

Guided GenAI Use in EFL Poetry Writing: Multimodal Literacy and Learner Agency

Rakhun Kim[†]

Hongik University

ABSTRACT

Prior research has raised concerns that generative AI (GenAI) may encourage learner passivity, although most studies have examined its unmediated use. To fill this gap, this exploratory comparative mixed-methods study examined Korean high school students' engagement with GenAI in guided and unguided poetry writing. Two intact classes received the same instruction in poetic form; only the guided class received multimodal-literacy and learner-agency instruction. A task-specific 12-item self-report questionnaire was administered at pretest, immediate posttest, and four-week delayed posttest. Guided learners' revisions were analyzed using a perceive-select-act-reflect framework. Mixed ANOVAs showed time-by-condition interactions: the guided class reported a larger increase in agency than the comparison class and remained above pretest at delayed posttest. Qualitatively, all guided learners modified AI-generated texts and explained revision decisions, although rationales varied in specificity, technical precision, and correspondence with revisions. Findings document contrasting self-reported trajectories across classes and selective GenAI engagement in the guided class, without establishing causal instructional effects.

Keywords: GenAI, learner agency, multimodal literacy, EFL writing, poetry writing

1. Introduction

Generative artificial intelligence (GenAI) has rapidly moved to the center of language education. Tools such as ChatGPT offer second-language writers immediate, individualized feedback, with documented gains in writing quality and motivation (Song & Song, 2023). In Korea, this development coincides with the implementation of the 2022 Revised National Curriculum, which emphasizes multimodal competence, self-directed learning, and the critical evaluation of digital

* Acknowledgment: ChatGPT (GPT 5.5, OpenAI) was used for language editing, as well as for the design and editing of Figure 1. All substantive content, analyses, interpretations, and conclusions were reviewed and verified by the author, who takes full responsibility for the final manuscript.

[†] Corresponding author: kimrhee02@hongik.ac.kr



information (Ministry of Education, 2022). The pedagogical challenge, therefore, is not simply to integrate GenAI into language learning but to ensure that learners critically evaluate its outputs and exercise agency in deciding how to use them.

However, a growing body of research has raised concerns that GenAI may encourage learners' passivity and overreliance. Learners who used large language models (LLM) during writing tasks have been reported to show lower levels of cognitive engagement than those writing under other conditions, a pattern described by Kosmyrna et al. (2025) as "cognitive debt." Similarly, students who relied heavily on GenAI support performed less successfully when the tool was no longer available (Bastani et al., 2025). Fan et al. (2025) further found that users may reduce their own monitoring and evaluation by deferring judgment to AI-generated responses. Thus, alongside documented benefits in writing and motivation (Song & Song, 2023), existing research points to risks associated with uncritical reliance on GenAI (Bastani et al., 2025; Kosmyrna et al., 2025). Yet these studies have focused primarily on outcomes of AI use rather than on how learners evaluate and act on AI-generated alternatives during composing. Emerging work has linked GenAI interaction design and learners' GenAI literacy to subsequent independent writing performance (Jin et al., 2026). What remains less understood is what learners actually do at specific revision points: how they evaluate AI-generated changes, decide what to retain, modify, restore, or reject, and justify those decisions.

The design of GenAI-supported writing tasks may be particularly important in creating opportunities for such decision making. From an ecological perspective, learner agency is not a fixed trait but emerges through learners' engagement with the opportunities and constraints of a learning environment (Mercer, 2011; van Lier, 2000). Revision provides a particularly useful site for examining this process because each AI-generated alternative creates a concrete decision point: learners can accept it, modify it, restore their original wording, or reject it. These textual decisions, together with learners' stated reasons for them, make otherwise abstract processes of evaluation and choice observable (Jiang & Lai, 2025). A focus on revision therefore shifts the question from whether GenAI improves a final text to how learners exercise judgment over the alternatives it provides.

English poetry writing offers a particularly revealing context for this analysis because poetic meaning is constructed through the interaction of linguistic, aural, and visual resources, including lexical and syntactic choices, rhyme, rhythm, and rhyme schemes, as well as lineation, stanzaic organization, typography, and other aspects of the poem's visual presentation (Kress, 2010; New London Group, 1996).

Revision in poetry consequently requires judgments across multiple, sometimes competing, semiotic criteria. A change that improves grammatical naturalness may disrupt rhythm; a stronger rhyme may alter intended meaning or voice; and a more elaborate expression may weaken line economy or stanzaic balance. Poetry therefore creates recurrent trade-offs among competing linguistic, aural, and visuospatial resources (Hanauer, 2012; Newfield & d'Abdon, 2015). Such trade-offs make learners' evaluative criteria especially visible and provide a concrete context in which multimodal literacy and learner agency can be examined together.

Accordingly, this exploratory comparative mixed-methods study examines patterns of GenAI engagement under guided and comparison conditions in Korean high school students' English poetry writing. It addresses two related gaps: whether explicit evaluative guidance is associated with different task-specific trajectories in self-reported learner agency and multimodal literacy, and how guided learners evaluate and intervene in GPT-revised texts. To address these questions, the study combines questionnaire data collected across three time points with process-level analysis of guided learners' initial drafts, GPT-revised texts, final revisions, and stated rationales. This design enables examination of both class-level self-reported trajectories and the specific evaluative decisions through which selective engagement with GenAI was enacted. Two research questions guide the study:

- RQ1. How do Korean high school students' self-reported multimodal literacy and learner agency change across the pretest, immediate posttest, and delayed posttest under guided and unguided GenAI-supported poetry-writing conditions?
- RQ2. Following instruction, how do guided learners participate in GenAI-revised versions of their poems, and how do their responses reflect multimodal literacy and learner agency?

2. Literature Review

2.1. GenAI in L2 writing

The growing integration of GenAI into language education has produced a rapidly expanding but mixed body of research. Studies suggest that tools such as ChatGPT can support L2 writing quality and learner motivation through immediate and

individualized feedback (Song & Song, 2023). Li et al.'s (2025) systematic review of 144 empirical studies confirms that writing and feedback have become major areas of GenAI research in language education, although the evidence remains heavily concentrated in higher education. At the same time, concerns have emerged regarding learner passivity and overreliance. LLM-assisted writers have shown reduced cognitive engagement relative to writers working under other conditions (Kosmyna et al., 2025), and students who relied heavily on GenAI during learning performed less successfully when subsequently required to work independently (Bastani et al., 2025). In L2 writing, learners may also defer monitoring and evaluation to AI-generated responses rather than critically scrutinizing them (Fan et al., 2025). Korean EFL research reflects this same combination of benefits and limitations. Lee's (2025) meta-analysis of 12 studies found an overall medium-to-large effect of ChatGPT-supported instruction on English writing but also highlighted the need for stronger research designs and greater representation of secondary-school contexts. Individual studies likewise report positive perceptions of ChatGPT for revision alongside concerns about feedback quality, authorship, and learning effectiveness (Bok & Cho, 2023), as well as perceived cognitive and affective benefits from its use in writing revision (Park, 2024). In narrative writing, Seo (2024) found gains in overall performance and fluency but nonuniform development in syntactic complexity, further suggesting that the benefits of GenAI-assisted writing vary across dimensions.

More recent research has begun to shift attention from whether GenAI improves final products to how learners engage with and make decisions about AI-generated suggestions during composing and revision. Koltovskaia et al. (2024) showed that graduate ESL writers' behavioral uptake of ChatGPT suggestions did not necessarily coincide with their cognitive evaluation of them. In a Korean EFL context, Hwang et al. (2025) found that learners frequently relied on general prompts and focused on surface-level concerns, with limited improvement in higher-order dimensions such as content, organization, and cohesion. Michel et al. (2025) likewise demonstrated that comparing learner-produced texts with GenAI-generated alternatives created opportunities for revision and deliberation in collaborative L2 writing. The scope of GenAI-supported writing is also expanding beyond conventional text revision, creating broader literacy demands for L2 writers; multimodal GenAI, for example, may support multimodal and prompt literacy through engagement with multiple modes of meaning-making (Kang & Yi, 2023). Yet process-oriented research remains less developed in creative writing, particularly poetry. Seo (2024) extended GenAI-supported L2 writing to narrative composition, while school-based poetry

studies have demonstrated the feasibility of ChatGPT-supported creative writing: Yang (2024) reported positive responses among 165 middle-school students writing poetry in Korean, and Kultas (2026) found improvements in poetry-writing performance and attitudes following ChatGPT-generated formative feedback and revision in primary education. Together, these studies establish the relevance of GenAI to creative writing but provide limited evidence about the specific decisions through which learners evaluate and selectively retain, modify, restore, or reject AI-generated language. Relatively little is therefore known about how pedagogical guidance relates to such evaluative processes in secondary-school EFL creative writing. The present study addresses this gap through Korean high-school English poetry writing, combining comparative self-report trajectories with a process-level analysis of guided learners' drafts, GPT-revised texts, final revisions, and stated rationales to examine learners' evaluative decisions during GenAI-supported revision.

2.2. Learner agency and GenAI

Within sociocultural and ecological perspectives, learner agency is understood not as a stable individual trait but as context-dependent engagement with environmental affordances (Mercer, 2011; van Lier, 2000). Building on this view, the present study operationalizes agency as a recursive perceive-select-act-reflect cycle: learners perceive available affordances, select among alternatives, act on their choices, and reflect on their effects to inform subsequent action (Mercer, 2011; van Lier, 2000, 2004, 2008). Self-regulated learning (SRL) provides a complementary perspective by emphasizing how learners regulate goals, strategies, monitoring, and self-evaluation during AI-mediated writing (Zimmerman, 2000). Recent evidence suggests that these regulatory processes are consequential for GenAI use: SRL strategies predict more advanced forms of AI-assisted writing (Jin et al., 2025), while EFL learners' self-regulated use of ChatGPT and machine translation is shaped by cognitive and sociocultural conditions, including instructional guidance and peer interaction (Wang, 2024). In the present study, GenAI output is therefore treated as a potential affordance: decisions to accept, modify, or reject suggestions represent selection, their textual implementation constitutes action, and reflection involves evaluating the completed revision in ways that may inform subsequent writing. SRL and learner agency are thus complementary but distinct: SRL foregrounds regulation and monitoring, whereas ecological agency foregrounds the perception of affordances and contextually situated choice and action.

