

Revisiting EFL Writing Instruction in the Digital Era A Proposed Quasi-Experimental Protocol for Integrating Claude, ChatGPT, and Qwen in an Algerian University Writing Classroom

Dr. Moulkheir GADA

Yahia Farès University of Médéa (Algérie), Laboratory of language and text didactics (LDLT)

Email: gada.moulkheir@univ-medea.dz ;  <https://orcid.org/0009-0001-2397-6335>

Received: 12/01/2026 ; Accepted : 03/05/2026 ; Published : 18/08/2026

Abstract:

The rapid diffusion of generative artificial intelligence (AI) tools, combined with Algeria's ongoing educational reforms, has renewed interest in how EFL writing is taught, practiced, and evaluated. This paper presents a research protocol investigating the pedagogical potential of three freely accessible generative AI tools (Claude, ChatGPT, and Qwen) in supporting EFL writing instruction at university level. The protocol proposes a quasi-experimental design in a control and in AI-mediated writing tasks assessed through a dedicated rubric and test. The paper formulates testable hypotheses on writing performance, learner autonomy, and retention, while foregrounding the ethical and academic-integrity safeguards required

Keywords: generative AI, EFL writing, academic integrity, digital pedagogy, university level

Introduction:

Algeria's higher education system is undergoing a significant linguistic and pedagogical transformation. Since the generalization of English teaching at the primary level and the adoption of English as a Medium of Instruction (EMI) in several scientific disciplines, English has progressively gained ground as a language of academic communication and scholarly production. This transition coincides with the rapid, largely uncontrolled diffusion of generative AI tools among students, who increasingly rely on such tools for writing assistance, regardless of their reliability or the pedagogical framing of their use.

Within this context, EFL writing instructors face a dual challenge: harnessing the genuine pedagogical potential of AI tools for personalized feedback and creative scaffolding, while safeguarding academic integrity, learner autonomy, and authentic language development. This paper responds directly to that challenge by proposing a structured, replicable research protocol for investigating the classroom integration of three generative AI tools — Claude, ChatGPT, and Qwen — in EFL writing instruction, using a controlled quasi-experimental design.

The choice to examine three tools rather than a single system is deliberate. Much of the existing literature on AI-assisted EFL writing has focused almost exclusively on ChatGPT, leaving open the question of whether observed benefits and risks generalise across different generative AI systems or are instead artefacts of a single tool's particular interface, output style, or safety configuration. By including Claude and Qwen alongside ChatGPT, this protocol is designed to test whether the pedagogical logic of guided, rubric-anchored AI use — rather than the specific properties of any one system — is what actually drives learning outcomes. This design choice also reflects the practical

reality of Algerian classrooms, where student and teacher access to any single proprietary tool cannot be assumed to be stable or universal, making a multi-tool, free-tier-based design more replicable across institutions with varying levels of digital infrastructure.

1. Literature Review

1.1 The changing status of English and digital transformation in Algeria

The evolution of language policy in Algerian higher education has been well documented. Idri and Bouguebs (2021) examine how French has historically maintained a hegemonic position in multilingual Algerian academic contexts, notably within English departments, even as English has begun to compete for space as a vehicular language of science and research. Their ethnographic study of the Department of English at Bejaia University shows that French continues to dominate professional and administrative exchanges even among EFL specialists, illustrating how deeply entrenched linguistic habits can outlast official policy shifts. Similarly, Otmani's study on teacher training at the Ecole Normale Supérieure Assia Djebbar, Constantine, found that an existing English course failed to meet teachers' disciplinary and research-oriented needs, and called for a radical revision of English-course content to make it professionally and academically beneficial (Otmani, n.d.). These findings underline a recurring theme: English instruction in Algeria has long struggled to align with learners' real communicative and academic needs — a gap that thoughtfully integrated AI tools might help to narrow, provided their use is pedagogically framed rather than left to unsupervised student initiative.

