Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Effects of Blended AI-Assisted and Teacher-Led Instruction on EFL Students' Academic Writing Development and Writing Self-Efficacy at Salale University, Ethiopia

Yonas Tamiru Gebre

yonas_tamiru@slu.edu.et

Salale University

Research Article

Keywords: academic writing, artificial intelligence, blended instruction, EFL, writing self-efficacy

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11367514/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Effects of Blended AI-Assisted and Teacher-Led Instruction on EFL Students' Academic Writing Development and Writing Self-Efficacy at Salale University, Ethiopia

Yonas Tamiru Gebre

Department of English Language and Literature, College of Social Sciences and Humanities, Salale University, Fitche, Ethiopia

yonas_tamiru@slu.edu.et

https://orcid.org/0009-0006-2132-7928

Abstract

Generative artificial intelligence has entered Ethiopian higher education faster than pedagogical frameworks for its use. This study examined the effects of a blended model combining AI-assisted feedback with teacher-led instruction on EFL students' academic writing development and writing self-efficacy. A one-group pre-test–post-test quasi-experimental design with an embedded qualitative strand was used. Thirty-seven third-year students of the Department of English Language and Literature at Salale University, enrolled in the Intermediate Writing course, received ten weeks of blended instruction in which AI tools supplied immediate draft-level feedback and the teacher supplied genre, content and revision instruction. Teaching materials were modified from the existing course textbook rather than newly prepared. Writing was assessed with an adapted analytic rubric and self-efficacy with the Second Language Writer Self-Efficacy Scale. Paired-samples t-tests showed significant gains in overall writing, from $M = 51.38$ to $M = 64.22$, $t(36) = 10.00$, $p < .001$, $d = 1.64$, and in writing self-efficacy, from $M = 2.84$ to $M = 3.73$, $t(36) = 9.33$, $p < .001$, $d = 1.53$. Gains were largest for content and organization and smallest for mechanics. Writing and self-efficacy gains correlated moderately, $r = .46$, $p = .004$.

Keywords: academic writing, artificial intelligence, blended instruction, EFL, writing self-efficacy

1. Introduction

Academic writing is the gatekeeping competence of Ethiopian higher education (Yonas & Bifa, 2026). Third-year students of English Language and Literature are expected to produce argumentative essays, literature reviews and research reports in a language they encounter chiefly in the classroom, under conditions of large class size and limited individual feedback. The structural problem is well known: formative feedback on writing is the single most effective lever on writing development and simultaneously the most labor-intensive, and the labor does not scale (Hyland, 2019; Ferris, 2023).

Generative artificial intelligence has altered this arithmetic. Large language models can return draft-level commentary within seconds, repeatedly, at any hour, and without the social cost that some learners attach to exposing a weak draft to a human reader (Kohnke et al., 2023). Early empirical work is encouraging. Asadi et al. (2025) found that combining ChatGPT feedback with teacher feedback improved EFL learners' argumentative writing more than either source alone, with gains across coherence, organization and language. Alnemrat (2025) compared AI- and teacher-generated feedback and reported complementary profiles: AI feedback was faster, more consistent and stronger on surface features, while teacher feedback was more contextually attuned and more effective on content and rhetorical purpose.

That complementarity is the pedagogical argument for blending rather than substituting. AI feedback is immediate, inexhaustible and non-judgmental, but it is also context-blind; it cannot know the assignment's disciplinary expectations, the learner's history of error, or the local rhetorical conventions of the course. Teacher feedback is contextually rich but scarce. A blended model assigns each source the work it does best: AI handles high-frequency, low-inference feedback on language, cohesion and surface organization during drafting, while the teacher handles genre instruction, content development, rhetorical purpose, and the metacognitive work of teaching students how to evaluate and act upon the feedback they receive.

A second rationale concerns affect. Writing self-efficacy, a writer's belief in their capability to execute the actions required to produce a text, is a strong predictor of writing performance, persistence and strategy use (Bandura, 1997; Teng et al., 2018). Teng et al. (2018) distinguish three dimensions: linguistic self-efficacy (confidence in handling grammar, vocabulary and mechanics), performance self-efficacy (confidence in executing writing tasks) and self-

regulatory self-efficacy (confidence in planning, monitoring and revising). Bandura's framework predicts that AI feedback should raise self-efficacy through two of its four sources: mastery experience, because low-stakes iterative revision generates visible success, and reduced physiological and affective arousal, because a machine reader removes evaluative threat. Lai (2025) found exactly this pattern, with AI-assisted blended learning strengthening writing self-efficacy and resilience relative to conventional instruction.

