International Journal of Learning, Teaching and Educational Research
Vol. 25, No. 8, pp. 231-265, August 2026
https://doi.org/10.26803/ijlter.25.8.9
Received May 30, 2026; Revised Jul 10, 2026; Accepted Jul 13, 2026

Enhancing Formal Writing Skills of Omani Post-Foundation EFL Students through QuillBot (an AI Tool) Integration

Zafar Iqbal Khattak[*], **Binu P. M.**,
Mohsen Ghorbanpoor, **Muhammad Babar Qureshi**,
University of Technology and Applied Sciences Al-Musannah
Oman

Malik Al-Zakwani
University of Technology and Applied Sciences Rustaq
Oman

Abstract. Artificial intelligence (AI) is being integrated into English as a foreign language (EFL) writing instruction, yet empirical evidence regarding its effectiveness in improving learners' writing performance remains inconclusive, particularly in under-researched EFL contexts such as Oman. This study investigated the impact of AI-assisted instruction on formal writing proficiency among Omani post-foundation EFL students (N = 58) using a quasi-experimental pre-test–post-test mixed-methods research design. The intervention integrated QuillBot as a supplementary paraphrasing and grammar feedback tool within structured, teacher-guided writing tasks. A one-way ANCOVA, controlling for pre-test performance, revealed no statistically significant difference, as the AI-assisted intervention did not produce a statistically significant improvement in total formal writing scores or in any individual writing category relative to conventional instruction. Furthermore, the raw post-test scores in several categories, including Organization, favored the control group rather than the experimental group. Qualitative findings obtained through semi-structured interviews with the five most active students revealed broadly positive perceptions of the intervention, including increased confidence, reduced anxiety, and reported gains in error awareness and revision practices. However, this qualitative pattern did not consistently align with the quantitative findings. Overall, the findings indicate that short-term AI-assisted instruction using a single corrective tool did not yield measurable quantitative gains in formal writing performance within the study's timeframe, even though some students reported a positive subjective experience with the tool. Thus, this study recommends the need for AI-tool integration for long-term

Citation :
Khattak, Z, I., Binu, P, M., Ghorbanpoor, M., Qureshi, M, B, & Al-Zakwani, M. (2026). Enhancing Formal Writing Skills of Omani Post-Foundation EFL Students through QuillBot (an AI Tool) Integration. *International Journal of Learning, Teaching and Educational Research*, *25*(8), 231–265. https://doi.org/10.26803/ijlter.25.8.9

[*]Corresponding author: Zafar Iqbal Khattak; *zafar.Iqbal.act@utas.edu.om*

interventions using larger samples, focusing on the measurement of student engagement with AI feedback.

Keywords: AI integration; AI tools; EFL formal writing skills; students' perceptions

1. Introduction

The integration of artificial intelligence (AI) into higher education writing instruction has generated substantial scholarly interest. This interest is driven by the proliferation of AI-assisted tools, from grammar-correction platforms to paraphrasing engines and generative chatbots, and by ongoing debates about their pedagogical value (Ahmadi et al., 2025; Luo et al., 2025). These AI tools provide immediate feedback to English as a foreign language (EFL) learners on language accuracy, organization, and tone as effectively as classroom instruction cannot readily replicate (Abd Rahim & Gregory, 2025; Binu, 2022; Song & Song, 2023). While the advantages of AI integration into language education are widely acknowledged, scholars have also cautioned that its adoption requires careful pedagogical design to avoid surface-level engagement and unintended dependency (Binu, 2024). Similarly, overreliance on AI-generated revisions risks displacing the cognitive engagement necessary for writing development, raising concerns about autonomous competence and academic integrity (Deep & Chen, 2025; Teng, 2024).

These issues are especially significant in formal writing, where register, structure, and genre conventions carry considerable academic weight. Formal writing refers to the use of clear, accurate, and professionally appropriate language that follows established organizational and genre conventions. As a distinct genre, it emphasizes coherence, grammatical correctness, appropriate register, and overall effective communication. The present study is situated within the General Requirement Programme at the University of Technology and Applied Sciences, Al-Mussanah (UTAS-A), Oman. The Technical Writing course (UNEN1203) is a diploma-level course aimed at developing post-foundation students' formal written communication skills for academic and professional purposes.

During a 16-week semester, students are required to produce genre-specific texts, such as formal emails, meeting minutes, scientific and business reports, and product information sheets. In the context of the present study, writing formal emails constitutes one of the main course components that is assessed in the midterm examination (MTE) that is held in Week 7 of the semester. In the MTE, students are tasked with two types of email: sending and replying to emails. Their emails are evaluated based on task achievement, organization, adherence to structural elements, grammar, and vocabulary. Overall, these tasks demand precise language, appropriate tone, and adherence to professional conventions.

Despite these expectations, Omani EFL students at this level frequently struggle with grammatical accuracy, organizational coherence, academic tone, and structural conventions. These difficulties are further compounded by students' limited ability to use digital tools purposefully rather than as replacements for

independent writing (Neiroukh et al., 2024). Against this backdrop, AI-assisted writing tools, particularly paraphrasing and grammar-revision platforms such as QuillBot, have gained growing attention as potential instructional supports for EFL writing development (Luo et al., 2025). However, empirical evidence from Gulf and specifically Omani higher education contexts remains scarce. The conditions under which AI tools can be effectively integrated into formal writing instruction for Omani EFL students remain largely underexplored.

