T I E Journal
Volume 4 Issue 3, 2026
http://ijtie.com/v403/n40309
ISSN: 2820-7521

The International Journal of Technology, Innovation, & Education

ChatGPT-Assisted Formative Feedback and Argumentative Essay Writing Performance in Moroccan Secondary Schools

Hamza Farhane

Department of English, Faculty of Letters and Human Sciences, Dhar Elmahraz, Sidi Mohammed Ben Abdellah University, Fez, Morocco

hamzafarhane96@gmail.com

Received: May 25, 2026; **Accepted**: August 01, 2026; **Published**: October 31, 2026

Abstract

This quasi-experimental study examined the effect of ChatGPT-assisted formative feedback on Moroccan secondary school students' argumentative essay writing performance. Thirty-one second-year Baccalaureate students participated in a four-week intervention: 22 students received teacher-mediated ChatGPT feedback, whereas 9 students received traditional teacher feedback. Pre-test and post-test essays were evaluated by two independent raters using a 20-point analytic framework based on Toulmin's Argumentation Model. Recalculation from the reported group sizes, means, and standard deviations showed no pre-test difference between the groups, $t(29) = 0.08$, $p = .938$. At post-test, the experimental group ($M = 11.06$, $SD = 1.77$) outperformed the control group ($M = 9.62$, $SD = 1.72$), $t(29) = -2.07$, $p = .047$, with a large standardized mean difference (Cohen's $d = 0.82$). A Welch sensitivity analysis produced a similar effect estimate but a marginal p value ($p = .053$), indicating that the inferential result should be interpreted cautiously because of the small, imbalanced groups. The findings provide preliminary evidence that carefully scaffolded ChatGPT-assisted feedback may support EFL argumentative writing, while underscoring the importance of teacher mediation and replication with larger, balanced samples.

Keywords: ChatGPT-assisted feedback, argumentative writing, secondary education, EFL writing, formative feedback

CC BY

1. Introduction

Critical thinking (CT) has become one of the most essential competencies in 21st-century education, particularly in language learning contexts where students are expected not only to acquire linguistic knowledge but also to analyze, evaluate, and communicate ideas effectively. In English as a Foreign Language (EFL) classrooms, critical thinking is increasingly viewed as a necessary skill for enabling learners to engage with information critically, construct logical arguments, and participate actively in academic discourse (Guo & Lee, 2023). As educational systems worldwide continue to emphasize higher-order thinking skills, the integration of critical thinking into writing instruction has gained growing scholarly attention.

Among the various forms of academic writing, argumentative essay writing occupies a central role in fostering critical thinking because it requires students to formulate claims, provide evidence, evaluate opposing viewpoints, and justify their reasoning coherently (Wang & Mat Said, 2024). Through argumentative writing, learners are expected to demonstrate analytical thinking, logical organization, and evidence-based reasoning. However, despite its pedagogical importance, many EFL learners struggle to produce effective argumentative essays. Previous research indicates that students frequently experience difficulties in constructing strong arguments, developing critical reasoning skills, and providing adequate justification for their claims (Wang & Mat Said, 2024). These challenges are particularly evident in EFL contexts where limited language proficiency may further hinder students' ability to express complex ideas critically and coherently.

Feedback plays a crucial role in supporting students' writing development and enhancing their critical engagement with written tasks. Traditionally, teacher feedback has been considered an essential component of writing instruction because it guides learners in identifying weaknesses and improving the quality of their written production. Nevertheless, providing detailed and individualized feedback on students' essays is often time-consuming and demanding for instructors, especially in classrooms with large numbers of learners. Moreover, traditional feedback practices may sometimes focus primarily on surface-level linguistic errors rather than promoting deeper critical engagement with ideas and arguments (Zhang & Hyland, 2018). As a result, students may not receive sufficient support in developing the higher-order thinking skills required for effective argumentative writing.

