminishing the human dimensions of learning.

CRedit authorship contribution statement

Seungjin Hong: Methodology, Formal analysis, Data curation. **Yu Kyoung Shin:** Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization.

Funding

This work was supported by Hallym University, South Korea [HRF-202501-004].

Appendices A and B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.system.2025.103820>.

References

- Abbas, M., Jam, F., & Khan, T. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21, 10.
- Albayati, H. (2024). Investigating undergraduate students' perceptions and awareness of using ChatGPT as a regular assistance tool: A user acceptance perspective study. *Computers and Education: Artificial Intelligence*, 6, Article 100203.
- Banihashem, S., Kerman, N., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21, 23.
- Barnes, M., & Tour, E. (2025). Teachers' use of generative AI: a 'dirty little secret'? *Language and Education*, 1–16.
- Chew, C., Idris, N., Loh, E., Wu, W., Chua, Y., & Bimba, A. (2019). The effects of a theory-based summary writing tool on students' summary writing. *Journal of Computer Assisted Learning*, 35(3), 435–449.
- Ding, L., Li, T., & Jiang, S. (2023). Students' perceptions of using ChatGPT in a physics class as a virtual tutor. *International Journal of Educational Technology in Higher Education*, 20, 63.
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, 57.
- Ghafouri, M. (2024). ChatGPT: The catalyst for teacher-student rapport and grit development in L2 class. *System*, 120, Article 103209.
- Graham, S. (2018). A revised writer(s)-within-community model of writing. *Educational Psychologist*, 53(4), 258–279.
- Hadwin, A. F., Bakhtiar, A., & Miller, M. (2018). Challenges in online collaboration: Effects of scripting shared task perceptions. *International Journal of Computer-Supported Collaborative Learning*, 13(3), 301–329.
- Haq, M., Cao, G., & Abukhait, R. (2025). Examining students' attitudes and intentions towards using ChatGPT in higher education. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13582>
- Hong, S., & Shin, Y. (2025). Comparative effects of AI and peer collaboration on L2 writing proficiency: An empirical study. *Studies in Foreign Language Education*, 39(4), 29–50.
- Hu, G. (2005). Using peer review with Chinese ESL student writers. *Language Teaching Research*, 9(3), 321–342.
- Jin, Y., & Fan, J. (2023). Test-taker engagement in AI technology-mediated language assessment. *Language Assessment Quarterly*, 20(4–5), 488–500.
- Kang, J. (2022). Transfer of knowledge across genres and media: Investigating L2 learners' multiple composing practices. *Journal of English for Academic Purposes*, 56, Article 101096.
- Kessler, G., & Bikowski, D. (2010). Developing collaborative autonomous learning abilities in computer mediated language learning: Attention to meaning among students in wiki space. *Computer Assisted Language Learning*, 22(1), 41–58.
- Kim, Y., Belcher, D., & Peyton, C. (2023). Comparing monomodal traditional writing and digital multimodal composing in EAP classrooms: Linguistic performance and writing development. *Journal of English for Academic Purposes*, 64, Article 101247.
- Kim, A., & Su, Y. (2024). How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. *System*, 122, Article 103256.
- Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of ChatGPT for academic text revision: Behavioral, cognitive, and affective engagement. *Journal of Second Language Writing*, 65, Article 101130.
- Langer, J. A. (2002). *Effective literacy instruction: Building successful reading and writing programs*. National Council of Teachers of English.
- Lee, K. W., Said, N., & Tan, C. K. (2016). Exploring the affordances of The Writing Portal (TWP) as an online supplementary writing platform. *Computer Assisted Language Learning*, 29(6), 1116–1135.
- McCarthy, S., & McMahon, S. (1992). From convention to invention: Three approaches to peer interactions during writing. In R. Hertz-Lazarowitz, & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 17–35). Cambridge University Press.
- McCarthy, S., & Mkhize, D. (2013). Teachers' orientations toward writing. *Journal of Writing Research*, 5(1), 1–33.
- Molenaar, I., Roda, C., Boxtel, C. V., & Sleepers, P. (2012). Dynamic scaffolding of socially regulated learning in a computer-based learning environment. *Computers & Education*, 59(2), 515–523.
- Mortensen, J. (2018). Language regulation in collaborative student writing: A case study. *Language and Education*, 32(6), 529–547.
- Newell, G., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly*, 46(3), 273–304.
- Newell, G., Bloome, D., & Hirvela, A. (2015). *Teaching and learning argumentative writing in high school English language arts classrooms*. Routledge.
- O'Sullivan, B. (2023). Reflections on the application and validation of technology in language testing. *Language Assessment Quarterly*, 20(4–5), 501–511.
- Pack, A., & Maloney, J. (2023). Using generative artificial intelligence for language education research: Insights from using OpenAI's ChatGPT. *Tesol Quarterly*, 57(4), 1571–1582.
- Pérez-Llantada, C. (2024). Approaching digital genre composing through reflective pedagogical praxis. *Journal of English for Academic Purposes*, 68, Article 101349.
- Prapahn, P., & Prapahn, K. (2023). AI technologies in the ESL/EFL writing classroom: The villain or the champion? *Journal of Second Language Writing*, 62, Article 101072.
- Prior, P. (2006). A sociocultural theory of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 54–66). The Guilford Press.
- Pu, S., & Xu, H. (2024). Self-assessment criteria as projected realities: A social cognitive study on student experience in postgraduate thesis writing. *Language and Education*, 38(5), 766–780.
- Ruiz-Funes, M. (2001). Task representation in foreign language reading-to-write. *Foreign Language Annals*, 34, 226–234.
- Shulgina, G., Costley, J., & Shcheglova, I. (2024). Online peer editing: The influence of comments, tracked changes and perception of participation on students' writing performance. *Smart Learning Environments*, 11, 30. <https://doi.org/10.1186/s40561-024-00315-8>.
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. (2024). Comparing the quality of human and ChatGPT feedback on students' writing. *Learning and Instruction*, 91, Article 101894.
- Teng, M. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, Article 100270.
- Voss, E., Cushing, S. T., Ockey, G. J., & Yan, X. (2023). The use of assistive technologies including generative AI by test takers in language assessment: A debate of theory and practice. *Language Assessment Quarterly*, 20(4–5), 520–532.
- Wambsganss, E., Janson, A., & Leimeister, J. (2022). Enhancing argumentative writing with automated feedback and social comparison nudging. *Computers & Education*, 191, Article 104644.
- Wang, X. (2017). Spatial and literacy practices of Chinese international students across a bridge writing classroom and WeChat. *Language and Education*, 31(6), 561–579.
- Wang, X., Liu, Q., Pang, H., Tan, S., Lei, J., Wallace, M., & Li, L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education*, 194, Article 104703.

- Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing*, 62, Article 101071.
- Won, D., Shin, Y., Kim, H., & Yoo, I. (2025). Advancing language assessment with GPT: Is it nonnative-language friendly? *Language Assessment Quarterly*, 22(1), 1–28.
- Zaccaron, R., & Khafaj, D. (2020). Knowing me, knowing you: A comparative study on the effects of anonymous and conference peer feedback on the writing of learners of English as an additional language. *System*, 95, Article 102367.

Seungjin Hong is Senior Researcher in Institute for Language and Information at Sogang University. Her work focuses on scalar adjectives, quantifiers, factive and non-factive verbs in experimental semantics and pragmatics perspectives. She has published articles in *Discourse and Cognition*, *Language and Information*, *Korean Journal of Linguistics*, *English Teaching*, and *Modern English Education*.

Yu Kyoung Shin (corresponding author) is Assistant Professor in the School of Global Studies at Hallym University. Her research interests include corpus analysis of academic registers and disciplinary variation, as well as AI chatbot-based instruction, particularly for applications to L2 learning and teaching. Her recent work has appeared in *Journal of English for Academic Purposes*, *Journal of Second Language Writing*, *TESOL Quarterly*, *System*, *English for Specific Purposes*, *Innovation in Language Learning and Teaching*, *Language and Intercultural communication*, *Corpora*, *Applied Linguistics Review*, and *Language Teaching Research*.