There is, however, a plausible counter-prediction. If learners outsource cognition to the tool, mastery experience is illusory and self-efficacy may become inflated and brittle, or self-regulatory efficacy may decline even as performance efficacy rises (Teng, 2024). Distinguishing these outcomes requires measuring self-efficacy dimensionally, which this study does.

Ethiopian evidence is almost non-existent. Ethiopian universities operate under constraints — intermittent connectivity, limited device ownership, large classes — that make unmodified importation of findings from well-resourced contexts unsafe. This study provides the first classroom-based evidence from an Ethiopian university on a blended AI–teacher writing model.

1.1. Statement of the Problem

Students in the Department of English Language and Literature at Salale University enter the Intermediate Writing course with persistent difficulties in thesis construction, paragraph-level development, cohesion, source integration and grammatical accuracy (Yonas & Bifa, 2026). The structural cause is feedback scarcity: with a single instructor serving a full section, each student can realistically receive detailed written feedback on no more than two drafts per semester, and that feedback typically arrives after the point at which it could inform revision. Delayed feedback on a completed assignment functions as justification of a grade rather than as instruction.

Meanwhile, AI tools have arrived in students' hands ahead of any institutional framework. Observation suggests that use is already widespread but undisciplined, oscillating between wholesale text generation at one extreme and avoidance at the other, with little of the structured, revision-oriented use that the literature associates with learning gains. Left unaddressed, the default trajectory is substitution rather than support: students submit AI-generated text, instructors respond with prohibition, and neither writing development nor AI competence improves.

What is missing is evidence on a structured middle path. Specifically, it is not known whether a designed blend, in which AI feedback is positioned at the drafting stage under explicit protocols and teacher instruction retains authority over genre, content and revision decisions, produces measurable gains in Ethiopian students' writing quality and whether those gains extend to the affective dimension of writing self-efficacy. Nor is it known which components of writing respond to such a blend and which do not, a question of direct practical importance because it determines where scarce teacher attention should be concentrated.

This study addresses these gaps by implementing and evaluating a ten-week blended intervention within an authentic Ethiopian university course.

1.2. Research Questions

1. Does blended AI-assisted and teacher-led instruction significantly improve students' overall academic writing performance?
2. Which components of academic writing (content, organization, vocabulary, language use, and mechanics) show the greatest and least improvement?
3. Does the intervention significantly improve students' writing self-efficacy across its linguistic, performance and self-regulatory dimensions?
4. What is the relationship between gains in writing performance and gains in writing self-efficacy?
5. How do students experience the blended model?

1.3. Theoretical Framework

The intervention is grounded in the convergence of three positions. Sociocultural theory supplies the mediation rationale: AI functions as a meditational artifact that provides graduated assistance within the learner's zone of proximal development, with the teacher responsible for the responsive, contingent scaffolding that a non-contingent tool cannot supply (Lantolf et al., 2020). Process-genre writing pedagogy supplies the instructional architecture: multiple drafts, explicit genre modeling and staged feedback (Hyland, 2019). Social cognitive theory supplies the affective mechanism: iterative, low-threat success experiences strengthen self-efficacy, which in turn sustains effort and strategy deployment (Bandura, 1997; Teng et al., 2018).

The framework predicts differential gains. Components governed by explicit, rule-like criteria and amenable to immediate automated commentary (organization, cohesion, lexical choice) should respond quickly. Components requiring deep contextual judgment (content quality, argumentative force) should respond to the teacher-led component. Components involving deeply entrenched inter-language features (complex grammatical accuracy) should respond most slowly, because noticing a correction is not equivalent to restructuring an inter-language rule.

2. Materials and Methods

2.1 Design

A one-group pre-test–post-test quasi-experimental design with an embedded qualitative strand was employed. Only an experimental group was used. The choice was dictated by institutional reality: the Department operates a single third-year section, and withholding an instructional innovation from a randomly selected half of a required course would have been both administratively impossible and ethically questionable. The design limitation this entails, the absence of a counterfactual, is acknowledged in Section 10 and partially offset by the use of a delayed baseline, standardized rubric, blind double-marking and effect-size reporting.