Recent advances in artificial intelligence (AI), particularly generative AI technologies, have introduced new possibilities for writing instruction and feedback practices. One of the most widely discussed AI tools in education is ChatGPT, a generative AI system capable of

producing human-like responses and interacting with users in a conversational manner. In writing instruction, ChatGPT can identify language and structural issues, generate suggestions for improvement, and provide immediate and personalized formative feedback to learners (Hasanein & Sobaih, 2023; Mahapatra, 2024). Unlike conventional feedback methods that may require substantial teacher time, AI-assisted feedback offers learners rapid responses that can support continuous revision and reflection during the writing process.

Existing studies suggest that EFL students generally perceive AI-generated feedback positively due to its immediacy, accessibility, and clarity (Jamshed et al., 2025). Consequently, AI-assisted feedback practices are becoming increasingly integrated into EFL writing classrooms. Despite this growing interest, research examining the pedagogical implications of ChatGPT in writing instruction remains limited in several important respects. Hikmah and Walida (2024) note that studies investigating the role of ChatGPT responsiveness in students' creative and critical thinking processes are still scarce. Similarly, relatively few quasi-experimental studies have explored the effectiveness of ChatGPT-assisted feedback specifically in argumentative writing contexts. Furthermore, research investigating the development of critical thinking through AI-assisted writing instruction in EFL and non-Western educational settings remains underrepresented in the literature (Shaikh, 2025). Although AI technologies are increasingly promoted as tools for enhancing learning outcomes, the causal impact of ChatGPT-assisted feedback on students' critical thinking and argumentative writing performance remains insufficiently explored.

Within the Moroccan secondary school context, this gap becomes even more significant. While educational reforms increasingly emphasize competency-based learning and critical thinking development, students continue to face challenges in argumentative writing and analytical reasoning in English language classrooms. At the same time, teachers often encounter practical constraints that limit their ability to provide extensive individualized formative feedback. In this regard, ChatGPT-assisted feedback may offer an alternative instructional support mechanism capable of enhancing students' writing performance while reducing some of the pressures associated with traditional feedback practices.

Therefore, the present study aims to investigate the impact of ChatGPT-assisted formative feedback on Moroccan high school students' argumentative essay writing performance. More specifically, the study seeks to answer the following research question:

RQ: What is the impact of ChatGPT-assisted formative feedback on Moroccan high school students' argumentative essay writing performance?

The significance of this study lies in its contribution to the growing body of research on artificial intelligence in language education. By examining the effectiveness of ChatGPT-assisted formative feedback in improving argumentative writing performance, the study provides empirical evidence regarding the pedagogical potential of generative AI in EFL contexts. Additionally, the findings may support writing instructors by offering alternative feedback practices that can help address time constraints while promoting students' higher-order thinking and argumentative writing abilities.

2. Literature Review

2.1. Formative Feedback in EFL Writing Instruction

Formative feedback (FF) plays a central role in the development of students' writing abilities, particularly in English as a Foreign Language (EFL) contexts where learners often require continuous support throughout the writing process. Unlike summative assessment, which primarily evaluates final performance, formative feedback is a continuous instructional process that provides learners with real-time guidance while they are writing (Zhu et al., 2020). Through formative feedback, students are able to identify weaknesses, revise their ideas, and gradually improve the quality of their written production.

In writing instruction, formative feedback can address multiple dimensions of writing performance, including content development, organization, grammar, vocabulary, and style (Bozorgian & Yazdani, 2021; Zhang et al., 2025). Effective feedback not only helps students correct linguistic errors but also encourages them to reflect critically on the structure and coherence of their arguments. Consequently, formative feedback is widely regarded as an important pedagogical tool for supporting both language development and higher-order thinking skills in argumentative writing tasks.

Despite its pedagogical value, implementing formative feedback effectively remains a major challenge in many educational contexts. Providing individualized and detailed feedback on students' writing is often highly time-consuming, particularly in large classrooms where teachers must respond to numerous learners simultaneously (Golzar et al., 2022). As a result, instructors may struggle to address students' individual learning needs adequately. Mahapatra (2024) further argues that meeting students' diverse feedback needs in large EFL classrooms represents a significant instructional challenge, especially when teachers face limited time and heavy workloads. These constraints have contributed to increasing interest in technological solutions that can support writing instruction and feedback practices.

