Integrating Generative AI into Strategy-Based Instruction: Adaptive Scaffolding for Self-Regulated Learning in Secondary EFL Writing

Chuang-Qi Chen
Institute of Language Sciences, Shanghai International Studies University
413806370@qq.com
Meng-Rui Li
School of Education, Shanghai International Studies University
LMR19115063061@outlook.com
Hui Jin
School of Education, Shanghai International Studies University
nmjinhui@163.com

Abstract: *Can GenAI do more than generate text—can it teach students to regulate their own writing? This study addresses that question through a randomized controlled trial (N = 80) examining whether a GenAI-powered adaptive dialogue agent, embedded in a Strategy-Based Instruction (SBI) framework, develops secondary EFL students' Self-Regulated Learning (SRL) strategies alongside writing performance. The agent was designed to foreground student agency: it deployed four context-sensitive modes—genre knowledge retrieval, ideation prompting, reflective self-assessment, and motivational support—mapped onto the SRL cycle. Over eight weeks, the experimental group (n = 40) significantly outperformed controls on writing scores, with the performance gap widening at a one-month delayed post-test. Crucially, the intervention produced sustained gains in text generation and self-monitoring strategies—competencies rarely improved through conventional instruction. These findings reframe GenAI's pedagogical role: not as a writing surrogate, but as a metacognitive scaffold that makes strategic thinking visible and keeps learners in the driver's seat of their own writing development.*

Keywords: Generative AI (GenAI), Self-Regulated Learning (SRL), Adaptive Scaffolding, EFL Writing, Strategy-Based Instruction (SBI)

1. Introduction

Generative Artificial Intelligence (GenAI) is fundamentally reshaping text production(Cerchione et al., 2026). While AI-driven efficiency offers transformative potential, it triggers a critical pedagogical concern: the risk of bypassing students' cognitive engagement and Self-Regulated Learning (SRL) capacities. If learners are relegated to passive recipients of AI-generated content, their essential writing skills and cognitive agency face a "hollowing out" effect in the AI era(Kim et al., 2025)(Adams, 2025).

To address this challenge, writing instruction must pivot from product evaluation to cultivating the underlying cognitive processes. Strategy-Based Instruction (SBI), grounded in SRL theory, provides a vital framework for this shift (Rubin et al., 2007). In an AI-augmented landscape, the core competency is no longer mere text production but the strategic management of the writing lifecycle—from goal setting and planning to monitoring and revising. However, implementing SBI in traditional large-scale secondary classrooms is logistically difficult(Del Mario & Tran, 2024); teachers often lack the bandwidth to monitor every student's "black box" of cognitive processes, resulting in delayed, product-oriented feedback.

Recent advancements in GenAI offer a unique opportunity to bridge this gap. Unlike static tools, GenAI can function as a "metacognitive partner" capable of delivering ubiquitous, adaptive scaffolding—such as genre knowledge retrieval or ideation guidance—tailored to individual needs(Tang & Putra, 2025). Despite extensive research on AI for error correction, few studies have explored the systematic integration of GenAI within an SBI framework to specifically foster secondary students' SRL strategies.

Addressing this gap, the present study employs a randomized controlled trial to evaluate the extent to which a GenAI-supported SBI intervention enhances writing performance and strategic deployment.

2. GenAI Agent Design and Implementation

Writing strategy instruction within an SBI framework demands support that is both responsive and differentiated. Students must simultaneously navigate genre conventions, construct logical arguments, monitor their own progress, and manage writing-related affect(Rubin et al., 2007)—each task calling for a qualitatively different kind of support. A one-size-fits-all chatbot cannot meet these varied demands: it risks delivering off-target feedback that impedes rather than builds strategic competence (Azevedo & Hadwin, 2005). This concern is especially acute for genre knowledge delivery, where factual accuracy is non-negotiable: even minor hallucinations can erode student trust and distort their understanding of writing conventions (Hattie & Timperley, 2007). Given inherent LLM limitations—output variability and the potential for confabulation—a structurally differentiated, modular design is essential to guarantee both pedagogical reliability and interaction-level adaptability.

To address these constraints, we constructed the agent around four task-specific functional modes, each instantiating a core scaffolding principle (Puntambekar & Hubscher, 2005) and activated in response to real-time learner cues:

1.Genre Knowledge Mode: Anchors instruction in verified content via Retrieval-Augmented Generation (RAG). Rather than relying on open-ended generation, the agent consults a curriculum-aligned knowledge base to supply structurally accurate guidance on text types such as formal letters and argumentative essays. This retrieval-grounded approach minimizes the risk of misleading information and ensures fidelity to established writing conventions.