At the University of Médéa, Kadik, Yaagoub, Gada, and Bouassal similarly document how instruction in a foreign language for specific purposes frequently fails to match the real disciplinary needs of Algerian science students, arguing for objectives-based, needs-driven language course design. Although their study concerns French rather than English, its methodological logic — anchoring language instruction in the specific academic and professional needs of Algerian students — transfers directly to the EFL writing context addressed here, and to the broader question of how any additional pedagogical resource, including AI, should be integrated in response to demonstrated learner needs rather than as an unreflective technological add-on.

The choice of Claude and Qwen, alongside ChatGPT, is further supported by recent institutional evidence from within the same journal's readership. In a survey of 21 teacher-researchers at the Department of Letters and English Language, University of Constantine 1, Hamani (2025) found that teachers already reported using a wide range of generative AI tools in their own academic work, with ChatGPT dominant but Claude and Qwen both explicitly mentioned among the tools in active use. The same study found that 95.2% of respondents considered AI a valuable tool for academic work, while 81% simultaneously rejected the idea that AI could replace human researchers, and a large majority called for clear institutional guidelines rather than either prohibition or unregulated adoption. This pattern — cautious, guideline-seeking acceptance rather than either rejection or uncritical enthusiasm — mirrors precisely the posture this protocol adopts toward AI-mediated student writing, and confirms that the three tools selected are neither exotic nor unfamiliar within the Algerian ENS network, but already part of the institutional landscape this study proposes to study systematically.

1.2 Generative AI in EFL writing: opportunities

A growing body of research documents the pedagogical value learners attribute to generative AI tools in academic writing. Werdiningsih, Marzuki, and Rusdin (2024) report that EFL students valued ChatGPT for helping them overcome uncertainties, clarify vocabulary, and generate content

suggestions, thereby allowing them to focus more of their effort on the creative and argumentative dimensions of their essays, while also cautioning that authenticity requires deliberate human judgment rather than passive acceptance of AI output. In a mixed-methods intervention study, Mahapatra (2024) found a significant positive impact of ChatGPT-based feedback on ESL students' academic writing skills, alongside overwhelmingly positive student perceptions of the tool as a dialogic feedback partner. Teng (2024) similarly reports that EFL learners perceived ChatGPT as a supportive "companion" rather than a threat when feedback was used to prompt revision rather than to replace it. Fetsi's classroom-based account of AI-integrated writing instruction at the Ecole Normale Supérieure Assia Djebbar likewise reports that structured guidelines on AI use — such as prohibiting learners from asking AI to generate entire essays — allowed a productive, creativity-oriented fusion of AI assistance and student composition, rather than outright substitution of student effort (Fetsi, 2025).

1.3 AI as scaffolded feedback and collaborative writing partner

A related strand of research has examined generative AI not simply as a source of content but as a structured scaffold for argumentative writing. Su, Lin, and Lai (2023) show that collaborating with ChatGPT during argumentative essay composition helped learners develop stronger claim-evidence-warrant structures when the collaboration was guided by explicit task prompts rather than left open-ended. Guo, Wang, and Chu (2022) similarly demonstrate that chatbot-based scaffolding can support EFL students in constructing more coherent argumentative writing, provided the scaffolding targets specific sub-skills rather than offering generic assistance. At a more evaluative level, Steiss et al. (2024) compared the quality of human and ChatGPT-generated feedback on student writing and found that AI-generated feedback could match human feedback on several dimensions of accuracy and specificity, while noting that human feedback retained an advantage in addressing higher-order rhetorical and contextual concerns. Taken together, these studies motivate the present protocol's decision to structure AI use around discrete diagnostic quizzes and bounded revision tasks rather than unrestricted, open-ended AI assistance.