2.2. AI-Assisted Feedback and Critical Thinking Development

Recent developments in artificial intelligence (AI) have transformed educational practices and introduced new possibilities for writing support and feedback. In particular, Automated Writing Evaluation (AWE) tools such as Grammarly and QuillBot have gained popularity for assisting students in identifying grammatical errors, improving sentence structure, and rephrasing written content (Mahapatra, 2021). These tools provide learners with rapid feedback that can facilitate revision and promote greater autonomy during the writing process.

More recently, generative AI technologies such as ChatGPT have expanded the capabilities of AI-assisted writing support beyond conventional automated correction tools. Unlike earlier AWE systems, ChatGPT can engage interactively with users and provide more comprehensive assistance during writing tasks. For instance, it can help students generate ideas, improve the organization of their essays, explain grammar rules, simplify or adjust text difficulty, and translate content when necessary. Such features position ChatGPT as a potentially valuable pedagogical tool in EFL writing classrooms, particularly for learners who require immediate and individualized support throughout the writing process.

Nevertheless, empirical findings concerning the impact of AI tools on critical thinking development remain mixed. Some studies suggest that AI-assisted feedback can support idea generation and enhance students' exploration of different perspectives during writing tasks (Guo & Lee, 2023). By encouraging students to expand their ideas and consider alternative viewpoints, AI tools may contribute positively to certain aspects of critical engagement in argumentative writing.

However, researchers also emphasize important limitations and concerns associated with unrestricted AI use in education. Guo and Lee (2023) note that while AI may assist students in exploring ideas, deeper reasoning and effective problem resolution often continue to require human guidance. Similarly, Kasneci et al. (2023) warn that excessive dependence on AI-generated suggestions may lead students to accept responses passively without sufficient reflection or evaluation. Such over-reliance may reduce learners' metacognitive engagement and weaken opportunities for independent critical reasoning. Therefore, although AI-assisted feedback offers promising instructional benefits, its effectiveness may depend largely on how it is integrated pedagogically within the learning environment.

2.3. Theoretical Framework

This study is grounded in Sociocultural Learning Theory, which conceptualizes learning as a socially mediated process in which cognitive development occurs through interaction,

guidance, and the use of mediating tools. According to this perspective, learners develop higher-order thinking skills through scaffolded support that enables them to perform tasks beyond their current level of independent ability. In educational settings, technological tools can serve as mediational means that facilitate learning and cognitive growth.

Within this framework, ChatGPT is conceptualized as a scaffolding tool that complements rather than replaces teacher feedback. Drawing on the work of Lantolf and Thorne (2006), the present study views AI-assisted formative feedback as a supportive mechanism that can guide students' engagement with argumentative writing tasks while still requiring teacher mediation and learner reflection. Through guided interaction with AI-generated feedback, students may receive additional support in developing argumentative structure, refining ideas, and engaging more critically with their writing processes.

3. Materials and Methods

3.1. Research Design

This study adopted a quasi-experimental pre-test/post-test control group design to investigate the impact of ChatGPT-assisted formative feedback on students' argumentative essay writing performance. Two intact groups were used without random assignment, allowing the study to be conducted within a natural classroom setting. The design enabled a comparison between students receiving ChatGPT-assisted feedback and those receiving traditional teacher feedback.

3.2. Participants and Context

The study was conducted at Imam El Ghazali High School in Sidi Bennour, Morocco. The participants were 31 second-year Baccalaureate (2nd BAC) students. The experimental group consisted of 22 students from the Mechanical Engineering stream, while the control group included 9 students from the Mathematical Sciences stream. Because intact classes were used, the groups could not be balanced through random allocation. The small control group and unequal group sizes reduce statistical power, make estimates more sensitive to individual observations, and limit the precision and external validity of between-group comparisons. Results are therefore interpreted as preliminary evidence from a pilot study.