2.2 Participants

Participants were all 37 third-year students of the Department of English Language and Literature at Salale University enrolled in the Intermediate Writing course in the 2026 academic year (comprehensive sampling; N = 37). Characteristics are summarized in Table 1.

Table 1: *Participant characteristics*

Variable	Category	Frequency	Percentage (%)
Sex	Male	21	56.8
	Female	16	43.2
Age	19–21 years	24	64.9
	22–24 years	11	29.7
	Above 24 years	2	5.4
Prior AI tool use for writing	Regular	9	24.3
	Occasional	19	51.4
	None	9	24.3
Personal smartphone access	Yes	34	91.9
	No	3	8.1
Personal laptop access	Yes	1	2.7
	No	36	97.3

The demographic and technology-access data show that the 37 participants were mostly male (56.8%) and between the ages of 19 and 21 (64.9%). In terms of past use of AI tools for writing, 51.4% reported occasional use, 24.3% used AI tools regularly, and an equal percentage had no prior experience. The majority of participants had personal smartphone access (91.9%), but only 2.7% had personal laptop access, which may limit their access to and use of AI-assisted writing technologies, especially for tasks requiring extensive academic writing.

2.3 Instructional Materials

Instructional materials were modified, not newly prepared. The existing Intermediate Writing course textbook prescribed by the Department served as the base text. Modification comprised four operations: (a) replacing decontextualized sample texts with topics drawn from students' academic and local context (Salale dairy value chain, campus water supply, Oromo oral literature, secondary school dropout in North Shewa); (b) restructuring single-submission assignments into three-draft cycles; © inserting AI-interaction protocol boxes specifying permitted prompts and required documentation at each drafting stage; and (d) adding revision logs and reflection prompts after each unit. No unit was removed and the course's official learning outcomes were unchanged. The modified materials were reviewed and approved by two senior staff of the Department prior to use.

2.4 The Intervention Procedure

The ten-week intervention ran four contact hours per week within the regular course timetable. Each two-week cycle addressed one genre (summary–response, cause–effect essay, argumentative essay, literature review section, research-report introduction) and followed a fixed five-stage sequence.

In Stage 1 (teacher-led, 90 minutes), the instructor presented genre models, rhetorical moves and evaluative criteria, and conducted joint deconstruction of a sample text. In Stage 2 (independent), students produced Draft 1 without AI assistance; this draft was retained as evidence of unassisted capability. In Stage 3 (AI-assisted, supervised lab session), students submitted Draft 1 to an AI assistant using prescribed feedback-seeking prompts and were explicitly prohibited from requesting generated text. Prompts followed a fixed frame: *"Do not rewrite my text. Identify the three weakest aspects of [coherence / thesis clarity / grammatical accuracy] in the paragraph below and explain why each is weak."* Students recorded all prompts and outputs in an AI

interaction log. In Stage 4 (teacher-led, 90 minutes), the instructor conducted whole-class and small-group conferencing focused on content quality, argument strength and the *evaluation* of AI suggestions, with students required to justify which AI recommendations they accepted or rejected. In Stage 5, students produced Draft 2, completed a revision log, and submitted the full portfolio.

To accommodate connectivity constraints, AI sessions were conducted in the university ICT laboratory during scheduled hours, and offline fallback worksheets containing pre-generated anonymous AI feedback samples were available when connectivity failed (used in weeks 3 and 7).

2.5 Instruments

2.5.1. Academic writing test

Parallel pre- and post-test prompts required a 450–500-word argumentative essay under 90-minute timed conditions without AI access were used. Prompt equivalence was established by expert judgment and by comparable readability and topic familiarity ratings. Scripts were scored on a 100-point analytic rubric adapted from Jacobs et al. (1981): content (30), organization (20), vocabulary (20), language use (25), and mechanics (5). Two trained raters scored all scripts blind to testing occasion; inter-rater reliability was high (ICC = .91 pre-test, .89 post-test), and discrepancies exceeding five points were resolved by discussion.