2.Ideation & Logic Mode: Supports pre-writing and drafting through strategic questioning rather than text generation. The agent poses probing prompts that push students to articulate their claims, identify supporting evidence, and tighten the logical structure of their arguments. By keeping students as the primary authors, this mode cultivates independent reasoning within the zone of proximal development.

3.Reflective & Monitoring Mode: Develops metacognitive self-regulation during revision. Rather than issuing evaluative judgments, the agent poses goal-referencing questions (e.g., “How well does this paragraph signal logical contrast through transitional language?”) that prompt learners to compare their draft against their original intentions. This structured reflection trains students to apply self-assessment criteria independently over time.

4.Motivational & Affective Mode: Attends to the emotional dimension of writing, which is particularly salient among adolescent EFL learners prone to writing anxiety and low self-efficacy. The agent offers affirmation and empathetic acknowledgment to sustain motivation and preserve a psychologically safe space—conditions that research identifies as foundational for productive intellectual risk-taking.

Taken together, these four modes operationalize the principle of contingent scaffolding—providing targeted support calibrated to the learner’s immediate cognitive and affective state rather than following a fixed instructional sequence (Azevedo & Hadwin, 2005).

2.1. Workflow Architecture

Figure 1 illustrates the technical infrastructure underpinning the agent. The system was built on Dify (v 0.13.2), an open-source LLM-ops platform selected for its support of multi-agent orchestration and RAG integration within a visual workflow environment.

Each incoming message first passes through a Context Manager and Memory Module, which reconstruct the learner’s interaction history and surface relevant metadata. A Routing mechanism then classifies the communicative intent of the message and directs it to the most contextually appropriate of the four specialized Retrieval Agents.

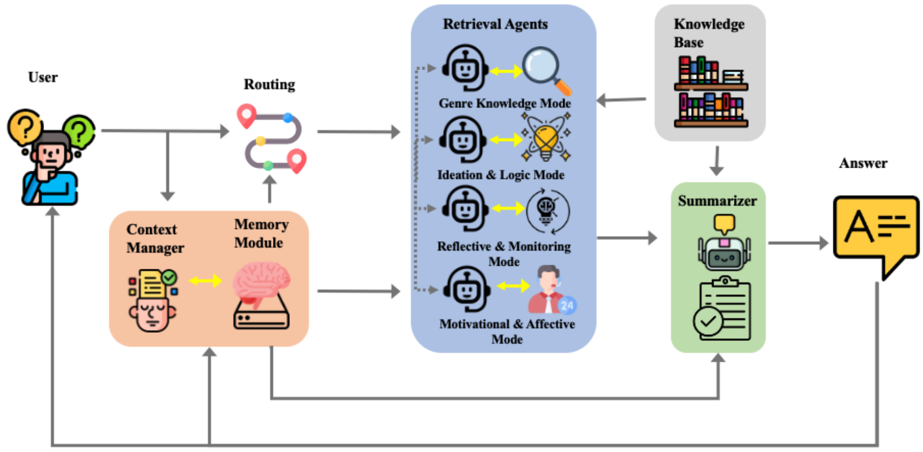

Figure 1 *GenAI Agent Workflow Architecture*

Agent outputs converge in a Summarizer component that consolidates multi-source responses, enforces content safety filters, and formats the final reply for fluent delivery. A pilot with 12 students (240 interactions) confirmed system viability: response accuracy reached 94%, all outputs passed safety review (100% compliance), and response latency remained under 5 seconds—sufficient to sustain natural turn-taking in dialogue.