1.4 Risks: academic integrity, overreliance, and digital literacy

Alongside these benefits, the literature consistently flags serious risks. Bin-Nashwan, Sadallah, and Bouteraa (2023) argue that the use of ChatGPT in academic settings places academic integrity itself in the balance, calling for institutional vigilance and for stakeholders — academic institutions, publishers, and AI developers alike — to specify clear ethical-use guidelines. Their findings, drawn from a large cross-institutional sample, further show that perceived academic-integrity risk directly and negatively affects adoption behaviour, suggesting that integrity concerns are not a marginal afterthought but a central variable shaping how AI tools are actually used in practice. At a broader policy level, UNESCO's guidance for generative AI in education stresses the need for human-centred, pedagogically grounded frameworks for AI adoption in educational institutions rather than unregulated diffusion, and recommends age-appropriate safeguards and institutional validation of AI tools before their classroom deployment (Miao & Holmes, 2023). These concerns directly justify the controlled, rubric-based, and ethically framed design proposed in Section 3 below.

1.5 Synthesis and research gap

Read together, this body of work suggests a consistent pattern: generative AI tools can meaningfully support EFL writing development — through idea generation, feedback, and scaffolded revision — but only when their use is bounded by explicit pedagogical guidelines and paired with

mechanisms that preserve learner agency and academic integrity. What remains comparatively underexplored, particularly in the Algerian context, is the durability of these gains beyond the immediate intervention period, and the specific behavioural patterns learners display when asked to critically evaluate rather than simply accept AI-generated suggestions. It is precisely this gap — the retention and critical-engagement dimension of AI-mediated EFL writing instruction — that the present protocol is designed to address, using three widely accessible tools (Claude, ChatGPT, and Qwen) rather than a single proprietary system, so as to maximise the ecological validity and replicability of the design across resource-constrained institutional settings.

A second, related gap concerns the unit of analysis typically used in prior studies. Much of the research reviewed above reports either aggregate writing-quality scores or general perception surveys, but comparatively few studies combine a rubric dimension specifically dedicated to autonomous revision with a delayed retention measure in the same design. By explicitly pairing these two elements — the autonomous-revision criterion in Table n° 02 and the one-month retention test described in Section 3.4 — the present protocol aims to generate evidence not only on whether AI-mediated instruction improves immediate writing outcomes, but on whether any such improvement reflects genuine skill consolidation or a more superficial, session-bound effect that dissipates once AI support is withdrawn.

2. Research Questions and Hypotheses

This protocol is guided by three research questions: RQ1 asks whether structured, guideline-based use of generative AI tools improves students' writing performance compared to conventional instruction; RQ2 asks whether AI-mediated writing tasks structured around AI-generated quizzes affect the one-month retention of targeted linguistic and rhetorical knowledge; RQ3 asks what patterns of autonomous revision and critical engagement with AI-generated suggestions can be observed among the experimental group.

Three hypotheses are proposed. H1: the experimental group will show significantly higher post-test writing scores than the control group. H2: the experimental group will demonstrate higher retention scores at the one-month follow-up, provided AI use is scaffolded rather than unrestricted. H3: learners in the experimental group will show measurable, though uneven, gains in autonomous revision behaviour.

3. Proposed Methodology

3.1 Research design

A quasi-experimental, pre-test/post-test control-group design is proposed. Random assignment at the individual level is not assumed to be feasible within standard institutional timetabling constraints; group allocation would instead follow existing class sections, a limitation discussed in Section 5.

3.2 Participants

The target population consists of 67 university-level EFL students, distributed into two intact groups: a control group receiving conventional writing instruction without AI mediation, and an experimental group receiving AI-mediated instruction as described below. Group sizes, proficiency levels, and disciplinary background should be documented and reported descriptively once the protocol is implemented, together with informed consent procedures approved by the relevant institutional authority.

3.3 Instructional tools

Three freely accessible generative AI tools are proposed for the experimental group: Claude (free tier), ChatGPT (free tier), and Qwen. Students would be introduced to each tool through a guided orientation session clarifying permissible and impermissible uses, consistent with the guideline-based approach documented in comparable classroom interventions (Fetsi, 2025).

3.4 Procedure

Over a four-week instructional period, both groups would complete parallel sequences of writing tasks. For the experimental group, each writing cycle would begin with a short diagnostic quiz generated by one of the three AI tools, targeting a specific writing sub-skill. Students would then complete a writing task addressing the identified sub-skill, followed by a self-revision phase in which they would justify or reject specific AI-suggested changes. The control group would complete the same writing prompts and quizzes in paper-based or non-AI digital form, with feedback provided solely by the instructor.