2.5.2. Second Language Writer Self-Efficacy Scale (SLWSES)

The 18-item scale of Teng et al. (2018) was used with minor contextual adaptation of three items, measuring linguistic self-efficacy (6 items), performance self-efficacy (6 items) and self-regulatory self-efficacy (6 items) on a five-point scale. Internal consistency in this sample was good (pre-test $\alpha = .88$; post-test $\alpha = .90$; subscale $\alpha = .81$–$.87$).

2.5.3. Semi-structured Interview

Semi-structured interviews were conducted with twelve purposively selected students (four each from the highest, middle and lowest gain terciles). Students' revision logs and AI interaction logs were collected as documentary data.

2.6 Data Analysis

Data were analyzed in SPSS version 27. Normality of difference scores was confirmed by Shapiro–Wilk tests (all p > .05) and inspection of Q–Q plots. Paired-samples t-tests compared pre- and post-test scores; Cohen's d for repeated measures was computed as the mean difference divided by the standard deviation of the differences, interpreted as 0.2 small, 0.5 medium, 0.8 large (Cohen, 1988). Pearson correlation examined the association between writing and self-efficacy gains. Because five rubric sub-scales were tested, a Bonferroni-adjusted alpha of .01 was applied to sub-scale comparisons. Qualitative data were analyzed through reflexive thematic analysis (Braun & Clarke, 2022).

2.7 Ethical Considerations

Ethical clearance was obtained from the Research Ethics Review Committee, College of Social Sciences and Humanities, Salale University (Ref. SLU/CSSH/ERC/052/2024). Students gave written informed consent, were told that participation was unrelated to course grades, and were free to withdraw their data at any stage. Students who preferred not to use AI tools were offered an equivalent peer-feedback pathway; none exercised this option. All data were anonymized.

3. Results and Discussion

3.1 Overall Academic Writing Performance

Table 2: *Paired-samples t-test results for overall academic writing performance*

Measure	Pre-test M (SD)	Post-test M (SD)	Mean diff.	SD diff.	95% CI	t(36)	*p*	d
Overall writing (100)	51.38 (8.42)	64.22 (7.16)	12.84	7.81	[10.24, 15.44]	10.00	<.001	1.64

Overall writing performance improved by 12.84 points, a statistically significant and very large effect (t (36) = 10.00, $p < .001$, d = 1.64). The 95% confidence interval excludes zero by a wide margin, and the narrowing of the standard deviation from 8.42 to 7.16 indicates that the intervention also reduced dispersion: weaker writers gained proportionally more, compressing the performance range. $H_0 1$ is rejected.

3.2 Component-Level Writing Gains

Table 3: *Paired-samples t-test results for writing sub-scales*

Sub-scale (max)	Pre-test M (SD)	Post-test M (SD)	Mean diff.	SD diff.	t(36)	p	d	Gain as % of max

Content (30)	14.86 (2.97)	18.92 (2.41)	4.06	2.96	8.34	<.001	1.37	13.5
Organization (20)	9.24 (1.88)	12.38 (1.56)	3.14	2.21	8.64	<.001	1.42	15.7
Vocabulary (20)	9.97 (1.74)	12.84 (1.52)	2.87	2.08	8.39	<.001	1.38	14.4
Language use (25)	13.46 (2.61)	15.86 (2.14)	2.40	2.54	5.75	<.001	0.94	9.6
Mechanics (5)	3.85 (0.74)	4.22 (0.61)	0.37	0.83	2.71	.010	0.45	7.4

Note. Bonferroni-adjusted $\alpha = .01$ for sub-scale comparisons.

All five components improved significantly. The gradient is theoretically coherent. Organization showed the largest proportional gain (15.7% of maximum, d = 1.42), followed by vocabulary (14.4%, d = 1.38) and content (13.5%, d = 1.37). Language use gained appreciably less (9.6%, d = 0.94) and mechanics least (7.4%, d = 0.45), the latter being the only effect in the small-to-moderate range and the only one approaching the adjusted significance threshold.

Two readings are warranted. First, the modest mechanics gain partly reflects a ceiling: the pre-test mean of 3.85 out of 5 left little headroom. Second, and more substantively, the comparatively small language-use gain indicates that ten weeks of feedback — however immediate and abundant — does not restructure entrenched inter-language grammar. Students corrected errors locally within drafts without converting those corrections into stable productive control, as the unassisted timed post-test reveals.