Although Korean research has rarely examined learner agency directly in GenAI-mediated writing, related studies of self-regulation provide relevant evidence about how learners manage AI-supported tasks. Lee and Davis (2024) reported gains in goal setting, time management, self-assessment, and task strategies among Korean university students using ChatGPT. Dillon (2024) likewise documented metacognitive strategy use during prompt-literacy training, while also noting concerns about AI accuracy and the continued need for human interaction. Chung and Ahn (2025), however, found that planning, monitoring, and evaluation declined over time, with greater cognitive offloading among lower-proficiency learners. Together, these studies suggest that GenAI can support regulatory activity but does not necessarily sustain learners' active monitoring and control. Importantly, this evidence concerns processes related to, rather than direct measures of, learner agency. Moreover, existing Korean research has focused largely on university-level English learning. Little is known about how secondary-school EFL learners exercise agency when evaluating and revising GenAI-generated alternatives in creative writing. The present study addresses this gap through task-specific self-reports across guided and comparison conditions and process-level analysis of guided learners' revision decisions in English poetry writing.

2.3. Multimodal literacy and GenAI

Multimodal literacy concerns the interpretation and production of meaning through the coordinated use of semiotic resources such as language, image, sound, and layout (Jewitt & Kress, 2003; Lim et al., 2022). From a social semiotic perspective, these resources are not neutral additions to written language; their selection and combination are shaped by communicative purposes and social contexts (Kress, 2010; New London Group, 1996). Accordingly, multimodal literacy involves not only producing multimodal texts but also interpreting and evaluating how different semiotic resources contribute to meaning. Three closely related dimensions are particularly relevant: critical interpretation, creative production, and assessment (Lim et al., 2022). Critical interpretation involves examining how linguistic, visual, aural, and spatial resources shape meaning and developing a metalanguage for discussing their functions (Kress, 2010; Kress & van Leeuwen, 1996). In GenAI-mediated communication, such interpretation also has an epistemic dimension, as learners must judge whether generated content is accurate, complete, and consistent with intended meaning. Creative production involves the purposeful orchestration of semiotic

resources rather than their decorative combination (Kress, 2010; Lim et al., 2022), consistent with Swain's (1995) account of output as a site for noticing and learning. Assessment concerns how effectively these resources work together, yet classroom assessment often continues to privilege written language over other modes (Lim et al., 2022; Towndrow et al., 2013). This issue is particularly relevant in Korea, where the 2022 Revised National Curriculum emphasizes self-directed learning and critical engagement with digital information (Ministry of Education, 2022).

Empirical research in L2 education suggests that the value of multimodal composing lies not simply in access to multiple modes but in how learners coordinate them for meaning-making. Jiang (2017) showed that digital multimodal composing afforded EFL learners technological, educational, and social opportunities that complemented print-based literacy learning. More recently, Jiang and Lai (2025) examined GenAI-assisted digital multimodal composing among school learners aged 11–14 in Hong Kong. GenAI-supported groups achieved higher scores for purpose, navigation, and overall composition quality, whereas no significant differences emerged for base units, layout, or rhetorical relations. These findings suggest that GenAI support does not benefit all dimensions of multimodal composition equally and reinforce the importance of examining how learners evaluate and coordinate semiotic resources rather than assuming that access to GenAI itself enhances multimodal literacy.

In Korean EFL contexts, evidence remains limited for secondary-school learners and for the processes underlying multimodal composing. Park's (2025) review of 54 empirical studies found that Korean research has concentrated mainly on university and elementary learners, with comparatively little attention to secondary students, composing processes, assessment, or interactions among modes. Hwangbo and Bae (2024) provide an important secondary-school example: Korean high-school students created English picture books using AI-generated images, ChatGPT-supported writing, and synthesized audio, with gains in writing performance and self-efficacy but not in reading performance. Their study demonstrates the feasibility of GenAI-supported multimodal production in secondary English education, but its main outcomes concerned performance, AI literacy, and self-efficacy rather than learners' ongoing evaluation of semiotic choices. Accordingly, relatively little is known about how Korean secondary EFL learners evaluate semiotic resources during GenAI-mediated writing and revision. The present study addresses this process-level gap by examining task-specific perceptions of multimodal literacy and, within the guided condition, the criteria learners used to evaluate and revise linguistic and aural resources in relation to meaning, voice, and textual organization in GPT-revised poetry.

2.4. Poetry writing, learner agency, and multimodal literacy

Poetry writing provides a particularly useful context for examining multimodal literacy because even a printed poem coordinates multiple semiotic resources. Linguistic choices interact with aural resources such as rhyme, rhythm, meter, and sound patterning, as well as visuospatial resources such as lineation, stanzaic organization, punctuation, typography, and white space (Attridge, 1995; New London Group, 1996; Newfield & d'Abdon, 2015; van Leeuwen, 1999, 2006). These resources jointly shape a poem's meaning, voice, and effect; poetry is therefore multimodal not simply because multiple resources are present, but because writers must coordinate them purposefully in designing meaning (Kress, 2010; Kress & van Leeuwen, 2001). Attention to features such as rhyme, rhythm, meter, lineation, and stanza structure can also provide learners with a metalanguage for discussing how formal choices contribute to meaning (Cook, 1997; Hanauer, 2012). Poetry writing thus involves critical interpretation, creative production, and assessment as learners interpret, combine, and evaluate semiotic resources in relation to intended meaning and voice (Lim et al., 2022).

GenAI-mediated poetry revision brings these multimodal judgments into direct contact with learner agency. An AI-generated change may improve grammatical naturalness while disrupting rhythm, strengthen rhyme while altering intended meaning, or modify lineation in ways that affect the poem's presentation. Learners must therefore evaluate alternatives across interacting and sometimes competing criteria rather than simply determine whether a suggestion is linguistically correct. Such revision creates observable decision points at which learners can retain, modify, restore, or reject AI-generated alternatives. From an ecological perspective, these alternatives can function as affordances that learners perceive, select among, and act upon, while subsequent evaluation can inform later action (Mercer, 2011; van Lier, 2000, 2004; Zimmerman, 2000). Research on GenAI-assisted multimodal writing likewise highlights the importance of examining how learners use AI-generated resources during composing rather than assuming that access to such resources itself produces productive engagement (Jiang & Lai, 2025).

The present study therefore brings ecological learner agency and multimodal literacy together through GenAI-mediated poetry revision. Learner agency is represented through the perceive-select-act-reflect cycle, while multimodal literacy concerns learners' interpretation, production, and evaluation of the semiotic resources involved in poetic meaning-making (Kress, 2010; Lim et al., 2022). The two

perspectives address complementary aspects of the same revision process: agency concerns how learners respond to available alternatives, whereas multimodal literacy concerns the criteria through which those alternatives are evaluated and redesigned. This integrated framework informed the guided instruction, questionnaire dimensions, qualitative analysis of revision processes, and interpretation of the findings.

3. Method

3.1. Participants

The participants were 46 Korean high school students in Grades 11 and 12 enrolled in voluntary extracurricular English programs in the Seoul-Gyeonggi metropolitan area. They comprised two intact classes: a guided class ($n=23$) and a comparison class ($n=23$). The instructional differences between the two classes are described in Section 3.3. Because the classes were pre-existing rather than randomly assigned, similarities in observed background characteristics should not be interpreted as establishing baseline equivalence. Table 1 summarizes participant characteristics. Both classes had identical grade and gender distributions, with 11 Grade 11 students and 12 Grade 12 students, and 11 male and 12 female students in each class. Individual age data were not collected.

Table 1. Participants' characteristics

	Guided class ($n=23$)	Comparison class ($n=23$)
Grade 11 / grade 12	11 / 12	11 / 12
Male / female	11 / 12	11 / 12
Self-rated English proficiency	Approximately novice-high	Approximately novice-high

Note. Standardized English-proficiency scores were not collected; proficiency is reported as self-rated background information only.

All participants reported prior experience using ChatGPT, but none had used it specifically for English writing. They also reported receiving at least six hours of private English education per week. Their self-rated English proficiency was approximately novice-high; no standardized proficiency measure was administered. Participation was voluntary, and student assent and parental or legal-guardian consent were obtained before data collection. Participants received financial compensation.

3.2. Procedure

The study followed a pretest–instruction–immediate–posttest–delayed–posttest sequence over seven weeks, as illustrated in Figure 1. In Week 1, participants completed the self-report survey questionnaire. In Week 2, both classes participated in two 50-minute instructional sessions and received the same instruction in poetic form, while the guided class additionally received explicit instruction in evaluating and selectively using GenAI-generated revisions (see Section 3.3). The immediate posttest was conducted in Week 3 over two consecutive days: On Day 1, students completed the questionnaire and the GenAI-assisted poetry-writing task, and on Day 2, they completed the final handwritten revision task without access to digital devices. No additional treatment was provided during Weeks 4–6. In Week 7, participants completed the delayed posttest questionnaire, four weeks after the immediate posttest.

