Augmenting Rather Than Replacing: Effects of Generative AI Use Modes on Intrinsic Motivation and Writing Performance in College English Writing

Yousong Hu[1, a], Yi Ma[1, b], Hongmei Yu[1, c], Letong Gao[1, d], Yaxin Xiao[1, e], Yang Liu[2, f*]

[1] School of Educational Science, Chengdu Normal University, Chengdu, 611130, China
[2] Publicity Department of School Beijing City University, Beijing, 100083, China,

[a]1509756249@qq.com, [b]3084866715@qq.com, [c]3084866715@qq.com, [d]1103721238@qq.com, [e]1559745062@qq.com, [*f]649192680@qq.com

Abstract. The rapid expansion of generative artificial intelligence has created both new opportunities and new tensions in foreign language writing instruction. This study examined how three instructional modes—traditional instruction, substitution-oriented GenAI use, and augmentation-oriented GenAI use—were associated with college students' intrinsic motivation and writing performance. A quasi-experimental design was employed with 120 first-year non-English majors. To strengthen internal validity, instructional time, writing topics, and teacher effects were controlled across groups. Results showed that the augmentation-oriented group significantly outperformed the other two groups in intrinsic motivation, writing performance, and short-term knowledge transfer. Notably, the substitution-oriented group scored significantly lower than the control group on the end-of-intervention transfer task, suggesting a possible short-term transfer disadvantage rather than evidence of long-term retention loss. ANCOVA further indicated that the advantage of the augmentation-oriented condition in intrinsic motivation remained significant after controlling for pretest intrinsic motivation and English proficiency. In addition, mediation analysis showed that intrinsic motivation partially mediated the relationship between augmentation-oriented GenAI use and writing performance. These findings suggest that GenAI functions most productively as a cognitive support tool when instructional design preserves learner authorship and active engagement, although the study did not directly observe students' interaction processes with AI.

Keywords: generative artificial intelligence, intrinsic motivation, college English writing, human–AI collaboration, knowledge transfer, writing performance

1 Introduction

Generative artificial intelligence (GenAI), especially large language model-based tools, is rapidly reshaping higher education. In foreign language learning, GenAI provides

Z. Ge et al. (eds.), *Proceedings of the 2026 6th International Conference on Education, Information Management and Service Science (EIMSS 2026)*, Atlantis Highlights in Computer Sciences 25,
https://doi.org/10.2991/978-94-6239-766-8_10

learners with immediate access to text generation, language refinement, feedback, and interactive support. These functions are particularly relevant to college English writing instruction, where students often struggle with idea generation, rhetorical organization, lexical choice, and revision. At the same time, while GenAI may lower the threshold for writing, it may also alter learner engagement in ways that are not uniformly beneficial.

A central concern is that students may use GenAI not as a scaffold for learning, but as a substitute for their own thinking. When learners rely heavily on AI to generate content directly, they may complete tasks more quickly yet become less involved in planning, monitoring, and revising their own writing. Recent work has raised similar concerns. For example, Fan et al. argued that overreliance on generative AI may foster metacognitive laziness by reducing learners' willingness to engage in effortful self-regulation [1]. More broadly, recent research on student adaptation to GenAI suggests that AI use may result in deskilling, reskilling, or upskilling, depending on how learners appropriate the technology and how instructional contexts shape that use [2].

This distinction points to two pedagogically meaningful modes of GenAI use. In substitution-oriented use, students employ GenAI to generate substantial portions of text directly, thereby outsourcing core aspects of the writing task. In augmentation-oriented use, by contrast, students use GenAI as a support tool for brainstorming, idea expansion, prompt refinement, and feedback-informed revision while remaining the principal authors of the final text. The latter mode is more consistent with learner-centered pedagogy and is more likely to support active short-term writing development when learners remain responsible for planning, drafting, and revision.

In second language writing, motivation is especially important because writing is a resource-intensive activity that requires persistence, self-regulation, and repeated revision. Thus, motivation provides a useful bridge between instructional design and observable writing outcomes. If GenAI changes how students experience effort, control, and competence during writing, it is likely to affect not only performance but also motivation. Yet the relationship between GenAI use mode and motivation remains underexamined in college English writing contexts. Existing studies on GenAI and L2 writing have increasingly documented the importance of engagement with AI-supported feedback and the role of pedagogical framing in shaping student experience, but direct comparison between substitution-oriented and augmentation-oriented use remains limited [3-5].