3.3 Writing Self-Efficacy

Table 4: *Paired-samples t-test results for writing self-efficacy*

Dimension	Pre-test M (SD)	Post-test M (SD)	Mean diff.	SD diff.	t(36)	p	d
Linguistic self-efficacy	2.84 (0.61)	3.78 (0.54)	0.94	0.69	8.29	<.001	1.36
Performance self-efficacy	2.96 (0.58)	3.92 (0.49)	0.96	0.66	8.85	<.001	1.45
Self-regulatory self-efficacy	2.71 (0.66)	3.49 (0.57)	0.78	0.74	6.41	<.001	1.05
Overall self-efficacy	**2.84 (0.55)**	**3.73 (0.47)**	**0.89**	**0.58**	**9.33**	**<.001**	**1.53**

Note. Scale range 1–5.

Writing self-efficacy rose significantly on all three dimensions and overall ($t(36) = 9.33$, $p < .001$, $d = 1.53$). $H_0 2$ is rejected. Pre-test scores clustered below the scale midpoint of 3.00, confirming that these students entered the course with low confidence; post-test scores moved into the moderately high band.

The dimensional pattern is informative. Performance self-efficacy gained most ($d = 1.45$) and self-regulatory self-efficacy least ($d = 1.05$). Students became markedly more confident in executing writing tasks and handling language and less markedly more confident in planning, monitoring and independently regulating their own writing process. This asymmetry is consistent with the theoretical concern that external feedback availability may substitute for, rather than build, self-regulatory capacity. The gain in self-regulation remains large by conventional standards, but its position at the bottom of the ranking is a signal worth heeding.

3.4 Relationship between Writing and Self-Efficacy Gains

Pearson correlation between writing gain scores and overall self-efficacy gain scores was positive, moderate and significant, $r(35) = .46$, $p = .004$, indicating that approximately 21% of the variance in self-efficacy gain was shared with writing gain. $H_0 3$ is rejected. Disaggregated, writing gain correlated most strongly with performance self-efficacy gain ($r = .49$, $p = .002$), then linguistic ($r = .41$, $p = .012$) and self-regulatory ($r = .33$, $p = .046$). The moderate rather than strong magnitude indicates that performance and confidence improved along related but partially independent trajectories.

3.5 Qualitative Findings

Four themes emerged. **Theme 1: Immediacy as the decisive affordance.** Students consistently identified feedback speed rather than feedback quality as transformative. S7 explained: “Before, I wrote and waited two weeks. By then I had forgotten what I was thinking. Now I get a comment in one minute and I fix it while the idea is still in my head.” This map onto the organization and cohesion gains, which depend on acting on feedback while the text’s structure remains mentally available.

Theme 2: Reduced evaluative anxiety. S14 reported: “I can show my worst paragraph to the machine. I cannot show my worst paragraph to the teacher or to my classmates.” Several

students described submitting drafts they would previously have abandoned. This supports the Bandurian prediction that lowered affective arousal contributes to efficacy growth.

Theme 3: The teacher as arbiter. Students did not treat AI feedback as authoritative. S3 observed: “Sometimes it changes my meaning to make nice English. The teacher told us to ask: did I want to say this? Now I refuse some corrections.” Revision logs confirmed this: across the cohort, students documented rejecting 18.6% of AI suggestions on average, most frequently on grounds of meaning distortion or inappropriate formality. The teacher-led evaluation stage appears to have been the active ingredient preventing passive acceptance.

Theme 4: Dependence anxiety. A persistent minority voiced concern about their unassisted capability. S21 stated: “When I write in the exam alone, I feel my hand is empty.” Six of twelve interviewees expressed some version of this. This triangulates with the comparatively weaker self-regulatory efficacy gain and identifies the principal risk of the model.

Infrastructural strain was reported by all interviewees as a background condition: scheduled laboratory access, intermittent connectivity and limited device ownership (only 32.4% owned laptops) meant the model depended on institutional provision rather than individual resources.

3.6. Discussion

The intervention produced large, statistically significant gains in both academic writing performance (d = 1.64) and writing self-efficacy (d = 1.53) over ten weeks. The direction and magnitude align with Asadi et al. (2025), who found combined ChatGPT-and-teacher feedback superior for EFL argumentative writing, and with Lai (2025), who reported enhanced writing self-efficacy under AI-assisted blended learning. The present study extends this evidence to an Ethiopian university operating under substantially tighter infrastructural constraints, demonstrating that the model is viable where access is institutionally scheduled rather than individually continuous.