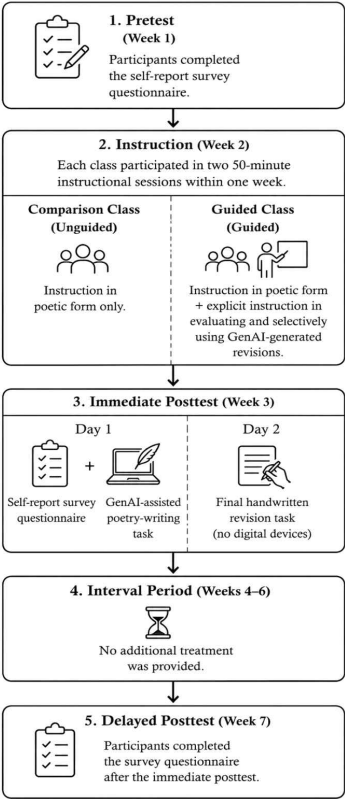


Figure 1. Procedure of the study.

For RQ1, questionnaire responses from both intact classes were compared across three measurement occasions—Weeks 1, 3, and 7. At each time point, participants completed the same 12-item Korean-language questionnaire online. The qualitative analysis for RQ2 focused on the guided class because written rationales documenting revision decisions were collected only from these students. Although students in both conditions produced final handwritten versions and submitted brief interpretations of their poems, equivalent rationale data were not collected from the comparison class. Accordingly, the process-level analysis for RQ2, which linked textual revisions to learners' articulated perceptions and decisions, was conducted within the guided class and was not used for a direct comparison of revision processes between the two conditions.

3.3. Instructional intervention

Instruction was delivered in two sessions over one week (Week 2), each lasting 50 minutes. Both classes received the same foundational instruction in poetic form, including rhyme, rhyme schemes (e.g., ABBA), meter, and rhythm as resources contributing to a poem's sound and musicality. Each session followed a common instructional sequence of teacher explanation, presentation of examples, and student practice using worksheets.

The instructional distinction between the two conditions concerned the guidance provided for evaluating and revising GenAI output. The comparison class was instructed to use GenAI to improve their draft poems but received no additional guidance on how to evaluate or revise the generated output beyond the common GenAI task instructions described below. The guided class received explicit instruction informed by multimodal literacy and learner agency.

Within the guided condition, multimodal literacy was incorporated across both instruction and revision. During Week 2, learners were taught to consider rhyme and other sound features, lexical choices, intended meaning, individual voice, and musicality as interrelated resources for poetic meaning-making. In the Week 3 GenAI-assisted writing task, they applied these criteria to GPT-revised poems by comparing AI-generated changes with their intended meaning and voice, considering their linguistic and aural effects, and deciding whether to retain, modify, or reject them. Consistent with the focus of the poetry task, this implementation emphasized linguistic and aural resources rather than the full range of visual-spatial modes.

When using GenAI, students were taught to (a) attend to rhyme and other sound

features as meaning-making resources; (b) evaluate AI-generated revisions in relation to their intended meaning, individual voice, lexical choices, and the poem's musicality; (c) selectively retain, modify, or reject AI suggestions; and (d) explain the reasons for their revision decisions. During the ChatGPT phase, guided learners were additionally instructed to ask ChatGPT why particular changes had been made, to question or discuss revisions they did not understand, and not to retain expressions whose meanings they could not themselves interpret. This instruction operationalized the perceive, select, and act stages of the agency framework at the decision points created by AI-generated suggestions. Learners were not required to make any particular revision or a prescribed number of changes; rather, they retained discretion over which suggestions to retain, modify, or reject and over the extent of their intervention. Because these evaluative options were explicitly foregrounded through instruction, however, the resulting revision behavior is interpreted as selective engagement within the guided task rather than as wholly spontaneous evidence of learner agency. Learners' specific choices and stated reasons therefore provide the primary basis for interpreting their engagement.

To maintain instructional consistency across the two conditions, the same teacher taught both classes using the same common instructional sequence and GenAI task instructions. The planned instructional difference was limited to the guided class's additional instruction in multimodal evaluation, selective use of AI-generated suggestions, and articulation of revision reasons. During the GenAI-assisted writing and final-revision phases, the same time limits, device-use rules, and procedures for collecting and returning materials were applied to both classes. The teacher supervised task completion but did not provide individualized feedback or advice on how particular poems should be revised.

Students accessed the free version of ChatGPT without logging into an account through their personal smartphones; the exact underlying model version was not recorded. The poetry-writing task proceeded in two stages on consecutive days. On the first day, students photographed their handwritten drafts and uploaded the images to ChatGPT. No fixed verbatim prompt was prescribed. In both conditions, however, students were instructed to ask ChatGPT to check the poem's language and rhyme, revise language, expressions, and rhyme where appropriate. Students could formulate the wording of the request themselves within these required components. They were given 15 minutes for the ChatGPT phase, and the number of requests or follow-up exchanges within this period was not restricted. Because students requested whole-poem revisions rather than line-by-line feedback, ChatGPT

typically produced substantially rewritten versions of their original drafts. The version available at the end of this 15-minute phase, following any permitted follow-up interaction, was retained as the GPT-revised version. The teacher then collected both the initial draft and the retained GPT-revised version. Intermediate prompts and ChatGPT responses were not retained.

On the following day, the teacher returned the initial draft and retained GPT-revised version to the students. Before the final revision began, students put away their personal smartphones and school-provided digital devices. Under teacher supervision, they then had 30 minutes to produce a handwritten final version while referring to the two archived versions, without further access to ChatGPT. All students also submitted a brief interpretation of their final poem as a procedural check on their understanding of the submitted text. Guided learners additionally submitted a written rationale explaining their revision decisions. The poem interpretations were not included in the RQ2 analysis.

3.4. Testing instruments

For RQ1, a brief, task-specific 12-item self-report questionnaire was developed from the integrated framework described in Sections 2.2–2.4. Drawing on the broader perceive–select–act–reflect framework, the questionnaire sampled perceive, select, and act, together with monitoring-related self-reports, for learner agency. For multimodal literacy, it sampled critical interpretation and creative production, the two dimensions most directly represented in the focal GenAI-mediated poetry task.

The learner agency component consisted of four two-item subscales. Three represented the perceive, select, and act stages of the agency framework (Mercer, 2011; van Lier, 2000, 2004), while the fourth measured monitoring, defined here as learners' review of their revisions and ability to explain their revision decisions. Monitoring was treated as conceptually distinct from reflection in the qualitative analysis, where reflection additionally required learners to identify what they had learned from the completed revision and how this learning would inform subsequent writing. Since several agency items asked what learners perceived they could notice, select, or do, they may partly reflect perceived capability or confidence rather than learner agency itself. The multimodal literacy component consisted of two two-item subscales: critical interpretation, including the critical evaluation of AI-generated text (Kress, 2010), and creative production (Kress, 2010; Lim et al., 2022). These items primarily targeted the aural and poetically formal dimensions emphasized in the

poetry task, particularly rhyme, meter, and rhythm, and did not comprehensively sample visuospatial features such as lineation, stanzaic arrangement, typography, or space.

All items were contextualized within GenAI-mediated poetry revision and rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Item 4 was negatively worded and reverse-coded as 6 minus the raw score. The items were initially written in Korean and pilot-tested once to assess their wording and clarity before administration. The pilot test was intended to assess item comprehensibility rather than to establish psychometric validity. Accordingly, the questionnaire was used as a brief self-report instrument specific to the focal GenAI-mediated poetry task, rather than as a comprehensive validated measure of learner agency or multimodal literacy. The complete questionnaire is provided in Appendix. The delayed posttest consisted only of the same self-report questionnaire; no performance-based writing or transfer task was administered at that time.

Written textual artifacts were collected from both conditions, including the initial handwritten draft, the retained GPT-revised version, and the handwritten final version produced on the following day. All students also submitted a brief interpretation of their final poem as a procedural check on their understanding of the submitted text. However, written rationales explaining revision decisions were collected only from the guided class.

Accordingly, the RQ2 process-level analysis used four sources from each guided learner: The initial handwritten draft, the retained GPT-revised version, the handwritten final version, and the written rationale. Since equivalent rationale data were unavailable for the comparison class, learners' perceptions and selection criteria could not be examined using the same analytic procedure, precluding a symmetric process-level analysis across conditions. The poem interpretations were not used for this purpose because they were collected only as a procedural check on comprehension rather than as accounts of revision decisions.

The final version was produced while learners referred to both the archived initial draft and the retained GPT-revised version. Because intermediate ChatGPT prompts and responses from the preceding GenAI phase were not retained, observable revision acts were identified through a systematic three-way comparison of the archived initial draft, the retained GPT-revised version, and the final version. It could not establish the ultimate provenance of wording appearing in neither archived version. Such wording was therefore described in relation to the archived texts rather than attributed to independent learner generation. The written rationales were used

as supplementary evidence to interpret the criteria and intentions associated with the observed revisions.

3.5. Analysis

For RQ1, questionnaire scores were analyzed using 2×3 mixed-design ANOVAs, with class (guided vs. comparison) as the between-subjects factor and time (T1, T2, and T3) as the within-subjects factor. The two composite indices were treated as the primary outcomes, whereas the subscale analyses were exploratory. Before inferential analyses, the reverse scoring of Item 4 was verified, and internal consistency and inter-item relationships were re-examined. Given the abbreviated and task-specific nature of the questionnaire, its scores were interpreted as self-reports of learners' perceptions and perceived capacities in the focal GenAI-mediated poetry task rather than as validated measures of stable learner-agency or multimodal-literacy constructs. The primary effect of interest was the Class $\times$ Time interaction. Baseline differences at T1 were examined using independent-samples t tests. Given the nonrandom assignment of the intact classes, separate ANCOVAs were conducted as sensitivity analyses for the T2 and T3 composite scores, with the corresponding T1 score entered as a covariate. Since these analyses could not control for unmeasured class-level differences or rule out possible regression-to-the-mean effects, the results were interpreted as comparisons of observed class trajectories rather than as estimates of causal instructional effects. When Mauchly's test indicated a violation of sphericity, Greenhouse–Geisser corrections were applied to the time and Class $\times$ Time effects. Significant interactions were followed by Bonferroni-adjusted simple-effects comparisons. Partial η^2 was reported for ANOVA and ANCOVA effects. Between-class comparisons were summarized using Cohen's d based on the pooled standard deviation, whereas within-class comparisons across measurement occasions were reported as d_{av} , calculated by dividing the mean difference by the average of the two occasions' standard deviations. For the principal pretest-to-immediate-posttest contrast, d_z , calculated by dividing the mean change by the standard deviation of the change scores, was also reported.