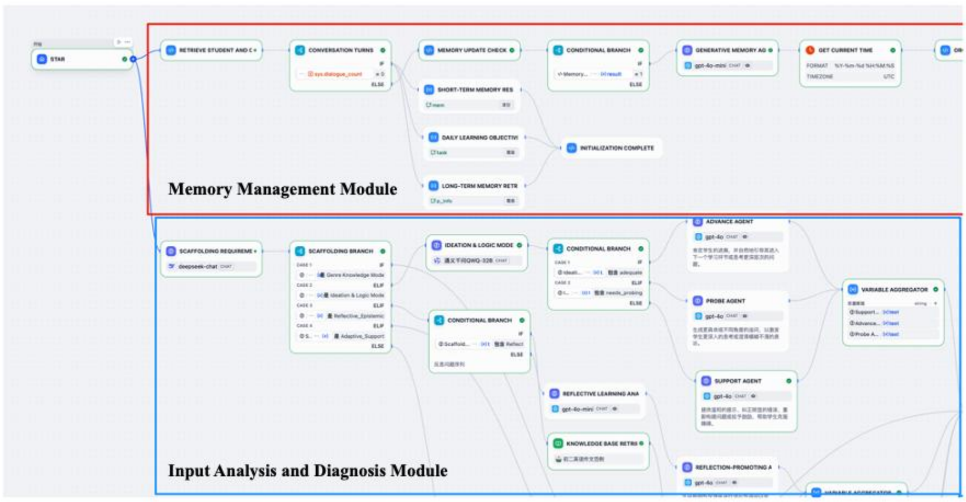

Figure 2 *Screenshot of the Dify Workflow Interface (Partial View)*

3. Method

3.1. Participants

Eighty eighth-grade students from a single middle school participated in a university-based summer camp program, following parental consent. Eighth graders were selected because this developmental stage is critical for writing proficiency gains in EFL contexts(Bai et al., 2022). Participants were randomly assigned to the experimental group (N = 40; 22 boys, 18 girls) or the control group (N = 40; 23 boys, 17 girls).

3.2. Procedure

The procedure spanned 8 weeks followed by a delayed assessment. Preparation (Week 1): A 40-minute session covering initial training and pre-tests on writing, SRL strategies. Intervention (Weeks 2–7): Over 6 weeks, students completed four recursive writing topics. In 80-minute sessions, both groups followed the SRL cycle (Forethought, Performance, Reflection). The control group relied on conventional teacher-led instruction and static checklists, while the experimental group utilized GenAI scaffolding for personalized goal setting, drafting support, and retrospective analysis. Assessment (Week 8 & Delayed): Immediate post-tests were conducted in Week 8, followed by a delayed post-test one month later to evaluate long-term retention.

3.3. Measures

Students' writing performance was assessed based on their composition scores. An analytical scoring rubric developed by(Hyland, 2003), widely recognized for its comprehensive description of writing proficiency, was adopted for assessment. The rubric covers three dimensions: sentence structure and vocabulary (40 pts), organization and coherence (20 pts), and format and content (40 pts). Two trained senior EFL teachers scored each essay independently; final scores were averaged to ensure inter-rater reliability.

Participants' deployment of SRL strategies in ESL/EFL writing was measured using a writing strategy questionnaire adapted from(Bai et al., 2022). The instrument comprises 17 items divided into four subscales: planning strategies (4 items, alpha = 0.87), text generation strategies (4 items, alpha = 0.71), revising strategies (4 items, alpha = 0.83), and self-monitoring strategies (6 items, alpha = 0.82). Responses were recorded on a 5-point Likert scale, where higher scores indicate a higher frequency of SRL strategy usage.

3.4. Data Analysis

Analyses were conducted in Jamovi 2.4. Normality (Shapiro-Wilk, Q-Q plots) and variance homogeneity (Levene's test) were confirmed, and baseline equivalence verified via independent t-tests. A two-way repeated measures ANOVA (Time × Group) was the primary analysis ($\alpha = .05$).

5. Results

Normality tests confirmed that the scores for writing performance and the SRL strategy use scale were normally distributed ($p > 0.05$). Table 1 presents the descriptive statistics for all variables and the results of baseline (T0) comparisons between the two groups. Mauchly's test indicated that the assumption of sphericity was met across all dimensions. Given the significant interaction effect observed between Group and Time, simple effects analyses were subsequently conducted.