The present study therefore investigated the effects of traditional instruction, substitution-oriented GenAI use, and augmentation-oriented GenAI use on college students' intrinsic motivation (IM), writing performance (WP), and short-term knowledge transfer (KT). It also examined whether intrinsic motivation mediated the relationship between GenAI use mode and writing performance. In doing so, the study aimed to clarify not only whether GenAI benefits writing instruction, but also how different pedagogical configurations of AI use may produce qualitatively different outcomes.

2 Literature Review and Hypoethesis

2.1 GenAI Use Modes in Educational Writing Contexts

The educational effects of GenAI depend not only on the technology itself but also on how learners use it. In writing instructions, this issue is especially consequential because writing development depends on sustained cognitive engagement rather than mere task completion. A substitution-oriented use mode may improve apparent efficiency but reduce learners' authorship, strategic thinking, and reflection. By contrast, an augmentation-oriented mode preserves the learner's central role and uses AI to extend, rather than replace, cognitive work.

This distinction aligns with recent debates on human–AI adaptation. Yang et al. argued that GenAI dependence may lead to deskilling, reskilling, or upskilling depending on learner agency and patterns of appropriation [2]. In L2 writing, the issue is therefore not simply whether students use AI, but whether AI functions as a shortcut or a scaffold. Recent writing research likewise shows that learners' engagement with AI-supported writing feedback is shaped by how AI is incorporated into revision and reflection processes rather than by AI access alone [3-5].

Accordingly, augmentation-oriented use should be more educationally productive than substitution-oriented use. Hypothesis 1: Students in the augmentation-oriented GenAI group will demonstrate significantly better writing performance and short-term knowledge transfer than students in the substitution-oriented GenAI group and the traditional instruction control group.

2.2 Intrinsic Motivation in AI-Supported Writing

Intrinsic motivation is a key psychological condition for successful writing development. In language learning, motivated students are more likely to invest effort, tolerate uncertainty, and persist in revising their work. From the perspective of self-determination theory, learning environments that support autonomy and competence tend to strengthen intrinsic motivation [6]. Augmentation-oriented use may support both. When students use AI to refine ideas, test rhetorical alternatives, and obtain feedback while remaining in control of the writing process, they may experience greater agency and efficacy. Substitution-oriented use, however, may weaken such experiences because success is attributed less to the learner's own effort than to AI-generated output.

Fan et al. suggested that AI-supported learning may improve visible performance while simultaneously weakening deep engagement if learners defer cognitive work to the tool [1]. This insight implies that the motivational consequences of GenAI are likely to vary by use mode. Recent L2 writing studies similarly suggest that GenAI is pedagogically valuable when it supports revision, reflection, and feedback engagement, but less so when it displaces active composing [4,5]. Hypothesis 2: Augmentation-oriented GenAI use will be associated with higher intrinsic motivation than substitution-oriented GenAI use and traditional instruction.

2.3 Intrinsic Motivation in AI-Supported Writing

Instructional conditions shape learning outcomes not only directly but also through motivational pathways. In writing, intrinsically motivated students are more likely to revise deeply, sustain effort, and remain engaged with the task. If augmentation-oriented GenAI use enhances learners' intrinsic motivation, then part of its positive effect on writing performance should operate through this increase in motivation. Intrinsic motivation was therefore treated as a theoretically grounded but nonexclusive mediator; other psychological mechanisms were outside the tested model. Hypothesis 3: Intrinsic motivation will mediate the relationship between augmentation-oriented GenAI use and writing performance

3 Methodology

3.1 Design

This study employed a quasi-experimental design. Three intact classes were selected, and each class was assigned to one of three instructional conditions. The primary independent variable was GenAI use mode, operationalized as a multicategorical variable with three levels: control, substitution-oriented use, and augmentation-oriented use. Because the study relied on intact classes rather than random assignment of individual students, the findings should be interpreted as evidence from a classroom-based quasi-experiment rather than a fully randomized trial.