At the end of the four-week period, both groups would sit a common post-test writing task, assessed via the rubric presented in Section 3.5. One month later, a retention test — targeting the same sub-skills but using new prompts — would be administered to both groups to test knowledge durability (RQ2).

3.5 Indicative timeline

Table 1 below summarises the indicative four-week instructional sequence, followed by the one-month retention interval, that structures the procedure described above. The timeline is deliberately compact so that it can be embedded within a single semester's teaching calendar without displacing other required coursework.

Table n° 01: Indicative four-week instructional timeline and retention interval

Week	Experimental group	Control group
1	Orientation session on Claude, ChatGPT, and Qwen; AI-generated diagnostic quiz on thesis-statement construction; guided writing task; self-revision phase	Diagnostic quiz (paper-based); writing task; instructor feedback
2	AI-generated quiz on paragraph cohesion; guided writing task; self-revision phase	Paper-based quiz; writing task; instructor feedback
3	AI-generated quiz on argument substantiation; guided writing task; self-revision phase	Paper-based quiz; writing task; instructor feedback
4	Consolidation session; common post-test writing task (both groups)	Consolidation session; common post-test writing task (both groups)
+4 weeks	One-month retention test (new prompts, same sub-skills; both groups)	One-month retention test (new prompts, same sub-skills; both groups)

Source: author's own elaboration, 2026

3.6 Evaluation rubric

The following analytic rubric is proposed for scoring both the post-test and the retention test. It combines standard EFL writing-assessment criteria with two dimensions specific to AI-mediated writing: autonomous revision and academic integrity, in line with recently proposed AI-inclusive assessment frameworks for argumentative writing (Khampusaen, 2025) and with evidence that AI-

generated feedback, while often accurate, still requires human judgment on higher-order rhetorical concerns (Steiss et al., 2024).

Table n° 02: Evaluation rubric for AI-mediated EFL writing tasks

Criterion	Weight	Descriptor (highest band)
Task achievement / content relevance	20%	Fully addresses the prompt; ideas are relevant, developed, and well supported
Organization & coherence	20%	Clear logical progression; effective use of cohesive devices and paragraphing
Lexical range & accuracy	20%	Wide, appropriate vocabulary used accurately with minimal L1 interference
Grammatical range & accuracy	20%	Varied structures used with a high degree of accuracy; few systematic errors
Autonomous revision (post-AI use)	10%	Evidence of critical evaluation and independent editing of AI-suggested content
Academic integrity / originality	10%	Transparent, appropriate acknowledgment of AI assistance; no unattributed AI-generated text passed off as original

Source: author's own elaboration, 2026

3.7 Data analysis plan

Quantitative rubric scores would be compared between groups using an independent-samples t-test (or Mann-Whitney U test if normality assumptions are not met) for the post-test and retention test separately, alongside a paired comparison of post-test versus retention scores within each group to assess decay or consolidation of learning. Effect sizes (Cohen's d) would be reported alongside p-values to allow meaningful comparison across future replications of this protocol in other Algerian institutions. Qualitative data drawn from the self-revision justifications would be analysed thematically, following an inductive coding procedure, to characterise patterns of critical engagement with AI suggestions (addressing RQ3). Inter-rater reliability for both the rubric scoring and the thematic coding should be established using a second, independent rater on a representative subsample, with Cohen's kappa reported as a measure of coding consistency.

4. Expected Outcomes and Discussion

Consistent with prior findings on AI-assisted second-language writing, the experimental group is expected to outperform the control group on task achievement, lexical range, and organizational coherence, reflecting the scaffolding and clarification functions previously documented for AI writing tools (Werdiningsih et al., 2024; Mahapatra, 2024; Su et al., 2023). Khampusaen (2025), working with a comparable EFL argumentative-writing population, documented substantial gains in content quality and organizational coherence when ChatGPT was used strategically during essay composition, which offers a useful benchmark against which the present protocol's post-test outcomes could later be compared.