For RQ2, revision-process data were analyzed using a hybrid deductive–inductive approach. The perceive–select–act–reflect framework presented in Section 2.2 provided the *a priori* analytic structure, while the categories within each stage were refined through iterative comparison across the 23 cases. Four related coding

schemes were applied. First, uptake type represented the learner's overall response to the GPT-revised poem at the select stage and was classified as partial modified uptake, extensive reworking, draft restoration, or mixed uptake. Second, the object of perception captured the principal problem or affordance identified in each learner's rationale. When multiple criteria were mentioned, the dominant concern was used for learner-level classification, while co-occurring criteria were retained for qualitative analysis. Third, intervention behavior represented the act stage. An intervention act was defined as any identifiable difference between the retained GPT-revised and final versions, established through systematic three-way comparison with the initial draft archived. For wording appearing in neither archived version, coding captured its textual relation to those versions without attributing its ultimate provenance or independent generation to the learner. Unchanged GPT material was recorded separately as retention because it produced no observable revision act. Fourth, reflection was coded using two indicators: retrospective evaluation of the completed revision (REF_BACK) and the formulation of a strategy for subsequent writing (REF_FWD). Statements that explained only why a particular revision had been made were treated as verbalizations of the perceive-select-act process rather than as evidence of reflection.

Following the categorical analysis, a cross-case thematic analysis examined how these patterns were realized in individual revision processes. To move beyond the occurrence or amount of revision alone, the cross-case analysis also considered (a) the specificity of the reasons learners articulated for their revision decisions, (b) where applicable, the technical accuracy of their linguistic or poetic judgments, and (c) the correspondence between their stated rationales and the observable GPT-to-final revisions. Cases in which a rationale was general, a technical judgment was imprecise, or the stated reason accounted for only part of the observable revision were retained as boundary cases rather than treated as equivalent instances of criterial engagement. The resulting themes were interpreted in relation to motivated design, affordance-based action, learners' control over textual choices, and self-regulation (Hanauer, 2012; Kress, 2010; van Lier, 2000; Zimmerman, 2000). Both coders initially coded the full corpus to develop and refine the categories, consolidated the coding scheme through discussion, and then independently recoded all cases using the finalized codebook. Final percentage agreement for these categorical codes exceeded 95%, and the remaining disagreements were resolved through consensus. Excerpts from the Korean-language rationales quoted in the article were translated for the reporting, whereas poem excerpts were reproduced verbatim.

4. Results and Discussions

4.1. RQ1: Changes in self-reported multimodal literacy and learner agency

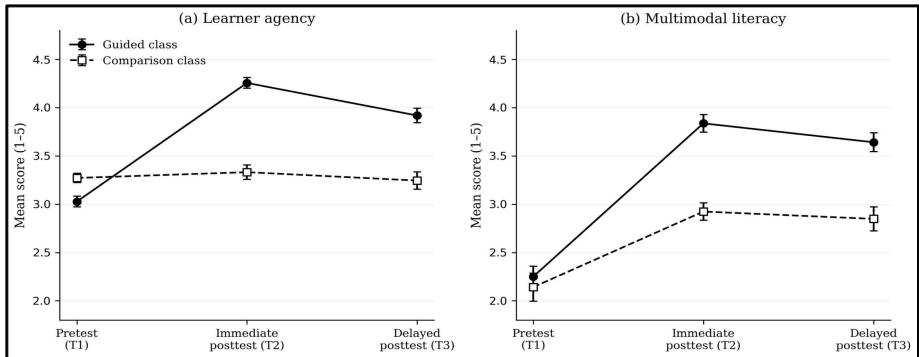
Internal consistency estimates for the abbreviated instrument were generally low. Across the six class-by-administration combinations, Cronbach's α ranged from .23 to .58 for the eight-item learner agency composite and from -.28 to .78 for the four-item multimodal literacy composite. The pooled Spearman-Brown coefficients for the two-item subscales ranged from .35 to .64. Re-examination of the raw responses confirmed that Item 4 had been correctly reverse-scored and that the reported reliability estimates were reproducible. Inspection of inter-item relationships further showed that some item pairs were weakly or negatively related within particular class-by-administration combinations, indicating limited internal consistency rather than a scoring error. Accordingly, as specified in Section 3.5, the composite indices were retained as the primary outcomes, the subscale analyses were treated as exploratory, and all scores were interpreted cautiously as brief self-reports of learners' perceptions and perceived capacities in the focal GenAI-mediated poetry task rather than as validated measures of stable constructs. Table 2 presents means and standard deviations by class and time.

Table 2. Descriptive statistics by class and time (n=23 per class)

Scale & subscale	Group	T1	T2	T3
		M (SD)	M (SD)	M (SD)
Learner agency	G	3.03 (0.27)	4.26 (0.27)	3.92 (0.35)
	C	3.27 (0.23)	3.33 (0.36)	3.24 (0.43)
Perceiving	G	3.26 (0.40)	4.63 (0.27)	3.93 (0.51)
	C	3.35 (0.35)	3.43 (0.48)	3.59 (0.83)
Selecting	G	2.83 (0.36)	4.17 (0.39)	3.57 (0.57)
	C	2.91 (0.44)	3.15 (0.53)	3.04 (0.77)
Acting	G	3.02 (0.53)	4.09 (0.42)	4.07 (0.48)
	C	3.52 (0.38)	3.46 (0.52)	3.04 (0.80)
Monitoring	G	3.00 (0.58)	4.13 (0.59)	4.11 (0.48)
	C	3.30 (0.36)	3.28 (0.65)	3.30 (0.91)
Multimodal literacy	G	2.25 (0.52)	3.84 (0.44)	3.64 (0.47)
	C	2.14 (0.70)	2.92 (0.44)	2.85 (0.60)
Critical interpretation	G	2.22 (0.62)	3.98 (0.55)	3.76 (0.54)
	C	2.41 (0.67)	3.13 (0.59)	2.85 (1.05)
Creative production	G	2.28 (0.67)	3.70 (0.54)	3.52 (0.70)
	C	1.87 (0.87)	2.72 (0.62)	2.85 (0.79)

Note. G=guided class; C=comparison class (unguided use). T1=pretest; T2=immediate posttest; T3=four-week delayed posttest.

At T1, the classes did not differ significantly on the multimodal literacy composite, $t(44)=0.60$, $p=.553$, $d=0.18$, or on the Perceiving, Selecting, and Critical Interpretation subscales (all $p\geq .308$). A significant difference emerged on the agency composite, with the guided class scoring lower (3.03 vs. 3.27), $t(44)=-3.29$, $p=.002$, $d=0.97$. Figure 2 displays the trajectories of the two composite scales by class.



Note. Error bars represent ± 1 standard error. G=guided class; C=comparison class (unguided use); $n=23$ per class.

Figure 2. (a) Self-reported learner agency and (b) multimodal literacy by class and time.

4.1.1. Mixed-design ANOVA

Table 3 summarizes the 2×3 mixed ANOVAs. The Class $\times$ Time interaction was significant and large for the agency composite, $F(2, 88)=41.62$, $p<.001$, $\eta^2=.49$, and significant for the multimodal literacy composite, $F(2, 88)=6.55$, $p=.002$, $\eta^2=.13$. At the subscale level, the interaction was significant for five of the six subscales ($\eta^2=.16-.34$); the exception was Creative Production, $F(2, 88)=1.66$, $p=.196$. Mauchly’s test indicated sphericity violations for Perceiving ($W=.79$, $p=.007$) and Selecting ($W=.84$, $p=.023$). Greenhouse–Geisser corrections applied to the time and Class $\times$ Time effects ($\epsilon=.83$ and $.86$, respectively) did not alter the significance of these effects (corrected $ps<.001$). Since the Monitoring subscale is conceptually distinct from reflection proper, the agency composite was also recomputed from the six perceive–select–act items alone; the Class $\times$ Time interaction was essentially unchanged, $F(2, 88)=37.33$, $p<.001$, $\eta^2=.46$.

Table 3. Mixed-design ANOVA results by scale and subscale

Scale & subscale	Group <i>F</i> (1, 44)	η^2	Time <i>F</i> (2, 88)	η^2	G $\times$ T <i>F</i> (2, 88)	η^2
Learner agency	63.27***	.59	47.38***	.52	41.62***	.49
Perceiving	34.73***	.44	23.24***†	.35	18.26***†	.29
Selecting	44.49***	.50	22.26***†	.34	10.86***†	.20
Acting	20.64***	.32	9.24***	.17	22.96***	.34
Monitoring	14.38***	.25	14.09***	.24	14.66***	.25
Multimodal literacy	63.06***	.59	58.45***	.57	6.55**	.13
Critical interpretation	25.86***	.37	36.78***	.46	8.29***	.16
Creative production	42.19***	.49	34.73***	.44	1.66	.04

Note. † Greenhouse–Geisser corrected (Perceiving $\epsilon=.83$; Selecting $\epsilon=.86$); *F* values are unaffected by the correction. ** $p<.01$. *** $p<.001$.