3.2 Participants and Procedure

Participants were 120 first-year non-English majors whose English proficiency ranged from elementary to lower-intermediate levels. The three intact classes were uneven in size, which reflected the natural class composition in the institutional setting: 30 students were assigned to the control group, 45 to the substitution-oriented group, and 45 to the augmentation-oriented group. To strengthen treatment fidelity and internal validity, several controls were implemented. First, all three groups were taught by the same instructor, thereby minimizing teacher-effect bias. Second, total instructional time and writing tasks were kept equivalent across groups, with all students receiving two 45-minute sessions per week. Third, a standardized classroom protocol was used to guide permitted GenAI use in the two experimental groups, but the study did not collect detailed interaction traces such as prompt logs, revision histories, or screen recordings.

The study spanned six weeks in total. In Week 1, students completed the pretest measures. Weeks 2–6 constituted the instructional phase, during which the three conditions were implemented. The control group received traditional process-oriented writing instruction without GenAI support. The substitution-oriented group was instructed to use GenAI (Wenxin Yiyan) primarily to generate full-text drafts from prompts, with only minimal modification. The augmentation-oriented group received explicit training in scaffolded human–AI collaboration, focusing on prompt design for

brainstorming, rhetorical comparison, feedback interpretation, and revision while maintaining authorship of the final text. The post-test and short-term transfer task were administered at the end of Week 6.

Before the intervention, group equivalence was examined using pretest measures of intrinsic motivation and English proficiency. Pretest intrinsic motivation scores and Oxford Placement Test (OPT) scores were later included as covariates in the ANCOVA analysis.

3.3 Measures

Intrinsic Motivation.

Intrinsic motivation was assessed using a revised IMI-based instrument adapted from the Academic Motivation Scale [12]. To better reflect the context of English writing in this study, the original items were modified and refined, resulting in a 13-item scale that measured students' interest, enjoyment, and active involvement in writing tasks. Responses were collected on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), and higher scores represented stronger intrinsic motivation. The reliability of the adapted scale was acceptable in the present study (Cronbach's $\alpha = .756$). The focus on intrinsic motivation was consistent with the study's theoretical foundation and with recent findings in AI-supported learning research [1].

Writing Performance.

Writing performance was assessed using an analytic scoring rubric adapted from ESL Composition Profile [11] and supported by more recent rubric revision research [10]. The rubric employed a 100-point scale and evaluated five dimensions of writing performance: content, organization, vocabulary, language use, and mechanics. This multidimensional scoring approach was adopted to provide a systematic assessment of both higher-order and lower-order aspects of students' English writing. All essays were independently rated by two experienced English teachers, each with more than five years of teaching experience, under double-blind conditions. Interrater reliability was strong, with an intraclass correlation coefficient (ICC) of .88.

Knowledge Transfer.

Knowledge transfer was assessed through a cross-topic English argumentative writing task administered at the end of the instructional phase. This task was designed to examine the extent to which students could transfer and apply the writing knowledge, genre awareness, and strategies developed during the intervention to a new writing context. Because the task was administered immediately after the intervention, it measured short-term transfer performance rather than long-term retention of learning outcomes. Higher scores indicated a stronger ability to generalize prior learning beyond the immediate instructional task. Such a transfer-oriented assessment is consistent with the view that writing transfer involves the reuse and reshaping of previously acquired knowledge in response to novel rhetorical situations [7], and with evidence that learners can apply acquired metacognitive and genre-related strategies when performing subsequent writing tasks [8]. This focus also aligns with recent GenAI research emphasizing

whether AI-supported learning can promote performance generalization beyond immediate task completion [1].

English Proficiency.

Participants' English proficiency was assessed using the Oxford Placement Test prior to the intervention. Consistent with recent research supporting Oxford placement testing for ability grouping and proficiency alignment in university EFL contexts [13], the test was administered to establish baseline comparability across groups. In the present study, OPT scores were used both as a screening measure and as a covariate in subsequent analyses to control for potential differences in initial English proficiency.

3.4 Weekly Instructional Tasks

The intervention was organized as a progressively scaffolded instructional sequence, drawing on the process-genre approach to L2 writing instruction [8]. As shown in Table 1, the tasks for Weeks 2–6 moved from sentence-level practice to full-text production and short-term transfer-oriented application. This progression was intended to support students' gradual development of writing knowledge, metacognitive strategy use, and genre awareness. Meanwhile, students in the augmentation-oriented group participated in scaffolded human–AI collaboration, with GenAI assistance integrated into each stage of the writing process according to the instructional focus of the week.