Retention outcomes are anticipated to be more variable. Where AI use remains tightly scaffolded and paired with mandatory self-revision, retention gains are plausible, consistent with the

scaffolding-based gains reported by Guo et al. (2022) and Su et al. (2023). Where students default to passive acceptance of AI suggestions, however, retention gains may be minimal or absent, an outcome consistent with concerns raised about over-reliance in the wider literature on AI-mediated academic work (Bin-Nashwan et al., 2023). This anticipated contrast underscores why the guided-use protocol, rather than unrestricted AI access, constitutes the central pedagogical variable of interest.

The qualitative self-revision data are expected to reveal heterogeneous patterns of engagement rather than a single dominant style. Some learners are likely to treat AI suggestions as a starting point for independent editing, consistent with the critical-engagement patterns Steiss et al. (2024) associate with higher-quality outcomes, while others may show a tendency toward uncritical acceptance, particularly under time pressure or lower initial proficiency. Triangulating these qualitative patterns with the quantitative rubric scores — rather than treating them as separate strands of evidence — should allow future implementations of this protocol to explain, and not merely document, any observed differences between the experimental and control groups. Any future empirical application of this protocol should report retention results transparently, including null or mixed findings, rather than selectively confirming the hypotheses above.

5. Ethical Considerations and Limitations

Several ethical and methodological safeguards are integral to this protocol, not optional additions. First, informed consent must be obtained from all participants, with clear disclosure of the AI tools used and data handled. Second, academic integrity guidelines must be established and communicated before the intervention begins, explicitly defining permissible AI assistance, consistent with recommendations that educational institutions establish clear ethical-use guidelines for generative AI (Bin-Nashwan et al., 2023; Miao & Holmes, 2023). This recommendation is further reinforced by Hamani's (2025) finding that Algerian teacher-researchers themselves overwhelmingly favour guided integration of AI over either prohibition or unrestricted use, suggesting that a rubric-based, guideline-anchored design of the kind proposed here is likely to align with prevailing institutional expectations rather than run against them.

Third, quasi-experimental designs using intact classes rather than individually randomized groups carry an inherent risk of selection bias, which should be explicitly acknowledged and, where possible, mitigated through pre-test score comparison and, where feasible, propensity-score matching between the two intact groups. Fourth, the reliance on free-tier access to Claude, ChatGPT, and Qwen introduces a further limitation: usage caps, model updates, or changes in free-tier availability during the intervention period could affect the consistency of the AI-mediated condition across participants and should be monitored and reported as a fidelity check. Fifth, teacher-researchers implementing this protocol should themselves receive prior orientation on the three tools, since Hamani (2025) found considerable variation among Algerian university teachers in confidence and prior exposure to specific AI systems, which could otherwise introduce implementation inconsistency across sections taught by different instructors.

Finally, any future implementation of this protocol must report actual, unmodified data; presenting simulated or illustrative figures as genuine empirical findings would constitute research misconduct and is explicitly excluded from the scope of this proposal. Researchers adopting this protocol are also encouraged to pre-register their hypotheses and analysis plan, in line with growing calls for transparency in applied linguistics and educational-technology research more broadly.

6. Pedagogical Implications and Conclusion

Even prior to empirical validation, this protocol carries direct pedagogical implications for EFL writing instruction in Algeria's evolving linguistic landscape. It suggests that generative AI tools are unlikely to function as neutral, self-regulating aids; their pedagogical value appears contingent on explicit scaffolding, clear integrity guidelines, and structured opportunities for critical, autonomous revision. As Algerian universities continue to navigate the transition toward English as a medium of instruction and scientific communication, EFL writing pedagogy will need to integrate — rather than ignore or uncritically embrace — generative AI, in ways that are answerable to both pedagogical effectiveness and academic integrity.