4.1.2. Follow-up and sensitivity analyses

First, baseline-adjusted sensitivity analyses were conducted using ANCOVAs, with the corresponding T1 score entered as a covariate. The class effect remained significant for learner agency at both T2, $F(1, 43)=72.64$, $p<.001$, $\eta^2=.63$, and T3, $F(1, 43)=29.36$, $p<.001$, $\eta^2=.41$. The class effect also remained significant for multimodal literacy at T2, $F(1, 43)=49.85$, $p<.001$, $\eta^2=.54$, and T3, $F(1, 43)=26.71$, $p<.001$, $\eta^2=.38$. Thus, the between-class differences at the immediate and delayed posttests remained after adjustment for the observed baseline scores. These analyses, however, do not rule out unmeasured class-level differences or possible regression-to-the-mean effects in this intact-class design.

Second, follow-up analyses of learner agency revealed sharply contrasting trajectories in the two classes. The guided class showed a significant change over time, $F(2, 44)=94.20$, $p<.001$, $\eta^2=.81$. Scores increased markedly from T1 to T2 ($d_{av}=4.58$, $d_z=2.88$, $p<.001$), declined significantly from T2 to T3 ($d_{av}=1.07$, $p=.005$), but remained substantially above the T1 level at T3 ($d_{av}=2.83$, $p<.001$). By contrast, the comparison class showed no statistically significant change over time, $F(2, 44)=0.41$, $p=.643$, $\eta^2=.02$, and none of the pairwise comparisons was significant. Between-class comparisons showed a crossover pattern: the guided class scored lower than the comparison class at T1 ($d=0.97$, $p=.002$), but higher at T2 ($M_s=4.26$ and

3.33, respectively; $d=2.91$, $p<.001$) and T3 ($d=1.71$, $p<.001$).

Third, multimodal literacy showed a different pattern, with both classes reporting increases from T1 to T2 but a larger increase in the guided class. The pretest-to-immediate-posttest effect sizes were $d_{av}=3.30$ and $d_z=2.30$ for the guided class and $d_{av}=1.34$ for the comparison class. Although the classes did not differ significantly at T1 ($d=0.18$), the guided class scored higher at both T2 ($M_s=3.84$ and 2.92 , respectively; $d=2.09$, $p<.001$) and T3 ($d=1.47$, $p<.001$). Neither class showed a statistically significant decline from T2 to T3.

4.1.3. Discussion for RQ1

The guided class demonstrated a marked increase in self-reported agency scores from T1 to T2, with scores remaining above baseline at T3, whereas the comparison class showed no statistically significant change across the three occasions. Given the use of one intact class per condition, this contrast should not be interpreted as demonstrating an instructional effect; rather, it documents different trajectories in learners' reported capacity to notice AI-generated alternatives, make selective decisions, and act on those decisions under the two classroom conditions. From an ecological perspective, agency depends not simply on the availability of technological affordances but on whether the learning environment makes meaningful choices salient and provides opportunities for learners to exercise them (Mercer, 2011; van Lier, 2000). One plausible mechanism is that the guided instruction made evaluative decision points more salient by explicitly directing learners to inspect GPT-generated changes, compare them with their intended meaning and voice, and decide whether to retain, modify, or reject them. This may have increased learners' perceived control over the revision process by positioning AI output as an object of monitoring and choice rather than as a default answer. This interpretation is consistent with recent SRL research showing that self-regulatory strategies predict more advanced forms of GenAI use in writing (Jin et al., 2025). It also accords with Chung and Ahn's (2025) Korean EFL study, in which learners displayed planning, monitoring, and evaluation during AI-assisted writing, but actual strategy use declined over time and cognitive offloading emerged, highlighting the importance of sustained support for strategic AI engagement. Because the present study did not directly test the mediating processes through which the instruction operated, this mechanism should be treated as a plausible interpretation of the observed trajectory rather than as an established causal pathway.

Both classes reported gains following their shared instruction in rhyme, rhyme schemes, meter, and rhythm, although the increase was larger in the guided class. The nonsignificant interaction for Creative Production was consistent with broadly similar changes across the two classes in their self-reported use of sound as a design resource.

A plausible mechanism for this shared change is that the common poetic-form instruction provided learners in both classes with explicit concepts and metalanguage for attending to rhyme, meter, rhythm, and related sound features, consistent with approaches that make formal and semiotic features available for explicit discussion (Hanauer, 2012; Kress, 2010; Lim et al., 2022). The guided condition, however, additionally required learners to evaluate these formal choices in relation to intended meaning, individual voice, lexical choices, and musicality. This additional evaluative focus may help explain why the stronger difference appeared in Critical Interpretation rather than Creative Production: the intervention foregrounded not simply the production of formal features but the judgment of how those features functioned in relation to meaning. This interpretation is consistent with multimodal-literacy accounts that distinguish critical interpretation of how semiotic choices contribute to meaning from the creative production of multimodal texts (Kress, 2010; Lim et al., 2022). This pattern is also compatible with recent GenAI-mediated multimodal research showing that the benefits of AI-supported composing are selective across dimensions rather than uniform (Jiang & Lai, 2025). It is also relevant to the Korean secondary-school context, where Hwangbo and Bae (2024) demonstrated the feasibility of integrating GenAI-supported multimodal production with English writing, AI literacy, and self-efficacy, underscoring the importance of instructional design in how learners engage with AI-supported meaning-making. These findings should be understood in relation to the questionnaire's relatively narrow coverage of multimodal literacy, which emphasized the aural and formal resources foregrounded in the poetry task rather than comprehensively assessing visual and spatial dimensions.

Given the limited internal consistency and restricted construct coverage of the abbreviated questionnaire, these results should be treated as preliminary evidence of contrasting self-reported trajectories rather than as stable measures of the underlying constructs or definitive instructional effects. Moreover, because each condition was represented by a single intact class, the observed differences may also reflect unmeasured pre-existing class characteristics. Since the delayed posttest consisted only of self-report measures, the persistence of elevated scores at T3 should

not be interpreted as evidence of durable change in enacted learner agency, improved writing ability, or transfer to writing without GenAI. RQ2 separately examines observable revision behaviors—including modification, rejection, and restoration of elements in GPT-revised texts—within the guided class.

4.2. RQ2: Guided learners’ engagement with GenAI-revised poems: Multimodal literacy and learner agency in revision

RQ2 examined how guided learners engaged with GPT-revised versions of their poems and how this engagement reflected multimodal literacy and learner agency. The analysis drew on 23 complete data sets, each comprising an initial draft, a GPT-revised version, the learner’s final version, and a written rationale.

4.2.1. Learner agency across the perceive–select–act–reflect cycle

The first strand of RQ2 examined which stages of the agency cycle were observable in learners’ revision processes and how each stage was realized. As summarized in Table 4, all 23 learners met the operational indicators for the perceive–select–act sequence: each verbalized at least one problem or affordance in the GPT-revised poem, made a codable uptake decision, and implemented at least one intervention.

Table 4. Frequency of operational indicators in the perceive–select–act–reflect cycle

Stage	Operational indicator	n	%
Perceive	At least one problem or affordance identified in the GPT-revised poem	23	100
Select	Codable decision to retain, modify, restore, or reject elements of the GPT revision	23	100
Act	At least one observable intervention in the final version	23	100
Reflect: possible retrospective evidence	Broader evaluation of the completed revision potentially relevant to REF_BACK	2	8.7
Reflect: complete loop	Retrospective evaluation accompanied by a forward-oriented strategy (REF_BACK + REF_FWD)	0	0

Note. REF_BACK=retrospective evaluation of the completed revision; REF_FWD=formulation of a strategy for subsequent writing. Possible REF_BACK evidence was coded cautiously because these statements could also be interpreted as extended rationales for completed revisions.

Since these evaluative and revision options had been explicitly foregrounded through instruction, these indicators are interpreted as engagement within the focal guided task rather than as wholly spontaneous behavior independent of instruction.

At the *perceive* stage, all 23 learners identified at least one aspect of the GPT-revised poem as warranting evaluation. Because individual rationales often invoked more than one criterion, each learner was classified according to the dominant concern underlying the intervention. As shown in Table 5, form or structure was the most frequent primary object of perception ($n=11$, 47.8%), followed by prosody or sound ($n=5$, 21.7%), meaning or narrative ($n=4$, 17.4%), and naturalness or lexis ($n=2$, 8.7%); one rationale was too general to classify ($n=1$, 4.3%). For example, P23 perceived an AI-expanded stanza as overly extended and judged that a longer expression weakened the rhyme, whereas P17 identified formal or unnatural wording that did not adequately convey her intended account of growth through mistakes.

Table 5. Primary perceived concern in the GPT-revised poem at the *perceive* stage

Primary perceived concern	Operational definition	n	%
Form and structure	Stanza organization, line length or economy, discourse order, and title	11	47.8
Prosody and sound	Rhyme, rhythm, and syllabic patterning	5	21.7
Meaning and voice	Preservation or development of intended meaning, theme, narrative, or individual voice	4	17.4
Naturalness and lexis	Idiomacity, lexical appropriateness, register, and grammatical naturalness	2	8.7
Unclassified	Rationale too general to permit classification	1	4.3

At the *Select* stage, no learner accepted the GPT-revised version verbatim or rejected it in its entirety. All learners instead showed selective uptake. Four subtypes were identified (Table 6): partial modified uptake ($n=11$, 47.8%), extensive reworking ($n=6$, 26.1%), draft restoration ($n=4$, 17.4%), and mixed uptake ($n=2$, 8.7%). Selective uptake was evident even in cases involving substantial textual retention. For example, although P01 retained 11 of the 12 GPT-revised lines, she reinstated her original title because she believed that “Toward Tomorrow” represented the poem’s overall content more accurately.