Table 1. Weekly instructional tasks during the intervention.

	Writing focus	GenAI intervention focus in the augmentation-oriented group
W2	Topic sentences	Role-setting in prompts and idea expansion
W3	Supporting ideas	Logic-chain checking and evidence elaboration
W4	Conclusion writing	Tone enhancement and restatement of the core claim
W5	First full draft	Cohesion improvement and academic lexical substitution
W6	Final revision and transfer test	Deep polishing and critical evaluation of AI output

3.5 Data Analysis

Data were analyzed in four stages. First, preliminary analyses were conducted, including Shapiro–Wilk tests for normality, Levene's tests for homogeneity of variance, and one-way ANOVAs for pretest group equivalence. Second, descriptive statistics and one-way ANOVAs were used to examine posttest group differences across the three instructional conditions, followed by LSD post hoc comparisons. Third, ANCOVA was conducted to test the effect of GenAI use mode on posttest intrinsic motivation while controlling for pretest intrinsic motivation and English proficiency. Finally, a mediation model was tested using PROCESS Model 4 with 5,000 bootstrap samples. For the mediation analysis, GenAI mode was recorded as a binary planned contrast comparing the augmentation-oriented condition with the two non-augmentation conditions (0 = control/substitution, 1 = augmentation), because the theoretical focus centered on whether scaffolded GenAI use conferred advantages beyond both traditional instruction and substitution-oriented use [9]. This recoding was used only for the mediation model;

descriptive analyses and ANOVA comparisons retained the original three-group design.

4 Results

4.1 Preliminary Analysis

Before conducting the main analysis, assumption checks were performed. Shapiro–Wilk tests indicated that the outcome variables were normally distributed, and Levene's tests suggested that the assumption of homogeneity of variance was satisfied for all posttest outcome variables (all *p*s > .05).

Pretest group equivalence was also examined. One-way ANOVA showed no significant between-group differences in pretest intrinsic motivation, $F(2, 117) = 1.12$, $p = .32$, or pretest English proficiency, $F(2, 117) = 0.95$, $p = .41$, indicating that the three groups were comparable prior to the intervention.

4.2 Descriptive Statistics and Posttest Group Differences

Table 2 presents the descriptive statistics and one-way ANOVA results for the posttest outcomes across the three instructional groups.

The ANOVA results showed a significant main effect of instructional mode on all three focal posttest measures. Specifically, significant group differences were found for intrinsic motivation, $F(2, 117) = 42.15$, $p < .001$, partial $\eta^2 = .41$; writing performance, $F(2, 117) = 35.82$, $p < .001$, partial $\eta^2 = .38$; and short-term knowledge transfer, $F(2, 117) = 29.41$, $p < .001$, partial $\eta^2 = .32$.

LSD post hoc comparisons indicated that the augmentation-oriented group scored significantly higher than both the control group and the substitution-oriented group across all three outcome variables (all *p*s < .001). For writing performance, the mean difference between the augmentation-oriented group and the substitution-oriented group was 13.13 points (*SE* = 1.34). By contrast, the substitution-oriented group scored significantly lower than the control group in knowledge transfer ($MD = -7.61$, $p < .01$), while the group difference in intrinsic motivation approached significance ($MD = -0.27$, $p = .061$).

Taken together, the results revealed clearly differentiated posttest patterns across the three instructional groups, with the augmentation-oriented group obtaining the highest scores on intrinsic motivation, writing performance, and short-term knowledge transfer.

Table 2. Descriptive statistics and ANOVA results for posttest outcomes.

Variable	Control	Substitution-oriented	Augmentation-oriented	*F*	*p*	*Partial η²*
IM	3.42 (0.55)	3.15 (0.62)	4.25 (0.48)	42.15	< .001	.41
WP	72.42 (5.8)	68.52 (7.2)	81.65 (6.4)	35.82	< .001	.38
KT	65.82 (6.2)	58.21 (8.4)	74.56 (5.9)	29.41	< .001	.32

[a] IM = intrinsic motivation, WP = writing performance, KT = short-term knowledge transfer. Control group: *n* = 30, substitution-oriented group: *n* = 45, augmentation-oriented group: *n* = 45.