Table 1. *Descriptive statistics and time effects analysis.*

	Group	Pre-test	Post-test	Delayed test
Composition Score	Exp Group	64.9 (6.11)	69.3 (5.83)	77.9 (5.28)
	Control Group	64.2 (5.40)	66.2 (5.82)	72.9 (4.51)
Planning Strategy	Exp Group	2.88 (0.68)	3.46 (0.84)	3.26 (0.72)
	Control Group	2.92 (0.85)	3.29 (0.89)	2.94 (0.78)
Text Generation Strategy	Exp Group	2.99 (0.40)	3.62 (0.31)	3.69 (0.38)
	Control Group	3.00 (0.38)	3.29 (0.39)	3.24 (0.34)
Revision Strategy	Exp Group	2.96 (0.38)	3.45 (0.36)	3.28 (0.32)
	Control Group	3.08 (0.37)	3.27 (0.36)	3.16 (0.33)
Self-Monitoring Strategy	Exp Group	3.00 (0.71)	3.51 (0.73)	3.25 (0.62)
	Control Group	3.09 (0.60)	3.17 (0.76)	3.16 (0.66)

5.1. Results on Writing Performance

Regarding writing performance, the mean scores for the experimental and control groups across the three time points (T0, T1, T2) were 64.9 ± 6.11, 69.3 ± 5.83, 77.9 ± 5.28 and 64.2 ± 5.40, 66.2 ± 5.82, 72.9 ± 4.51, respectively. Simple effects analysis of Time indicated significant improvements within both groups ($p < 0.001$). However, pairwise comparisons (see Table 3) revealed that the control group showed no significant growth from pre-test (T0) to post-test (T1) ($p > 0.05$), whereas all other time intervals for both groups showed significant gains ($p < 0.01$).Simple effects analysis of Group demonstrated that the experimental group significantly outperformed the control group at T1 ($t = 2.50$, $p = 0.013$) and T2 ($t = 4.00$, $p < 0.001$), with the mean difference widening from 3.11 to 4.97.

Table 2. *Two-factor repeated measures ANOVA results.*

	Effect	SS	df	MS	F	p	$\eta 2p$
Composition Score	Time	4961	2	2480.5	92.98	< .001	0.414
	Group	510	1	509.8	13.5	< .001	0.068
	Group x Time	179	2	89.6	3.36	0.037	0.025
Planning Strategy	Time	8.90	2	4.45	23.81	< .001	0.057
	Group	1.31	1	1.31	0.86	0.358	0.009
	Group x Time	1.34	2	0.67	3.59	0.030	0.009
Text Generation Strategy	Time	11.21	2	5.61	75.00	< .001	0.263
	Group	3.84	1	3.84	14.9	< .001	0.109
	Group x Time	2.15	2	1.07	14.4	< .001	0.064
Revision Strategy	Time	4.63	2	2.31	59.6	< .001	0.137
	Group	0.18	1	0.18	0.61	0.437	0.006
	Group x Time	1.02	2	0.51	13.2	< .001	0.034
Self-Monitoring Strategy	Time	3.39	2	1.69	10.13	< .001	0.031
	Group	0.77	1	0.77	0.72	0.40	0.007
	Group x Time	1.84	2	0.92	5.51	0.005	0.017

5.2. Results on SRL Strategy Use

Analysis of the four SRL sub-dimensions (planning, text generation, revising, and self-monitoring) revealed distinct patterns of improvement. Longitudinally, both groups achieved significant gains in planning, text generation, and revising strategies across the time points ($p < 0.001$). However, a notable divergence occurred in self-monitoring strategies, where only the experimental group demonstrated significant growth ($F = 15.02$, $p < 0.001$), while the control group showed no significant change ($F = 0.43$, $p = 0.652$).

Regarding between-group comparisons, the experimental group demonstrated superior immediate outcomes at the post-test (T1), significantly outperforming the control group in text generation ($t = 3.95$, $p < 0.001$), revising ($t = 2.25$, $p = 0.026$), and self-monitoring ($t = 2.22$, $p = 0.028$), though no significant difference was found in planning ($p > 0.05$). By the delayed post-test (T2), the experimental group's advantage in text generation widened further ($t = 5.33$, $p < 0.001$). However, the performance gaps in revising and self-monitoring narrowed and were no longer statistically significant ($p > 0.05$), aligning with planning strategies which remained non-significant throughout. These findings indicate that the intervention was particularly effective in fostering sustained improvements in text generation strategies.