Table 6. Uptake decision at the select stage

Uptake pattern	Operational definition	n	%
Partial modified uptake	GPT-revised version retained as the primary textual base, with targeted changes to a minority of lines or discrete structural elements	11	47.8
Extensive reworking	GPT-revised version retained as an overall scaffold, but the majority of lines substantially changed in the final version	6	26.1
Draft restoration	Learner's initial draft reinstated as the primary textual base, with selective incorporation of GPT-generated elements	4	17.4
Mixed uptake	Different uptake patterns applied across stanzas or sections, with no single textual base predominating	2	8.7

Note. Uptake patterns were assigned at the learner level through three-way comparison of the initial draft, GPT-revised version, and final version. No learner accepted the GPT-revised version verbatim or rejected it in its entirety.

At the *Act* stage, 56 discrete intervention acts were identified (Table 7). Replacement of GPT wording was the most frequent behavior ($n=16$, 28.6%). Restoration of draft material, deletion or compression, and addition each occurred 12 times (21.4%), while reordering occurred four times (7.1%). Learners therefore did more than edit within the boundaries of the GPT-revised text: they also reinstated elements of their original drafts and included material appearing in neither the archived initial draft nor the retained GPT-revised version. For example, P18 restored the draft rhyme pair “rest–chest” in place of GPT’s “ache–shake” couplet, whereas P17 added new closing expressions that appeared in neither the archived initial draft nor the retained GPT-revised version.

Codable evidence for the *Reflect* stage was much less frequent. As shown in Table 8, twenty-one learners (91.3%) provided no evidence beyond explanations of particular revision decisions. Two learners produced broader retrospective evaluations that were coded cautiously as possible REF_BACK: P16 considered whether the completed revision had preserved the “honest mood” of her initial draft, and P17 evaluated the strengths and weaknesses of her original poem as a whole. These statements could also be interpreted as extended rationales, however, and neither learner articulated a forward-oriented strategy for subsequent writing. No case therefore showed evidence of REF_FWD or completion of the reflective loop. This result should be interpreted in relation to the elicitation prompt, which asked why learners had made particular revisions rather than what they had learned or

would do differently in a future task. It consequently represents an absence of codable evidence for reflection, not evidence that reflection did not occur.

Table 7. Intervention behaviors at the act stage (56 acts across 23 learners)

Intervention act	Operational definition	n	%
Replacement	GPT-generated wording replaced with alternative or substantially reformulated wording in the final version, excluding restoration from the initial draft	16	28.6
Restoration	Material from the learner’s initial draft reinstated in place of GPT-generated material	12	21.4
Deletion or compression	GPT-generated material removed or shortened without equivalent replacement	12	21.4
Addition	Material appearing in neither the archived initial draft nor the retained GPT-revised version	12	21.4
Reordering	Sequence of existing textual units, such as lines or stanzas, changed without substantial alteration to their wording	4	7.1

Note. Each observable difference between the retained GPT-revised and final versions was coded as one intervention act, based on three-way comparison with the archived initial draft. These categories describe textual relations among the archived versions and do not establish the ultimate provenance or independent authorship of wording absent from the first two versions. Percentages were calculated from the 56 identified acts.

Table 8. Distribution of codable reflection evidence

	Category	n	%
R0	rationale articulation only; no additional evidence of reflection	21	91.3
R1	possible retrospective reflection evidence (REF_BACK)	2	8.7
R2	forward strategy only (REF_FWD)	0	0
R3	complete loop (REF_BACK + REF_FWD)	0	0

Overall, the observable pattern across cases was $P \rightarrow S \rightarrow A \rightarrow$ rationale articulation. All learners made and implemented discernible decisions within the focal guided revision task, but no re-entry into a subsequent cycle through forward-oriented reflection was documented in the collected data.

4.2.2. Multimodal literacy

Three overlapping dimensions were coded at the learner level by considering the final revision together with the written rationale: semiotic critical interpretation, epistemic critical interpretation, and creative production. Unlike the *perceive* components reported in Section 4.2.1, these features were not mutually exclusive, because a single revision episode could simultaneously involve the interpretation, evaluation, and redesign of semiotic resources. As shown in Table 9, semiotic critical interpretation was identified in 22 of the 23 cases.

Table 9. Realization of multimodal literacy dimensions in learners’ revisions and rationales

Dimension	Operational evidence	n	%
Semiotic critical interpretation	Explicit connection between a linguistic, aural, or visuospatial feature and its perceived effect on meaning, sound, organization, or presentation	22	95.7
Epistemic critical interpretation	Explicit evaluation, verification, correction, selective acceptance, or rejection of at least one GPT-generated choice	23	100
Creative production	Observable redesign or addition of semiotic material in relation to the learner’s intended meaning, voice, organization, or formal effect	23	100

Note. The dimensions were coded independently and were not mutually exclusive; percentages therefore do not sum to 100%. Creative production refers to purposeful redesign and does not necessarily indicate an improvement in the linguistic or poetic quality of the final version.

Cases of semiotic critical interpretation connected a textual feature to its perceived effect on the poem. Such judgments concerned rhyme and rhythm, lexical length, title, stanza organization, discourse flow, imagery, and individual voice. P01, for example, evaluated GPT’s changes to both the title and the order of the final two stanzas.

- (1) Semiotic critical interpretation: title and discourse organization [P01]
 - a. Initial draft
Title: Toward Tomorrow
“*This is the path I chose today ...*” stanza
followed by
“*Dreaming about my future ...*” stanza
 - b. GPT-revised version

Title: Dream

“Dreaming of my future sure ...” stanza

followed by

“This is the path I choose with pride ...” stanza

c. Learner’s final version

Title: Toward Tomorrow

The original stanza order was restored.

d. Rationale

“I changed the title back to the draft’s title because I thought ‘Toward Tomorrow’ better represented the overall content of the poem. I also changed the order of the third and fourth stanzas so that the flow of the poem would be natural.” (translated)

P01 related the title to the poem’s global meaning and the spatial ordering of the stanzas to its discourse flow. The intervention therefore involved an interpretation of how linguistic and visuospatial organization contributed to the poem as a whole, rather than a preference for isolated words. Her removal of line-final punctuation to make the overall format more consistent provided further evidence that learners’ semiotic attention extended beyond the aural dimension emphasized during instruction.

Epistemic critical interpretation was coded in all 23 cases. In each case, at least one GPT-generated choice was evaluated rather than treated as inherently authoritative. A second revision episode from P01 illustrates this evaluation at the level of grammar and sound.

(2) Epistemic critical interpretation: verifying a GPT-generated expression [P01]

a. Initial draft

Brings new hope everywhere

b. GPT-revised version

Bring new hope, a healing light

c. Learner’s final version

Bringing new hope, a healing light

d. Rationale

“I changed Bring to Bringing to make a choice that satisfied both rhyme and grammar.” (translated)

The learner did not assume that GPT’s formulation was correct merely because it was machine-generated. Instead, she subjected the expression to explicit

grammatical and prosodic criteria and modified it accordingly. The grammatical basis of the change was reasonably supported because “*Bringing*” created greater structural parallelism with the following “*Guiding*”. The claim that the change also improved rhyme was less clearly supported, however, because the revision added a syllable. In this case, the learner’s rationale was specific and corresponded directly to the observable revision, but the technical basis of one part of that rationale was only partially supported.

Creative production was also identified in all 23 cases, although it ranged from highly localized compression or reordering to extensive rewriting and restoration of the initial draft. The clearest cases involved learners retaining GPT’s broad scaffold while redesigning its linguistic and semantic realization. In Example 3, the learner preserved GPT’s three-stanza structure and thematic progression but revised every line of the AI-generated version.

(3) Creative production: redesigning direct language as figurative imagery [P02]

a. Initial draft

Everything changes

Everything moves so fast

New things coming

Old things don’t last

b. GPT-revised version

Everything changes, everything flies

Racing forward before our eyes

New things rising, old things die

Nothing here is meant to stay

c. Learner’s final version

Everything shifts, everything soars

Rushing ahead beyond closed doors

Fresh sparks rising, fading lights fall

Nothing is ever meant to stay at all

d. Rationale

“I thought GPT’s opening was too simple, so I revised it to emphasize change and development. I used *beyond closed doors* to convey the uncertainty of the future, and I replaced the direct expressions with *fresh sparks* and *fading lights* to create a more figurative and poetic image of technological change.” (translated)

The final version retained GPT's macrostructure but transformed its relatively direct account of change into a more coherent figurative field. The replacements of *changes/flies* with *shifts/soars* and *new things/old things* with *fresh sparks/fading lights* altered the imagery, tone, and thematic representation of technological development. Creative production was thus realized through the purposeful transformation of available resources rather than through either complete acceptance or rejection of the GPT version. The written rationale specifically articulated reasons for several central changes rather than providing an exhaustive account of every intervention. For the changes it did address—such as the revised opening, “*beyond closed doors*,” and “*fresh sparks/fading lights*”—the stated intentions corresponded closely to the observable textual changes.

Across cases, however, the extent of textual intervention and the quality of the accompanying rationale did not always coincide. For example, P15 extensively reworked the GPT-revised poem but provided a specific reason for only one change, explaining that just escape felt abrupt and was therefore changed to trying to escape, while the remaining changes were described only generally as “content revision.” In a different pattern, P18 articulated specific judgments concerning rhythm, lexical repetition, grammar, and rhyme, but these explanations primarily accounted for GPT-generated changes that were retained in the final version; several observable restorations made between the GPT-revised and final versions were not explicitly addressed in the rationale. These cases indicate that rationale specificity, technical accuracy, and rationale–revision correspondence were analytically distinct aspects of learners’ engagement with the GPT-revised texts.