3.3. Instruments

The primary instrument for data collection was students' argumentative essays, which were analyzed using Toulmin's Argumentation Model. The model includes six components:

- Claim (main argument)

- Evidence (supporting data or facts)
- Warrant (logical connection between claim and evidence)
- Backing (additional justification for the warrant)
- Qualifier (degree of certainty)
- Rebuttal (consideration of counterarguments)

To align with the Moroccan grading system, a 20-point analytic scoring framework was developed around the six Toulmin elements. Claim and evidence received greater emphasis than the remaining elements because the assessment prioritized argument quality and reasoning development.

To ensure reliability and objectivity, essays were independently scored by two experienced teacher-researchers who were not involved in the instruction process. The raters were blind to participants' identities to reduce bias. A shared analytic rubric was used to ensure consistency in scoring across raters.

3.4. Treatment Description

Before the intervention, both groups received an introductory session on argumentative essay writing. This session focused on essay structure, key argumentative components, and basic outlining strategies.

The intervention lasted four weeks. In Week 1, both groups completed a pre-test in the form of an argumentative essay. During Weeks 2 and 3, the experimental group received training on the use of ChatGPT for formative feedback. This training lasted one hour and followed guidelines adapted from Mahapatra (2024). The training aimed to ensure consistent use of ChatGPT across students and included prompting strategies designed to:

- Improve argument quality and reasoning,
- Standardize AI interaction,
- Align feedback use with assessment criteria.

Throughout the intervention, the teacher monitored students' use of ChatGPT to ensure appropriate application. The teacher's role included checking the relevance of AI-generated feedback, preventing over-reliance or direct copying, and clarifying misleading or inaccurate suggestions. In this way, ChatGPT functioned as a guided scaffolding tool rather than an autonomous evaluator.

The control group received traditional teacher feedback during the same period. Feedback focused on key argumentative elements, including thesis clarity, adequacy of evidence, strength

of reasoning, and organization. However, limited attention was given to qualifiers and rebuttals. No AI tools were used with the control group.

In Week 4, both groups completed a post-test argumentative essay under similar conditions to the pre-test.

3.5. Data Collection Procedures

Data were collected in three stages: pre-test, intervention (feedback phase), and post-test. Both groups followed the same timeline and completed equivalent writing tasks to ensure comparability of results.

3.6. Data Analysis

Independent-samples t-tests were used to compare the experimental and control groups. The reported group sizes, means, and standard deviations were used to recalculate pooled-variance tests with df = n1 + n2 - 2 = 29. Cohen's d was calculated using the pooled standard deviation. Because the groups were small and unequal in size, a Welch t-test was also calculated as a sensitivity analysis. Shapiro-Wilk results from the original analysis are reported below.

3.7. Scoring Reliability

To ensure scoring reliability, essays from both pre-test and post-test stages were evaluated by two independent teacher-researchers using the Toulmin-based analytic rubric. Inter-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC). The results indicated good agreement in the pre-test phase (**ICC = .80**) and very high agreement in the post-test phase (**ICC = .90**) (Koo & Li, 2016). Final scores used in the analysis were the averaged ratings of both raters.

4. Results

4.1. Pre-test and Post-test Descriptive Results

The descriptive statistics for both groups in the pre-test and post-test phases are presented in Table 1. At baseline, both groups showed very similar performance, indicating comparable initial levels of argumentative writing ability. However, after the intervention, the experimental group demonstrated noticeably higher improvement compared to the control group.

Table 1. *Pre-test and Post-test Mean Scores of Control and Experimental Groups*

Group	Pre-test Mean	Pre-test SD	Post-test Mean	Post-test SD
Control	7.95	1.38	9.62	1.72
Experimental	7.91	1.25	11.06	1.77

A Shapiro-Wilk test was conducted in the original analysis to assess normality. The reported pre-test p values were .414 for the control group and .108 for the experimental group. At post-test, the control group remained consistent with normality (p = .860), whereas the experimental group showed a slight deviation (p = .044).

4.2. Pre-test Comparison (Independent-Samples t-test)

The independent-samples t-test showed no statistically significant pre-test difference between the control group (M = 7.95, SD = 1.38) and the experimental group (M = 7.91, SD = 1.25), t(29) = 0.08, p = .938. The groups therefore showed comparable mean argumentative writing performance at baseline, although the non-randomized design does not establish equivalence on all relevant characteristics.