4.3 ANCOVA on Posttest Intrinsic Motivation

To further examine the effect of instructional mode on posttest intrinsic motivation while controlling baseline differences, an ANCOVA was conducted with posttest intrinsic motivation as the dependent variable, instructional mode as the independent variable, and pretest intrinsic motivation and OPT scores as covariates.

The analysis revealed a significant main effect of instructional mode on posttest intrinsic motivation, $F(2, 115) = 15.64$, $p < .001$, partial $\eta^2 = .21$. As shown in Table 3, the adjusted mean intrinsic motivation score was highest in the augmentation-oriented group ($M = 4.25$, $SE = 0.09$, 95% CI [4.07, 4.43]), followed by the substitution-oriented group ($M = 3.71$, $SE = 0.09$, 95% CI [3.53, 3.89]), and the control group ($M = 3.40$, $SE = 0.11$, 95% CI [3.18, 3.62]). Post hoc LSD comparisons showed that the augmentation-oriented group scored significantly higher than the control group ($MD = 0.85$, $p < .001$) and the substitution-oriented group ($MD = 0.54$, $p < .01$). The difference between the substitution-oriented group and the control group was not significant ($p = .12$). These results indicate that the advantage of the augmentation-oriented condition in intrinsic motivation remained significant after controlling for baseline motivation and English proficiency.

Table 3. Adjusted means for posttest intrinsic motivation from ANCOVA.

Group	*N*	Observed Mean (*SD*)	Adjusted Mean (*SE*)	95% Confidence Interval
Control	30	3.42 (0.55)	3.40 (0.11)	[3.18, 3.62]
Substitution-oriented	45	3.15 (0.62)	3.71 (0.09)	[3.53, 3.89]
Augmentation-oriented	45	4.25 (0.48)	4.25 (0.09)	[4.07, 4.43]

[b] Adjusted means are estimated marginal means controlling for pretest intrinsic motivation and English proficiency (OPT scores).

4.4 Correlation Analysis

Prior to the mediation analysis, Pearson correlation coefficients were computed to examine the associations among the primary study variables. As shown in Table 4, GenAI mode was dummy coded as a planned contrast (0 = control/substitution; 1 = augmentation). This coding tested the theoretically central augmentation effect and should not be interpreted as evidence that the control and substitution-oriented conditions were identical.

The augmentation-oriented mode was positively correlated with intrinsic motivation, $r = .48$, $p < .01$, writing performance, $r = .52$, $p < .01$, and short-term knowledge transfer, $r = .45$, $p < .01$. In addition, intrinsic motivation was positively associated with writing performance, $r = .58$, $p < .01$, and short-term knowledge transfer, $r = .51$, $p < .01$. Writing performance and short-term knowledge transfer were also strongly correlated, $r = .62$, $p < .01$.

Table 4. Correlations Among Study Variables.

Variables	1	2	3	4
1. GenAI planned contrast (augmentation = 1)	1			

Variables	1	2	3	4
2. IM	.48**	1		
3. WP	.52**	.58**	1	
4. KT	.45**	.51**	.62**	1

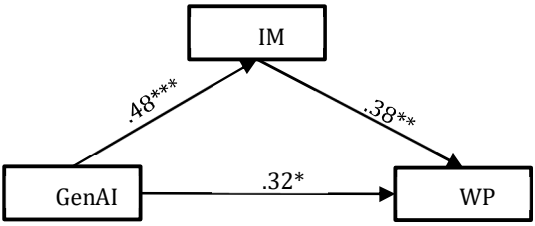
[c] GenAI = generative artificial intelligence. $N = 120$. ** $p < .01$.

4.5 Mediation Analysis

To test Hypothesis 3, a mediation model was estimated to be using PROCESS Model 4. GenAI mode was entered as the planned-contrast independent variable, intrinsic motivation as the mediator, and writing performance as the dependent variable. The mediation pathway is illustrated in Figure 1.