4.2.3. Discussion for RQ2

First, within the focal guided task, all 23 learners engaged in selective rather than wholesale uptake of GPT-generated revisions. Their rationales generally referred to identifiable criteria rather than simple preference or deference. However, the specificity of these rationales, the technical precision of the judgments expressed in them, and their correspondence with observable revisions varied across cases. This pattern is relevant to concerns that GenAI use without sufficient evaluative engagement may encourage reliance on fluent output (Fan et al., 2025; Kosmyna et al., 2025). At the same time, the uniform occurrence of revision should not by itself be taken as evidence of spontaneous learner agency. Although learners retained discretion over which GPT-generated elements to retain, modify, or reject, these

evaluative options had been explicitly foregrounded through instruction. The observed behavior therefore documents selective engagement within the guided task but cannot disentangle learners' spontaneous agency from the influence of the instructional context. Because equivalent rationales were not collected from the comparison class, RQ2 also does not permit a symmetric process-level comparison across conditions. The comparison condition should therefore not be treated as a qualitative counterfactual for the revision processes examined in RQ2.

Within the guided task, one plausible mechanism is that the instruction repositioned GPT-generated text from a potential endpoint of revision to a provisional object of evaluation. Learners were explicitly directed to compare GPT-generated changes with their intended meaning and individual voice, consider linguistic and aural qualities, and decide which elements to retain, modify, or reject. By making evaluative criteria and decision options salient, the guidance may have supported monitoring and selection before revision. This interpretation extends the ecological account of agency by specifying how task design may structure the perception and evaluation of available affordances (Mercer, 2011; van Lier, 2000). It is also consistent with recent AI-assisted revision research showing that behavioral uptake of ChatGPT suggestions and cognitive evaluation of those suggestions need not coincide (Koltovskaia et al., 2024), and with SRL research indicating that self-regulatory strategies are associated with more advanced forms of GenAI use in writing (Jin et al., 2025). In human-AI collaborative writing, Michel et al. (2025) similarly showed that comparing learner-produced texts with GenAI models can create occasions for revision and deliberation. The present study did not directly isolate these processes as mediators, however, so this mechanism remains an interpretation of the observed guided-task behavior rather than a demonstrated causal pathway.

Second, learner agency and multimodal literacy operated jointly in the revision process. Within the study's task-specific operationalization, agency-related engagement was reflected less in the amount of revision itself than in the selectivity of learners' decisions and in how those decisions were articulated and enacted. Importantly, the cross-case analysis showed that these dimensions did not necessarily converge: a learner could articulate a specific evaluative criterion yet implement it only partially accurately, make extensive revisions while explaining only some of them, or provide detailed reasons that corresponded more closely to retained GPT changes than to subsequent GPT-to-final interventions. This distinction between articulated evaluation and textual action parallels recent process-oriented research

showing that productive engagement with GenAI cannot be inferred from the mere presence or amount of revision. Koltovskaia et al. (2024), for example, distinguished behavioral engagement with ChatGPT feedback from learners' cognitive evaluation of that feedback, while Hwang et al. (2025) found that learners' prompting objectives and actual prompting approaches were frequently misaligned, with predominantly surface-level interaction corresponding to limited improvement in higher-order writing dimensions.

Multimodal literacy provided many of those criteria. Learners interpreted how linguistic, aural, and visuospatial choices affected meaning and poetic form, evaluated GPT-generated expressions as open to evaluation and verification rather than simply accepting them, and redesigned the available resources in accordance with their intended meanings and voices (Kress, 2010; Kress & van Leeuwen, 2001; New London Group, 1996). The observed relationship between learners' perceived concerns and the extent of their revisions further suggests that action on the AI-generated text was structured by what learners first noticed: concerns about form or sound were generally associated with localized intervention, whereas concerns about meaning or voice were associated with more extensive redesign. This pattern is consistent with the ecological view that agentic action is shaped by the affordances learners perceive in a given environment (van Lier, 2000). The guidance may therefore have operated not by prescribing particular revisions but by providing criteria through which learners could differentiate among GenAI-generated alternatives. In this sense, human-AI collaboration involved judgment over AI-generated possibilities rather than simple acceptance of AI-generated text.

Third, the findings also identify limits in the depth and verbalization of learners' engagement. Although the perceive-select-act sequence was observable across all cases, no learner articulated a forward-oriented strategy that would indicate completion of the reflective cycle. This absence should not be interpreted as evidence that reflection did not occur, because the rationale prompt asked learners why they had made particular revisions rather than what they had learned or would do in a subsequent writing task. In addition, several learners applied relevant criteria imprecisely or provided rationales that did not fully correspond to their observable interventions. These cases show that criterial awareness, technical precision, and systematic self-explanation did not always coincide within the focal task. Pedagogically, explicit attention should therefore extend beyond encouraging learners to question and revise AI output; it should also help them evaluate the consequences of their revisions and formulate strategies for future composing. This implication is

consistent with recent SRL research suggesting that productive GenAI use depends on learners' capacity to regulate their engagement with AI rather than merely access increasingly powerful forms of assistance (Jin et al., 2025). Prompts asking what learners learned, which criteria were most useful, and what they would examine first in a subsequent poem may help elicit the transfer-oriented reflection emphasized in accounts of agency and self-regulation (van Lier, 2004; Zimmerman, 2000).

5. Conclusions

This exploratory comparative mixed-methods study examined patterns of learner engagement with GenAI across guided and unguided conditions in two intact high-school classes. The findings revealed contrasting trajectories across the two classes, although the present design does not permit these differences to be attributed causally to the instructional condition. The comparison class showed no statistically significant change in self-reported learner agency across the three measurement occasions, whereas the guided class demonstrated substantial increases in both self-reported agency and multimodal literacy at the immediate posttest, with scores remaining above baseline four weeks later. These delayed scores indicate persistence in task-specific self-reported perceptions rather than demonstrated long-term improvement in enacted learner agency, writing ability, or transfer. Within the focal guided revision task, their revision data further showed that GPT-generated texts were treated as provisional resources for evaluation and redesign. Since evaluative choices had been explicitly foregrounded through instruction, however, these revision behaviors should be interpreted as selective engagement within the guided task, rather than as wholly spontaneous evidence of agency.

Learners interpreted the semiotic effects of choices involving sound, meaning, organization, and voice, evaluated AI-generated language rather than accepting it without evaluation, and creatively transformed available linguistic, aural, and visuospatial resources. Given the abbreviated instrument's limited internal consistency and construct coverage, these questionnaire findings should be understood as changes in task-specific self-reports rather than as evidence of change in the underlying constructs themselves.

The study makes three contributions. Theoretically, it operationalizes learner agency in AI-mediated revision through learners' selective decisions about what to retain, modify, restore, or reject and the criteria they articulated for those decisions.

It also extends multimodal literacy in GenAI-supported writing through the related dimensions of semiotic critical interpretation, epistemic critical interpretation, and creative production. Methodologically, the study demonstrates the value of a three-way comparison among learners' initial drafts, GPT-revised versions, and final texts, supplemented by written rationales. This approach makes observable textual differences across the archived versions visible even when complete interaction logs are unavailable, while not establishing the ultimate provenance of wording absent from the initial draft and retained GPT-revised version. Pedagogically, the findings suggest that the risks GenAI may pose to learner autonomy need not be addressed through restriction alone. The behavioral patterns observed in the guided class suggest that positioning AI output as an object of semiotic analysis, epistemic evaluation, and purposeful redesign warrants further investigation as a means of encouraging more selective and learner-directed engagement with GenAI.

These conclusions are subject to several methodological limitations. First, the use of one intact class per condition, together with the baseline difference in self-reported agency scores, limits causal interpretation; the ANCOVA results adjusted for the observed baseline scores but could not rule out possible regression to the mean or unmeasured class-level differences. Second, the abbreviated questionnaire showed limited internal consistency within individual administrations and provided only task-specific self-reports closely aligned with the instructional content: Some agency items may have reflected perceived capability or confidence rather than enacted agency, while the multimodal-literacy items emphasized aural and formal features of poetry and did not comprehensively represent visual and spatial dimensions. Third, the qualitative analysis was based on one focal guided revision task. Because evaluation and selective use of GPT output had been explicitly foregrounded during instruction, the observed revision behaviors cannot fully disentangle learners' spontaneous choices from the influence of the instructional context. Since written rationales documenting revision decisions were collected only from the guided class, the resulting asymmetry precluded a symmetric process-level comparison linking observable textual revisions to learners' articulated perceptions and decisions across the two conditions. The qualitative analysis also did not include interaction logs and used percentage agreement rather than a chance-corrected index of intercoder agreement. The retrospective rationale prompt was not designed to elicit future-oriented reflection, making it impossible to determine whether reflection was absent or occurred but was not verbalized. The delayed posttest consisted only of self-report measures and did not include a poetry-writing task completed without

GenAI or a transfer task in another genre; the delayed scores therefore cannot establish durable changes in enacted learner agency, writing performance, or transfer of the instructed practices. Finally, the small sample, single educational context, and focus on English poetry further constrain the transferability of the findings. Future research should include multiple classes in each condition, use more fully validated measures, retain learner–GenAI interaction logs, and directly elicit both retrospective and forward-oriented reflection. It should also incorporate delayed performance tasks completed without GenAI and cross-genre transfer tasks to examine whether learners apply the evaluative and revision strategies beyond the focal instructional task. Multi-class and repeated-task studies should further examine how guidance is enacted through teacher, peer, and learner–GenAI interaction, using interaction logs, screen capture, classroom recordings, or stimulated recall to trace how evaluative criteria and revision decisions develop over time. Where feasible, matched multi-class or cluster-randomized designs would strengthen causal inference, while replication across genres, proficiency levels, and instructional settings would test the broader applicability of the observed patterns.