The component gradient is the study’s most practically useful finding. Gains were largest for organization, vocabulary and content, and smallest for language use and mechanics. This contradicts a common assumption that AI feedback principally improves grammatical accuracy. Two mechanisms explain the discrepancy. First, grammatical accuracy in timed, unassisted conditions depends on automatized procedural knowledge, which develops through extensive

output practice over long periods rather than through the receipt of corrective information (Ellis & Shintani, 2014; Ferris, 2023). Noticing a correction is a necessary but insufficient condition for restructuring. Second, organization and content respond to explicit, teachable rhetorical schemata that can be modeled, applied and checked within a single drafting cycle. The implication for practice is pointed: teacher attention in such a blend should be concentrated on content and argumentation, where AI is weakest and where returns are high, while sustained accuracy development requires a longer instructional horizon than a single semester.

The self-efficacy findings require careful reading. The overall gain is unambiguous, but the dimensional ranking — performance highest, self-regulation lowest — mirrors the concern raised by Teng (2024) that readily available AI assistance may bolster confidence in task execution without building the metacognitive independence that sustains writing development beyond the supported environment. The qualitative "dependence anxiety" theme converges on the same point: a substantial minority of students doubted their unassisted capability even as their measured unassisted performance improved. Their doubt is empirically unfounded — the post-test was AI-free — but pedagogically significant, since efficacy beliefs drive persistence. This suggests that blended designs should include explicit unassisted-performance milestones that give students evidence of their own independent growth.

The moderate correlation between writing and self-efficacy gains ($r = .46$) is consistent with the reciprocal rather than deterministic relationship posited by social cognitive theory (Bandura, 1997). Confidence and competence moved together, but each had substantial independent variance, reinforcing the case for measuring both.

The documented rejection of 18.6% of AI suggestions is, arguably, the study's most encouraging result, because it indicates that students developed evaluative stance rather than compliance. Critically, the structural features most plausibly responsible for this were the prohibition on text generation, the requirement to log prompts, and the teacher-led stage in which students had to justify acceptance or rejection decisions. In the absence of these constraints, there is no reason to expect the same outcome, and the literature on uncritical AI reliance suggests the opposite would occur (Chan & Hu, 2023; Teng, 2024). The blend's benefits should therefore be attributed to the protocol, not to the tool.

4. Conclusions

Ten weeks of blended AI-assisted and teacher-led instruction produced large, statistically significant improvements in third-year Ethiopian EFL students' academic writing performance (12.84 points; d = 1.64) and writing self-efficacy (0.89 scale points; d = 1.53), with gains in the two domains moderately correlated (r = .46). Improvement was unevenly distributed across writing components: organization, vocabulary and content improved substantially, whereas language use and mechanics improved considerably less. The blend is therefore best understood as a rhetorical and organizational intervention supported by language feedback, rather than as an accuracy remedy.

Self-efficacy gains were weakest in the self-regulatory dimension, and a minority of students reported anxiety about unassisted performance. The model's principal risk is the cultivation of supported rather than autonomous competence. Finally, the outcomes are attributable to a structured protocol, no text generation, mandatory prompt logging, and teacher-led evaluation of AI suggestions, rather than to AI access as such. Unstructured availability of the same tools should not be expected to produce comparable results.

5. Recommendations

Departments of English should formally integrate structured AI-assisted feedback into writing courses using an explicit protocol that prohibits text generation, requires documentation of all AI interactions, and makes the evaluation of AI output a graded component of the course.

Instructors should reallocate their scarce feedback capacity toward content development, argumentation and rhetorical purpose, where AI performs weakest and where the present data show strong responsiveness, while using AI for high-frequency surface-level commentary during drafting.

Because grammatical accuracy responded least, departments should treat accuracy development as a multi-semester objective supported by sustained output practice and targeted form-focused instruction, rather than expecting feedback technology to resolve it within one course.

To protect self-regulatory development, blended courses should schedule periodic unassisted writing milestones with transparent score reporting, so that students obtain direct evidence of their independent capability.

Salale University should guarantee scheduled ICT laboratory access and basic connectivity for writing courses, since the model demonstrated here depends on institutional provision given that only 32.4% of students owned laptops.