4.3. Post-test Comparison (Independent-Samples t-test)

The pooled-variance t-test indicated that the experimental group (M = 11.06, SD = 1.77) scored higher at post-test than the control group (M = 9.62, SD = 1.72), t(29) = -2.07, p = .047. The standardized mean difference was large (Cohen's d = 0.82). However, a Welch sensitivity analysis yielded t(15.34) = -2.10, p = .053. Thus, the estimated difference is substantial, but its statistical significance is sensitive to the test specification and should be interpreted as preliminary rather than conclusive evidence.

5. Discussion

The present study investigated the effect of ChatGPT-assisted formative feedback on Moroccan secondary school students' argumentative essay writing performance. The pre-test comparison showed no meaningful mean difference between the experimental group (M = 7.91) and the control group (M = 7.95). This result supports comparability of the observed baseline writing scores, but it should not be treated as proof that the non-randomized groups were equivalent on every relevant characteristic.

At post-test, the experimental group (M = 11.06) scored higher than the control group (M = 9.62). The pooled-variance comparison was marginally significant, t(29) = -2.07, p = .047, and the standardized difference was large (d = 0.82). Nevertheless, the Welch sensitivity analysis produced p = .053. The observed effect is therefore educationally promising but statistically fragile. It may reflect benefits of immediate, iterative feedback for revision and argument development, but the small, imbalanced sample prevents a strong causal or generalizable conclusion.

The pattern is consistent with research describing the potential value of formative and AI-supported feedback in writing instruction (Mahapatra, 2021; Mahapatra, 2024). However, this study measured performance on argumentative essays through a Toulmin-based framework; it did not administer an independent, validated measure of critical thinking. Accordingly, the findings concern argumentative writing performance and only suggest possible implications for critical engagement. They do not demonstrate that ChatGPT redefines or independently develops critical thinking.

The findings also support examining AI-generated feedback together with teacher mediation. Previous research suggests that AI tools can assist with surface-level writing features, while deeper argumentation and critical reasoning may still require human guidance. In this study, teacher monitoring was intended to encourage students to evaluate ChatGPT-generated feedback rather than accept or copy it passively. This guided integration may have contributed to the observed post-test difference, although the design cannot isolate the effects of the AI tool from those of training, monitoring, and additional feedback opportunities.

Overall, the results offer preliminary evidence that teacher-mediated ChatGPT feedback may support argumentative writing in EFL classrooms. The intervention should be understood as a scaffolded complement to instruction, not a replacement for teacher judgment. Replication with larger, balanced samples and analyses that model change over time is needed before firm claims can be made about effectiveness or critical-thinking development.

6. Conclusion

This pilot study examined ChatGPT-assisted formative feedback in Moroccan secondary-school argumentative writing. The experimental group obtained a higher post-test mean than the control group (11.06 vs. 9.62), with a large standardized mean difference ($d = 0.82$). The pooled t-test was marginally significant ($p = .047$), whereas the Welch sensitivity analysis narrowly exceeded the conventional threshold ($p = .053$). The results therefore indicate a promising difference, not definitive evidence of effectiveness.

From a sociocultural perspective, ChatGPT can be interpreted as a mediating tool within a teacher-guided feedback process. In this study, it provided immediate suggestions while the teacher monitored relevance, discouraged copying, and clarified misleading feedback. Because the outcome was argumentative essay performance rather than an independent measure of critical thinking, theoretical implications for critical-thinking development remain tentative.

Pedagogically, ChatGPT may extend opportunities for timely revision in EFL classrooms where individualized teacher feedback is difficult to sustain. Responsible integration requires teacher oversight, prompt literacy, verification of generated suggestions, and explicit academic integrity guidance. AI feedback should be treated as material for evaluation and revision rather than as an authoritative answer or substitute for students' reasoning.