As shown in Table 5, the total effect of GenAI mode on writing performance was significant, $c = .50$, $SE = .08$, $t = 6.25$, $p < .001$, 95% CI [.34, .66]. GenAI mode also significantly predicted intrinsic motivation, $\beta = .48$, $SE = .07$, $t = 6.86$, $p < .001$, 95% CI [.34, .62]. When intrinsic motivation was included in the model, the direct effect of GenAI mode on writing performance remained significant but was reduced, $c' = .32$, $SE = .09$, $t = 3.55$, $p = .012$, 95% CI [.14, .50]. At the same time, intrinsic motivation significantly predicted writing performance, $\beta = .38$, $SE = .11$, $t = 3.45$, $p < .001$, 95% CI [.16, .60].

The bootstrapped indirect effect was .18, with a 95% confidence interval of [.08, .31], which did not include zero. Within the planned contrast, these results indicate a significant partial mediation effect, with intrinsic motivation accounting for approximately 36% of the total effect. As shown in Figure 1, augmentation-oriented GenAI use was associated with better writing performance both directly and indirectly through stronger intrinsic motivation.



[d] GenAI was coded as a planned contrast (0 = control/substitution, 1 = augmentation). Standardized coefficients are reported.

Fig. 1. Mediation model of intrinsic motivation in the relationship between GenAI use mode and writing performance.

Table 5. Mediation results for intrinsic motivation between GenAI mode and writing performance.

Path / Model	*β*	*SE*	*t*	*p*	95% CI
Model 1: WP (total effect)					
GenAI → WP	.50	.08	6.25	< .001	[.34, .66]
Model 2: IM					

Path / Model	*β*	*SE*	*t*	*p*	95% CI
GenAI → IM	.48	.07	6.86	< .001	[.34, .62]
Model 3: WP (direct and indirect effects)					
GenAI → WP (*c′*)	.32	.09	3.55	.012	[.14, .50]
IM → WP	.38	.11	3.45	< .001	[.16, .60]
Indirect effect (bootstrap)	**Effect**	**Boot *SE***	**Boot 95% CI**		
GenAI → IM → WP	.18	.06	[.08, .31]		

[c] The indirect effect was significant based on 5,000 bootstrap samples.

5 Discussion

5.1 Why Was Augmentation-Oriented GenAI Use More Effective?

The most salient finding of this study is that augmentation-oriented GenAI use was significantly more effective than substitution-oriented use in promoting both intrinsic motivation and writing performance. This result supports the interpretation that the pedagogical value of GenAI lies less in automating task completion than in supporting cognitively productive engagement.

In the substitution-oriented condition, students were allowed to use GenAI to generate substantial portions of text directly. While this may have reduced immediate task difficulty, it likely also weakened learners' sense of authorship and responsibility for the writing process. As a result, students may have experienced lower levels of control, competence, and ownership, all of which are central motivational conditions in writing development from the perspective of self-determination theory [6]. Because the study did not collect process traces, this explanation should be read as a theoretically informed interpretation rather than direct observation of individual AI-use behavior. This interpretation is also consistent with research findings warning that generative AI may undermine deep learning by inducing metacognitive laziness, as well as with studies demonstrating that the educational outcomes of AI use depend on whether such use produces deskilling or more constructive forms of adaptive skill development [1,2].

By contrast, the augmentation-oriented condition preserved the learner's central role while using AI as a structured support tool. Through prompt design, idea expansion, selective adoption of feedback, and revision guidance, students were encouraged to remain active participants in composing and refining their texts. In this sense, AI functioned as a scaffold for higher-order writing processes rather than as a replacement for them. This interpretation is in line with recent L2 writing research showing that AI-supported writing is most beneficial when learners engage critically with feedback and revision rather than adopting AI-generated content passively [3-5].

5.2 Intrinsic Motivation as An Explanatory Mechanism

The mediation findings provide additional insight into why augmentation-oriented GenAI use led to better writing outcomes. Intrinsic motivation partially mediated the relationship between GenAI use mode and writing performance. As illustrated in Figure 1, augmentation-oriented use was associated with writing performance both directly

and through its positive association with intrinsic motivation. This pattern indicates that the benefits of augmentation-oriented AI use were not purely technical; they were also motivational.