Future research should employ a control or comparison group, include a delayed post-test to test retention, and extend the intervention beyond one semester to determine whether accuracy gains emerge over a longer horizon.

6. Limitations

The absence of a control group is the principal limitation: observed gains cannot be unambiguously attributed to the intervention rather than to maturation, course exposure or testing effects, although the large effect sizes exceed what a ten-week maturation effect would plausibly produce. The single-institution sample of 37 students limits generalizability. No delayed post-test was administered, so retention is unknown. Self-efficacy was measured by self-report and may be subject to demand characteristics. Finally, the instructor also served as researcher, introducing possible experimenter expectancy, partially mitigated by blind double-marking of all scripts.

CRediT Author Statement

Yonas Tamiru Gebre: Conceptualization, Methodology, Investigation, Resources, Formal analysis, Data Curation, Writing – original draft, Writing – review and editing, Supervision, Project administration.

Declaration of Competing Interests

The author declares no competing financial or personal interests. No funding was received from any AI technology provider, and no commercial product was endorsed.

Ethical Approval

Approved by the Research Ethics Review Committee, College of Social Sciences and Humanities, Salale University (Ref. SLU/CSSH/ERC/052/2024), in accordance with the 1964 Declaration of Helsinki and its later amendments. Written informed consent was obtained from all participants.

Data Availability Statement

Anonymized datasets are available from the corresponding author upon reasonable request.

Acknowledgements

The author thanks the third-year students of the Department of English Language and Literature for their participation, and departmental colleagues who served as second raters and materials reviewers. Besides, the author extends his deepest thanks to his wife, Yemarshet Bekele Techan, for her cooperation in the work of this study, and his son, Leul Yonas Tamiru, for his affection.

References

Alnemrat, A. (2025). AI vs. teacher feedback on EFL argumentative writing. *Frontiers in Education, 10*, 1614673. https://doi.org/10.3389/feduc.2025.1614673

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity, 56*, 101766. https://doi.org/10.1016/j.tsc.2025.101766

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman. https://psycnet.apa.org/record/1997-08589-000

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. Sage. https://doi.org/10.1177/1035719X211058251

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education, 20*, 43. https://doi.org/10.1186/s41239-023-00411-8

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum. https://www.scirp.org/reference/ReferencesPapers?ReferenceID=2041144

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). Sage. https://www.scirp.org/reference/referencespapers?referenceid=3784840

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. Routledge. https://doi.org/10.4324/9780203796580

Ferris, D. R. (2023). *Treatment of error in second language student writing* (3rd ed.). University of Michigan Press. https://doi.org/10.3998/mpub.14491572

Field, A. (2024). *Discovering statistics using IBM SPSS Statistics* (6th ed.). Sage. https://www.scirp.org/reference/referencespapers?referenceid=3925246

Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press. https://catdir.loc.gov/catdir/samples/cam041/2003041957.pdf

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Newbury House. https://www.scirp.org/reference/referencespapers?referenceid=2835595

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Lai, Z. C. C. (2025). The impact of AI-assisted blended learning on writing efficacy and resilience. *Journal of Language Teaching and Research*, advance online publication. https://doi.org/10.1177/07356331261483943

Lantolf, J. P., Poehner, M. E., & Swain, M. (Eds.). (2020). *The Routledge handbook of sociocultural theory and second language development*. Routledge. https://www.scirp.org/reference/referencespapers?referenceid=3860229

Teng, L. S., Sun, P. P., & Xu, L. (2018). Conceptualizing writing self-efficacy in English as a foreign language contexts: Scale validation through structural equation modeling. *TESOL Quarterly, 52*(4), 911–942. https://doi.org/10.1002/tesq.432

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence, 7*, 100270. https://doi.org/10.1016/j.caeai.2024.100270

Tomlinson, B. (2023). *Developing materials for language teaching* (3rd ed.). Bloomsbury Academic.https://books.google.com.et/books?id=qNlzFdoD380C&printsec=frontcover#v=onepage&q&f=false

Yonas Tamiru Gebre, & Bifa Telila Beji. (2026). Computer-Assisted Process-Genre Approach in EFL Writing: University Students' Practices and Challenges. *Science, Technology and Arts Research Journal*, *15*(1), 56–66. https://doi.org/10.20372/star.V15.i1.05