The convergence between the pedagogical literature reviewed in Section 1 and the institutional evidence reported by Hamani (2025) — namely, that Algerian teacher-researchers themselves already use these tools and overwhelmingly call for guided rather than unregulated integration — suggests that this protocol responds to an existing institutional readiness rather than proposing a speculative departure from current practice. This protocol offers one replicable framework through which Algerian EFL researchers and practitioners can begin to generate the situated, empirically grounded evidence this transition requires, and its rubric, timeline, and consent procedures are deliberately designed to be adaptable to other language-teaching contexts within the Algerian ENS network and beyond.

Appendices

Appendix A. Indicative diagnostic-quiz prompt (Week 1, thesis-statement construction)

Sample instruction to be entered into the AI tool by the instructor when generating the Week 1 diagnostic quiz: "Generate five short multiple-choice questions, appropriate for an intermediate university-level EFL learner, that test the ability to distinguish a strong, arguable thesis statement from a purely descriptive one. Provide one worked example with a brief explanation." The instructor would review and, if necessary, adapt the AI-generated items before distribution, consistent with the guided-use principle established in Section 3.3.

Appendix B. Sample self-revision justification prompt

After receiving AI-generated suggestions on a draft, students in the experimental group would complete a short written justification for each accepted or rejected suggestion, using a sentence frame such as: "The AI suggested changing ____ to _____. I accepted / rejected this suggestion because _____." These justifications constitute the primary qualitative data source for RQ3 and for the autonomous-revision dimension of the rubric in Table n° 02.

Appendix C. Informed consent — indicative outline

An informed-consent form for this protocol should, at minimum, disclose: (a) the identity of the three AI tools used (Claude, ChatGPT, Qwen) and their free-tier data-handling practices as publicly documented by each provider; (b) the voluntary nature of participation and the absence of any grade penalty for students who prefer the non-AI condition, where institutionally feasible; (c) the anonymisation procedure for writing samples and self-revision justifications used in analysis; and (d) contact details for the responsible instructor and the relevant institutional ethics or scientific committee. The final wording of the consent form should be validated by the host institution's own ethics procedures before implementation, consistent with the safeguards discussed in Section 5.

References

- Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society*, 75, Article 102370. <https://doi.org/10.1016/j.techsoc.2023.102370>
- Fetsi, E. (2025, May). From quills to quantum: Rethinking writing composition in the age of AI [Conference presentation]. *Enseigner l'écrit et l'argumentation à l'université : Enjeux et solutions*, Université Echahid Cheikh Larbi Tébessi, Tébessa, Algeria.
- Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing*, 54, Article 100666. <https://doi.org/10.1016/j.asw.2022.100666>
- Hamani, N. (2025). Teachers' perspectives on AI tools in educational research: A case study at the University of Constantine. *Forum de l'Enseignant / The Teacher's Forum*, 21(1), 875–889.
- Idri, N., & Bouguebs, R. (2021). French linguistic and cultural hegemony in a multilingual context: Case of the Department of English, Bejaia University. *Forum de l'Enseignant / The Teacher's Forum*, 17(1), 328–340.
- Kadik, D., Yaagoub, L., Gada, M., & Bouassal, A. (n.d.). Pour un enseignement/apprentissage du français sur objectifs spécifiques : le cas des étudiants des filières scientifiques à l'université de Médéa (Algérie). *ASJP*, 106–124.
- Khampusaen, D. (2025). The impact of ChatGPT on academic writing skills and knowledge: An investigation of its use in argumentative essays. *LEARN Journal: Language Education and Acquisition Research Network*, 18(1), 963–988. <https://doi.org/10.70730/PGCQ9242>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, Article 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Otmani, A. (n.d.). Satisfying the teachers' needs and expectations from learning English through the content of the English course: The case of teachers at the Teachers' Training School Assia Djebbar, Constantine. *Forum de l'Enseignant / The Teacher's Forum*, 766–781.
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, Article 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, Article 100752. <https://doi.org/10.1016/j.asw.2023.100752>
- 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, Article 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- Werdiningsih, I., Marzuki, & Rusdin, D. (2024). Balancing AI and authenticity: EFL students' experiences with ChatGPT in academic writing. *Cogent Arts & Humanities*, 11(1), Article 2392388. <https://doi.org/10.1080/23311983.2024.2392388>