Despite these limitations, the guided class provides an illustrative example of learners treating GenAI-generated text as an object of interpretation, evaluation, and redesign rather than accepting it uncritically within a structured instructional context.

References

- Attridge, D. (1995). *Poetic rhythm: An introduction*. Cambridge: Cambridge University Press.
- Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- Bok, E., & Cho, Y. (2023). Examining Korean EFL college students' experiences and perceptions of using ChatGPT as a writing revision tool. *Journal of English Teaching through Movies and Media*, 24(4), 15-27. <https://doi.org/10.16875/stem.2023.24.4.15>
- Chung, Y. H., & Ahn, B. K. (2025). Metacognitive strategy use in AI-assisted writing: A mixed-methods study of EFL college students' dialogues with ChatGPT. *Studies in English Education*, 30(2), 29-56. <https://doi.org/10.22275/SEE.30.2.02>
- Cook, G. (1997). Language play, language learning. *ELT Journal*, 51(3), 224-231. <https://doi.org/10.1093/elt/51.3.224>
- Dillon, T. (2024). Korean university students' prompt literacy training with ChatGPT: Investigating

- language learning strategies. *English Teaching*, 79(3), 123-157. <https://doi.org/10.15858/engtea.79.3.202409.123>
- Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489-530. <https://doi.org/10.1111/bjet.13544>
- Hanauer, D. I. (2012). Meaningful literacy: Writing poetry in the language classroom. *Language Teaching*, 45(1), 105-115. <https://doi.org/10.1017/S0261444810000522>
- Hwang, M., Jeens, R., & Lee, H.-K. (2025). Exploring learner prompting behavior and its effect on ChatGPT-assisted English writing revision. *The Asia-Pacific Education Researcher*, 34(3), 1157-1167. <https://doi.org/10.1007/s40299-024-00930-6>
- Hwangbo, E., & Bae, J. (2024). The impacts of creating English picture books using generative AI on English reading and writing skills, AI literacy, and self-efficacy of Korean high school students [In Korean]. *English Teaching*, 79(4), 79-98. <https://doi.org/10.15858/engtea.79.4.202412.79>
- Jewitt, C., & Kress, G. (Eds.). (2003). *Multimodal literacy*. New York: Peter Lang.
- Jiang, L. (2017). The affordances of digital multimodal composing for EFL learning. *ELT Journal*, 71(4), 413-422. <https://doi.org/10.1093/elt/ccw098>
- Jiang, L., & Lai, C. (2025). How did the generative artificial intelligence-assisted digital multimodal composing process facilitate the production of quality digital multimodal compositions: Toward a process-genre integrated model. *TESOL Quarterly*, 59(S1), S52-S85. <https://doi.org/10.1002/tesq.3390>
- Jin, F., Lin, C.-H., & Lai, C. (2025). Modeling AI-assisted writing: How self-regulated learning influences writing outcomes. *Computers in Human Behavior*, 165, 108538. <https://doi.org/10.1016/j.chb.2024.108538>
- Jin, Y., Yang, K., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2026). The agency gap in AI-supported writing: How reactive and proactive agent designs shape multimodal reasoning. *Computers and Education: Artificial Intelligence*, 11, 100655. <https://doi.org/10.1016/j.caeai.2026.100655>
- Kang, J., & Yi, Y. (2023). Beyond ChatGPT: Multimodal generative AI for L2 writers. *Journal of Second Language Writing*, 62, 101070. <https://doi.org/10.1016/j.jslw.2023.101070>
- 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, 101130. <https://doi.org/10.1016/j.jslw.2024.101130>
- Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). *Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task* (arXiv preprint). <https://doi.org/10.48550/arXiv.2506.08872>
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Oxon and New York: Routledge.
- Kress, G., & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. London:

Routledge.

- Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.
- Kultas, E. (2026). Integration of artificial intelligence into primary school classrooms: The role of ChatGPT in improving poetry writing skills and attitudes towards poetry. *BMC Psychology*, 14, 1086. <https://doi.org/10.1186/s40359-026-04844-4>
- Lee, J.-Y. (2025). A meta-analysis of ChatGPT's effects on English writing in Korean classrooms. *English Teaching*, 80(5), 175-193. <https://doi.org/10.15858/engtea.80.5.202512.175>
- Lee, Y. J., & Davis, R. O. (2024). A case study of implementing ChatGPT for university's general English courses: Focusing on English language learners' self-regulated learning and grit. *Secondary English Education*, 17(3), 73-87. <https://doi.org/10.20487/kasee.17.3.202406.73>
- Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence*, 9, 100445. <https://doi.org/10.1016/j.caeai.2025.100445>
- Lim, F. V., Toh, W., & Nguyen, T. T. H. (2022). Multimodality in the English language classroom: A systematic review of literature. *Linguistics and Education*, 69, 101048. <https://doi.org/10.1016/j.linged.2022.101048>
- Mercer, S. (2011). Understanding learner agency as a complex dynamic system. *System*, 39(4), 427-436. <https://doi.org/10.1016/j.system.2011.08.001>
- Michel, M., Bazhutkina, I., Abel, N., & Strobl, C. (2025). Collaborative writing based on generative AI models: Revision and deliberation processes in German as a foreign language. *Journal of Second Language Writing*, 67, 101185. <https://doi.org/10.1016/j.jslw.2025.101185>
- Ministry of Education. (2022). *2022 revised national curriculum for English* [In Korean]. Sejong: Ministry of Education of Korea.
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60-92. <https://doi.org/10.17763/haer.66.1.17370n67v22j160u>
- Newfield, D., & d'Abdon, R. (2015). Reconceptualising poetry as a multimodal genre. *TESOL Quarterly*, 49(3), 510-532. <https://doi.org/10.1002/tesq.239>
- Park, E. (2025). Multimodal composing in Korean EFL contexts: A systematic literature review [In Korean]. *Secondary English Education*, 18(5), 79-104. <https://doi.org/10.20487/kasee.18.5.202510.79>
- Park, H. (2024). ChatGPT-integrated English writing: Writing strategies and perceptions. *The Linguistic Association of Korea Journal*, 32(3), 85-103. <https://doi.org/10.24303/lakdoi.2024.32.3.85>
- Seo, J.-Y. (2024). Exploring the educational potential of ChatGPT: AI-assisted narrative writing for EFL college students. *Language Teaching Research Quarterly*, 43, 1-21. <https://doi.org/10.32038/ltrq.2024.43.01>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>

- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in Honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.
- Towndrow, P. A., Nelson, M. E., & Yusuf, W. F. B. M. (2013). Squaring literacy assessment with multimodal design: An analytic case for semiotic awareness. *Journal of Literacy Research*, 45(4), 327-355. <https://doi.org/10.1177/1086296X13504155>
- van Leeuwen, T. (1999). *Speech, music, sound*. London: Macmillan. <https://doi.org/10.1007/978-1-349-27700-1>
- van Leeuwen, T. (2006). Towards a semiotics of typography. *Information Design Journal*, 14(2), 139-155. <https://doi.org/10.1075/idj.14.2.06lee>
- van Lier, L. (2000). From input to affordance: Social-interactive learning from an ecological perspective. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 245-259). Oxford: Oxford University Press.
- van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Boston: Kluwer Academic Publishers. <https://doi.org/10.1007/1-4020-7912-5>
- van Lier, L. (2008). Agency in the classroom. In J. P. Lantolf & M. E. Poehner (Eds.), *Sociocultural theory and the teaching of second languages* (pp. 163-186). London: Equinox.
- Wang, Y. (2024). Cognitive and sociocultural dynamics of self-regulated use of machine translation and generative AI tools in academic EFL writing. *System*, 126, 103505. <https://doi.org/10.1016/j.system.2024.103505>
- Yang, I. (2024). A study on poetry writing using ChatGPT: Focusing on user perception and educational approaches [In Korean]. *The Journal of Learner-Centered Curriculum and Instruction*, 24(6), 619-634. <https://doi.org/10.22251/jlcci.2024.24.6.619>
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego: Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>

Rakhun Kim

Lecturer

Department of English Education

Hongik University

94 Wausan-ro, Mapo-gu, Seoul, 04066, Korea

E-mail: kimrhee02@hongik.ac.kr

Received: July 15, 2026

Revised version received: August 15, 2026

Accepted: August 15, 2026

Appendix. The 12-Item Questionnaire

Response format: 5-point Likert scale (1=strongly disagree, 5=strongly agree). Administered online in Korean at all three time points; English translations are provided for reporting.

Number	Subscale	Item description (Korean)	Reverse
1	Perceiving	I can notice how GPT's suggestions affect the rhyme or meter of my poem.	
2	Perceiving	I can identify specifically what was changed and how in GPT's suggested revision.	
3	Selecting	Even if GPT's suggestion looks smoother, I can reject it if it differs from my intention.	
4	Selecting	When GPT suggests a revision, I usually accept it as it is.	R
5	Acting	Rather than using GPT's suggestions as they are, I modify them to fit my intention.	
6	Acting	I selectively use only the parts of GPT's suggestions that I need.	
7	Monitoring	After revising, I reread my poem to check whether it conveys the feeling I originally intended.	
8	Monitoring	I can explain in words why I revised my poem the way I did.	
9	Critical interpretation	I can explain the sound structure of a poem using terms such as rhyme scheme (AABB, ABAB).	
10	Critical interpretation	Even when a GPT-generated line looks plausible, I can check for myself whether it follows metrical rules.	
11	Creative production	When writing a poem, I deliberately choose rhyme and rhythm to match the feeling I want to convey.	
12	Creative production	I can give reasons for my word choices in terms of sound (rhyme/meter).	

Note. Items 1–8 form the learner agency composite (Perceiving, Selecting, Acting, Monitoring); items 9–12 form the multimodal literacy composite. Item 4 is reverse-coded (6 – raw score).