This finding matters because writing improvement is inseparable from engagement. Students do not develop writing skills simply by producing a better product; they improve through repeated effort, revision, and strategic reflection. If augmentation-oriented AI use increases intrinsic interest, perceived agency, and writing-related efficacy, improved writing performance is likely to follow. However, because interaction processes were not directly recorded, the present study cannot determine exactly how individual students selected, revised, or rejected AI-generated suggestions. This interpretation is compatible with self-determination theory and with emerging studies suggesting that GenAI can enhance writing outcomes when it supports engagement, feedback uptake, and sustained revision rather than replacing learner effort [1,4,5].

At the same time, the mediation was partial rather than complete. This suggests that GenAI use also influenced writing performance through additional pathways, such as idea support, language enhancement, and revision efficiency. In other words, augmentation-oriented GenAI use appears to have operated through both psychological and process-based mechanisms. Because the model included only intrinsic motivation, these additional pathways remain theoretical possibilities rather than empirically tested mediators in the present study.

5.3 Theoretical Contributions

This study makes three main contributions to the literature on AI-supported language learning. First, it provides empirical evidence that different GenAI use modes are associated with substantially different motivational and performance outcomes. Second, it clarifies that intrinsic motivation is one important, but not exhaustive, mechanism linking AI-supported instructional design to writing achievement. Third, it strengthens a growing line of argument that educational AI should be understood as a pedagogically mediated resource rather than as a neutral productivity tool.

More broadly, the findings support a view of GenAI as beneficial only when instructional design preserves learner agency. Whether GenAI strengthens or weakens learning depends on how cognitive labor is distributed between learner and tool, and on whether students remain authors rather than mere operators. In this sense, the present results extend recent concerns about metacognitive laziness by suggesting that AI-related risks are not fixed properties of the technology itself but are shaped by use mode and pedagogy.

5.4 Practical Implications

The findings also carry clear practical implications for college English writing instruction. Most importantly, the issue is not whether teachers should allow GenAI in writing classrooms, but how GenAI should be pedagogically configured. If AI is introduced mainly as a shortcut for draft production, it may improve apparent efficiency while weakening short-term transfer and learner ownership. However, it is introduced as a

structured support tool for brainstorming, rhetorical comparison, feedback interpretation, and revision, it may enhance both motivation and writing performance.

For teachers, this means that GenAI integration should be accompanied by explicit task design and process guidance. Students need support in using AI critically, selectively, and reflectively. Instructors can operationalize augmentation-oriented use through prompt-design training, staged revision tasks, comparison of AI-generated alternatives, and reflection on which suggestions are accepted or rejected. Instructors should therefore frame GenAI as a collaborator in idea development and revision rather than as a substitute author, and they should ask students to document how AI suggestions are evaluated and revised. This recommendation is in line with recent research calling for more deliberate and pedagogically grounded uses of GenAI in L2 writing contexts [3-5].

5.5 Limitations and Future Directions

Several limitations should be acknowledged. First, the study used a quasi-experimental design with intact classes, so causal inference should be made cautiously despite the statistical controls and baseline equivalence checks. Second, the sample consisted of first-year non-English majors from a single institutional context, which limits generalizability. Third, knowledge transfer was measured once at the end of Week 6 through a cross-topic task; therefore, the findings indicate short-term transfer performance rather than long-term retention of learning outcomes. Fourth, the study focused on intrinsic motivation and did not directly examine other potentially relevant constructs such as self-efficacy, anxiety, feedback literacy, or metacognitive strategy use. Fifth, although the design distinguished substitution-oriented and augmentation-oriented use, it did not collect prompt logs, revision histories, screen recordings, or other process-tracing data, so the discussion of cognitive engagement should be understood as an interpretation of group-level outcomes rather than direct evidence of students' moment-by-moment interaction patterns. Finally, the mediation analysis used a planned binary contrast to test the theoretically central augmentation effect, which reduced information about subtler differences between the traditional instruction and substitution-oriented conditions.

Future research could address these limitations by using multi-site randomized designs, delayed transfer tests, richer process data, and longitudinal follow-up measures of writing development and retention. It would also be valuable to compare multiple forms of augmentation-oriented AI use, including AI-supported feedback, AI-supported peer review, and collaborative human–AI revision, while testing additional mediators such as self-efficacy, anxiety, feedback literacy, and metacognitive strategy use.