Table 3. *Two-factor repeated measures ANOVA results.*

Comparison	Time	Group	Difference	SE	t	p
Composition Score	T0-T1	Exp	4.38	1.17	3.75	0.004
		Control	2.00	1.15	1.73	1.000
	T1-T2	Exp	8.64	1.17	7.39	< .001
		Control	6.77	1.15	5.87	< .001
	T0-T2	EXP	13.03	1.17	11.14	< .001
		Control	8.77	1.15	5.87	< .001
Planning Strategy	T0-T1	EXP	0.58	0.10	5.96	< .001
		Control	0.36	0.10	3.75	0.004

	T1-T2	EXP	-0.20	0.10	-2.03	0.661
		Control	-0.34	0.10	-3.56	0.007
	T0-T2	EXP	0.38	0.10	3.93	0.002
		Control	0.02	0.10	0.19	1.000
Text Generation Strategy	T0-T1	EXP	0.62	0.06	10.04	< .001
		Control	0.29	0.06	4.70	< .001
	T1-T2	EXP	0.07	0.06	1.14	1.000
		Control	-0.04	0.06	-0.71	1.000
	T0-T2	EXP	0.69	0.06	11.18	< .001
		Control	0.24	0.06	3.99	0.002
Revision Strategy	T0-T1	EXP	0.49	0.04	11.06	< .001
		Control	0.19	0.04	4.25	< .001
	T1-T2	EXP	-0.17	0.04	-3.88	0.002
		Control	-0.11	0.04	-2.41	0.257
	T0-T2	EXP	0.32	0.04	7.18	< .001
		Control	0.08	0.04	1.84	1.000
Self-Monitoring Strategy	T0-T1	EXP	0.51	0.09	5.48	< .001
		Control	0.08	0.09	0.85	1.000
	T1-T2	EXP	-0.26	0.09	-2.80	0.087
		Control	-0.01	0.09	-0.10	1.000
	T0-T2	EXP	0.25	0.09	2.68	0.121
		Control	0.07	0.09	0.75	1.000

6. Discussion

The present study demonstrates that integrating GenAI into a Strategy-Based Instruction (SBI) framework can significantly enhance secondary students' writing performance by fostering the acquisition and enactment of writing strategies, rather than merely improving surface-level textual features. Quantitative results showed that while both groups improved over time, the GenAI-supported group not only outperformed the control group at the immediate post-test but also exhibited a widening performance gap at the delayed post-test. This sustained advantage suggests that the observed gains were not transient but reflected deeper changes in students' writing processes.

Crucially, the strongest and most durable effects were observed in students' use of text generation strategies, aligning closely with the Ideation & Logic Mode. This mode addressed a persistent EFL bottleneck—difficulty generating ideas and building coherent argumentation (Pessoa et al., 2017)—by using dialogic questioning to scaffold reasoning rather than supplying text directly. The control group's non-significant T0–T1 gains reflect the inadequacy of delayed, product-focused feedback for this cognitively demanding phase (Morris et al., 2021). Students' own characterization of the agent as a "Socratic tutor" or "co-pilot" signals a shift from passive text reception to active meaning construction—an immediacy and personalization unattainable in large teacher-led classrooms (Tran & Ma, 2025). RAG-grounded genre support further ensured accuracy and contextual alignment, mitigating risks of unreliable guidance (Li et al., 2025).

Self-monitoring emerged as the most distinctive SRL divergence: the experimental group improved significantly ($F = 15.02$, $p < 0.001$) while the control group showed no meaningful change—a contrast expected given that self-monitoring is among the hardest SRL skills to cultivate conventionally (Azevedo, 2009).

The Reflective & Monitoring Mode drove these self-monitoring gains by requiring students to evaluate drafts against self-set goals rather than receive summative corrections—a form of feed-forward planning rarely sustained in time-constrained classrooms. The absence of between-group differences in planning strategies further suggests that GenAI's value lies in sustaining in-process regulation rather than initiating pre-writing (Su et al., 2023). Notably, students who questioned AI suggestions and cross-checked them against textbooks exhibited epistemic vigilance rather than passivity—evidence that well-designed GenAI can reinforce, rather than erode, learner agency.

7. Conclusion

This randomized controlled trial provides robust evidence that a GenAI-powered dialogue agent embedded within an SBI framework can effectively enhance secondary EFL students' writing performance and self-regulated learning. Students in the experimental group demonstrated significantly higher writing performance than the control group, with the performance gap widening at the delayed post-test. Beyond writing outcomes, the intervention yielded substantive improvements in students' writing processes. The experimental group showed marked and sustained gains in text generation strategies and self-monitoring strategies—two SRL components often difficult to cultivate in conventional classrooms. These results indicate that the intervention did more than improve final texts; it successfully supported students in regulating idea development and monitoring their progress. Ultimately, well-designed GenAI scaffolding can help secondary students transition from dependent users of AI-generated text to active, self-regulated writers.