Several limitations constrain interpretation. The study involved one institution, lasted four weeks, and used intact groups of 22 and 9 students without random assignment. The small control group and unequal group sizes reduce power and precision, increase sensitivity to individual scores, and contributed to a result that changed slightly across pooled and Welch tests. No standardized proficiency measure or delayed post-test was used, and no within-group analysis was available.

In conclusion, teacher-mediated ChatGPT formative feedback shows promise as support for EFL argumentative writing, but the present pilot evidence is not conclusive. Careful scaffolding and ethical guidance are required before the approach can be recommended broadly.

Acknowledgement

Artificial Intelligence Statement: ChatGPT was used as a complementary tool that helps the participants get formative feedback on their drafts rather than generate content for them.

Ethics Approval Statement: This study was conducted in accordance with the principles outlined in the Belmont Report and the Declaration of Helsinki. The research involved minimal risk to participants and was limited to standard educational practices within the classroom.

Disclosure Statement

The authors declare that there is no conflict of interest regarding the publication of this article. No financial, personal, or professional relationships have influenced the research, analysis, or conclusions presented in this work.

Funding Statement

This research paper was not funded by any institution.

Notes on Contributors

Hamza Farhane holds a PhD in Applied Linguistics from the Faculty of Letters and Human Sciences, Dhar El Mahraz, Sidi Mohammed Ben Abdellah University, Fez, Morocco. His research interests include English language teaching (ELT) and the integration of artificial intelligence (AI) in education. He serves as an adjunct professor at the Higher School of Education and Training (ESEF) in Berrechid, Morocco.

hamzafarhane96@gmail.com

ORCID

Hamza Farhane *https://orcid.org/0009-0000-7701-1128*

References

Bozorgian, H., & Yazdani, A. (2021). Direct written corrective feedback with metalinguistic explanation: Investigating language analytic ability. *Iranian Journal of Language Teaching Research*, *9*(1), 65-85.

Cohen, J. (1988). Set correlation and contingency tables. *Applied psychological measurement*, *12*(4), 425-434.

Golzar, J., Momenzadeh, S. E., & Miri, M. A. (2022). Afghan English teachers' and students' perceptions of formative assessment: A comparative analysis. *Cogent Education*, *9*(1), 2107297.

Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. *Journal of Chemical Education*, *100*(12), 4876-4883.

Hasanein, A. M., & Sobaih, A. E. E. (2023). Drivers and consequences of ChatGPT use in higher education: Key stakeholder perspectives. *European journal of investigation in health, psychology and education*, *13*(11), 2599-2614.

Hikmah, D., & Walida, B. (2024). The role of ChatGPT in enhancing students' critical thinking in academic writing. *Ethical Lingua: Journal of Language Teaching and Literature*, *11*(2).

Jamshed, M., Albedah, F., Ansari, M. S., & Banu, S. (2025). Assessing the Efficacy of AI-Driven Corrective Feedback via WhatsApp Application to Improve ESL Learners'

Writing Skills: An Experimental Study. *International Journal of Interactive Mobile Technologies*, *19*(7).

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, *11*(1), 9.

Mahapatra, S. K. (2021). Online formative assessment and feedback practices of ESL teachers in India, Bangladesh and Nepal: A multiple case study. *The Asia-Pacific Education Researcher*, *30*(6), 519-530.

Shaikh, H. (2025). Feedback as a tool for academic independence: A study on EFL students' research project writing at NUBTK. *Englisia: Journal of Language, Education, and Humanities*, *13*(1), 71-87.

Thorne, S. L., & Lantolf, J. P. (2006). A linguistics of communicative activity. *Bilingual Education and Bilingualism*, *62*, 170.

Wang, Y., & Said, S. M. (2024). The cognitive challenges in argumentative writing for EFL learners: A scoping review. *International Journal of Academic Research in Progressive Education and Development*, *13*(10.6007).

Zhang, H., Quan, J., Zhang, N., & Zhang, L. (2025). Investigating the effects of formative assessment on EFL students' achievement and motivation: a Self-Determination Theory perspective. *Frontiers in Psychology*, *16*, 1664871.

Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, *36*, 90-102.

Zhu, M., Liu, O. L., & Lee, H. S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, *143*, 103668.