6 Conclusion

Within this quasi-experimental, single-institution study, augmentation-oriented GenAI use was associated with higher intrinsic motivation, writing performance, and short-term knowledge transfer than the other conditions, whereas substitution-oriented use

showed weaker short-term transfer than traditional instruction. Intrinsic motivation partially mediated the association between augmentation-oriented use and writing performance. Although limited to short-term outcomes, the findings suggest that GenAI-supported writing pedagogy should augment rather than replace learner writing.

References

1. Y. Fan, L. Tang, H. Le, K. Shen, S. Tan, Y. Zhao, and D. Gašević, "Beware of metacognitive laziness: effects of generative artificial intelligence on learning motivation, processes, and performance," British Journal of Educational Technology, vol. 56, no. 2, pp. 489–530, 2025. doi: 10.1111/bjet.13544.
2. B. Yang, Y. Sun, Z. Zeng, and Q. Li, "Deskilling, reskilling, or upskilling? Unpacking the pathways of student adaptation to generative artificial intelligence," International Journal of Information Management, vol. 87, p. 103002, 2026. doi: 10.1016/j.ijinfomgt.2025.103002.
3. S. Koltovskaia, "Student engagement with automated written corrective feedback provided by Grammarly: a multiple case study," Assessing Writing, vol. 44, p. 100450, 2020. doi: 10.1016/j.asw.2020.100450.
4. M. A. A. Mekheimer, "Generative AI-assisted feedback and EFL writing: a study on proficiency, revision frequency and writing quality," Discover Education, vol. 4, p. 170, 2025. doi: 10.1007/s44217-025-00602-7
5. S. Yeung, "University students' engagement with generative AI-supported automated writing evaluation (AWE) feedback," Journal of Second Language Writing, vol. 68, p. 101203, 2025. https://doi.org/10.1016/j.jslw.2025.101203.
6. E. L. Deci, and R. M. Ryan, "The 'what' and 'why' of goal pursuits: human needs and the self-determination of behavior," Psychological Inquiry, vol. 11, no. 4, pp. 227–268, 2000. doi: 10.1207/S15327965PLI1104_01.
7. C. Rinnert, H. Kobayashi, and A. Katayama, "Argumentation text construction by Japanese as a foreign language writers: a dynamic view of transfer," The Modern Language Journal, vol. 99, no. S1, pp. 213–245, 2015. doi: 10.1111/modl.12210.
8. Y. Huang, and L. J. Zhang, "Facilitating L2 writers' metacognitive strategy use in argumentative writing using a process-genre approach," Frontiers in Psychology, vol. 13, p. 1036831, 2022. doi:10.3389/fpsyg.2022.1036831.
9. A. F. Hayes, and N. J. Rockwood, "Regression-based statistical mediation and moderation analysis in clinical research: observations, recommendations, and implementation," Behaviour Research and Therapy, vol. 98, pp. 39–57, 2017. doi: 10.1016/j.brat.2016.11.001.
10. Y. T. Chang, A. T. Choe, D. Holden, and D. R. Isbell, "Making each point count: revising a local adaptation of the Jacobs et al.'s (1981) ESL composition profile rubric," Language Testing, vol. 41, no. 3, pp. 443–455, 2023. doi: 10.1177/02655322231217979.
11. H. L. Jacobs, S. A. Zinkgraf, D. R. Wormuth, V. F. Hartfiel, and J. B. Hughey, Testing ESL composition: a practical approach, Newbury House, Boston, 1981.
12. R. J. Vallerand, L. G. Pelletier, M. R. Blais, N. M. Brière, C. Senécal, and E. F. Vallières, "The academic motivation scale: a measure of intrinsic, extrinsic, and amotivation in education," Educational and Psychological Measurement, vol. 52, no. 4, pp. 1003–1017, 1992. doi: 10.1177/0013164492052004025.
13. L. T. Yu, M. C. Chen, C. W. Chiu, C. C. Hsu, and Y. P. Yuan, "Examining English ability-grouping practices by aligning CEFR levels with university-level general English courses in Taiwan," Sustainability, vol. 14, no. 8, p. 4629, 2022. doi: 10.3390/su14084629.