Despite these encouraging findings, several limitations should be acknowledged. First, the study involved a relatively small sample over an eight-week period. Although significant effects were observed, this duration may be insufficient for writing strategies to become fully automatized. Future longitudinal studies with larger, more diverse samples are needed to examine the long-term transfer and stability of SRL strategy use across different writing tasks. Second, the effective use of the dialogue agent appeared contingent on students' ability to interact with the system. Challenges related to prompt literacy suggest that future GenAI-supported instruction should incorporate explicit training on how to communicate needs effectively to AI systems to maximize the scaffolding benefits. Finally, the present study relied primarily on self-report questionnaires to capture SRL strategy use. While informative, these measures provide only indirect evidence of cognitive processes. Future research should integrate objective behavioral metrics—such as system log data, keystroke analysis, or eye-tracking—to obtain fine-grained insights into learners' real-time regulatory behaviors and cognitive load during AI-supported writing.

References

Adams, R. (2025, October 15). Pupils fear AI is eroding their ability to study, research finds. The Guardian. https://www.theguardian.com/technology/2025/oct/15/pupils-fear-ai-eroding-study-ability-research

Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. Metacognition and Learning, 4(1), 87–95. https://doi.org/10.1007/s11409-009-9035-7

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition – Implications for the design of computer-based scaffolds. Instructional Science, 33(5/6), 367–379.

Bai, B., Wang, J., & Zhou, H. (2022). An intervention study to improve primary school students' self-regulated strategy use in English writing through e-learning in Hong Kong. Computer Assisted Language Learning, 35(9), 2265–2290. https://doi.org/10.1080/09588221.2020.1871030

Cerchione, R., Liccardo, G., & Passaro, R. (2026). Artificial knowledge generation: Investigating the revolutionary role of generative AI in knowledge management. Journal of Innovation & Knowledge, 11, 100866. https://doi.org/10.1016/j.jik.2025.100866

Del Mario, J., & Tran, H. (2024). A literature review: Analyzing barriers hindering the implementation of self-regulated learning in science classrooms. Frontiers in Education, 9. https://doi.org/10.3389/feduc.2024.1330449

Hattie, J., & Timperley, H. (2007). The Power of Feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Hyland, K. (2003). Genre-based pedagogies: A social response to process. Journal of Second Language Writing, 12(1), 17–29. https://doi.org/10.1016/S1060-3743(02)00124-8

Kim, S., So, H.-J., & Park, K. (2025). Supporting learner agency in collaborative writing with generative AI. British Journal of Educational Technology, n/a(n/a). https://doi.org/10.1111/bjet.70015

Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., & Wang, F. L. (2025). Retrieval-augmented generation for educational application: A systematic survey. Computers and Education: Artificial Intelligence, 8, 100417. https://doi.org/10.1016/j.caeai.2025.100417

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. Review of Education, 9(3), e3292. https://doi.org/10.1002/rev3.3292

Pessoa, S., Mitchell, T. D., & Miller, R. T. (2017). Emergent arguments: A functional approach to analyzing student challenges with the argument genre. Journal of Second Language Writing, 38, 42–55. https://doi.org/10.1016/j.jslw.2017.10.013

Puntambekar, S., & Hubscher, R. (2005). Tools for Scaffolding Students in a Complex Learning Environment: What Have We Gained and What Have We Missed? Educational Psychologist, 40(1), 1–12. https://doi.org/10.1207/s15326985ep4001_1

Rubin, J., Chamot, A. U., Harris, V., & Anderson, N. J. (2007). Intervening in the use of strategies. Language learner strategies, 30, 29-45.

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. Assessing Writing, 57, 100752. https://doi.org/10.1016/j.asw.2023.100752

Tang, K.-S., & Putra, G. B. S. (2025). Generative AI as a Dialogic Partner: Enhancing Multiple Perspectives, Reasoning, and Argumentation in Science Education with Customized Chatbots. Journal of Science Education and Technology. https://doi.org/10.1007/s10956-025-10240-1

Tran, T. T. T., & Ma, Q. (2025). Technology-enhanced self-regulation training: A dynamic training model to facilitate second language Vietnamese learners' self-regulated writing skills. System, 130, 103625. https://doi.org/10.1016/j.system.2025.103625