These interconnected gaps informed the present study. Despite growing interest in AI writing tools, existing research has focused primarily on general writing accuracy rather than formal writing proficiency as a composite of academic register, genre convention, and structural precision (Marzuki et al., 2023; Nhan et al., 2025). Additionally, controlled studies have been conducted predominantly in East Asian or Western contexts, leaving Gulf higher education settings underrepresented (Ozfidan et al., 2024). Empirical evidence on QuillBot's specific effects on formal writing proficiency within controlled instructional designs also remains scarce.

To address these gaps, a quasi-experimental pre-test–post-test mixed-methods research design study was conducted at UTAS-A. A total of 37 post-foundation students in Section 18 formed the experimental group and received QuillBot-assisted instruction in a Technical Writing course. An equivalent group of 37 students in Section 16 served as the control group, receiving traditional writing instruction. Data was collected through pre- and post-intervention writing assessments, semi-structured interviews with five active experimental group students, and teacher observations across the five-week instructional period.

Moreover, recent research has increasingly emphasized that the effectiveness of AI-assisted writing tools should not be evaluated solely through improvements in writing performance, but also through their usefulness, considering users' perceptions. Pedagogical adaptability refers to the extent to which AI tools can be integrated into instructional contexts, accommodate learner needs, and support diverse writing tasks and learning objectives (Luo et al., 2025; Ozfidan et al., 2024). Studies suggest that adaptable AI tools can provide personalized feedback, promote learner engagement, and complement teacher instruction when appropriately integrated into the curriculum. However, evidence regarding the practical usefulness of such tools within the Gulf-region higher education contexts remains limited.

Similarly, student perception has emerged as a critical factor influencing the successful adoption and sustained use of AI-assisted learning technologies. Previous studies have reported generally positive perceptions of students regarding AI writing tools, highlighting their usefulness for improving language accuracy, vocabulary development, and confidence in writing. For instance, the motivational aspect of AI-assisted writing tools had notably strong, favorable feedback from students, with 86.88% responding favorably to increased motivation in writing (Abd Rahim & Gregory, 2025). Nevertheless, concerns regarding overreliance, reduced critical thinking, and academic integrity continue

to shape learners' attitudes toward these technologies (Deep & Chen, 2025; Teng, 2024). Research examining Omani EFL students' perceptions of AI-assisted writing tools remains scarce, creating a need for context-specific investigation.

In response to these gaps, the present study adopted a quasi-experimental pre-test–post-test mixed-methods research design to investigate the impact of QuillBot-assisted instruction on formal writing proficiency among Omani post-foundation EFL students while simultaneously examining its usefulness considering students' perceptions.

This study contributes both practically and theoretically to the growing field of AI-assisted language learning. Practically, it provides EFL instructors, curriculum designers, and higher education institutions with empirical evidence regarding the effectiveness of QuillBot for supporting formal writing instruction in Omani higher education. The findings may inform pedagogical decisions concerning the integration of AI-assisted writing tools into Technical Writing courses and similar EFL programs. Furthermore, insights into students' perceptions regarding the usefulness of the tool can guide the development of instructional strategies that maximize educational benefits while minimizing potential risks associated with AI dependence.

Theoretically, the study extends the emerging body of literature on AI-assisted writing by examining writing proficiency, pedagogical usefulness, and student perception within a single conceptual framework. It also addresses a notable geographical gap in the literature by providing evidence from the Omani and Gulf higher education context, where empirical studies on AI-assisted writing technologies remain limited. By exploring these interrelated dimensions, the study contributes to a more comprehensive understanding of how AI tools can support language learning and writing development in EFL settings.

2. Literature Review

Research on AI writing tools in EFL contexts expanded rapidly over the past decade, concentrating predominantly on university-level populations (Ahmadi et al., 2025; Luo et al., 2025). In their review of 52 studies, Luo et al. (2025) identified grammar checkers and paraphrasing tools as the most often studied categories, evaluated primarily on accuracy and coherence outcomes. An important distinction running through this literature was drawn by Aljuaid (2024), who differentiated between corrective tools that operated on student-produced text, such as QuillBot, and generative tools that produced text on behalf of the student, arguing that this functional difference carried pedagogical implications the field had not yet fully theorized. The uptake among EFL students was consistently high.

Ozfidan et al. (2024) reported that Saudi undergraduate students used AI tools regularly for grammar checking and vocabulary enhancement, while Amani and Bisriyah (2025) found that students valued them primarily for supporting self-regulated revision. Selim (2024) similarly found that EFL university students held broadly positive perceptions of AI-powered writing tools, associating them with

improved drafting confidence and reduced language anxiety. Furthermore, in Syafei and Nuraeningsih's study (2025), English language teaching (ELT) students reported enhanced engagement and greater willingness to revise when AI tools were incorporated into writing tasks. Critically, however, most of this evidence addressed general writing accuracy rather than the specific demands of formal writing, and the outcome measures employed rarely captured academic register, genre adherence, or professional structural convention.

Turning to writing quality outcomes, the empirical evidence was broadly positive but contextually variable. Song and Song (2023) reported significant gains in content development and linguistic accuracy among Chinese EFL students using ChatGPT. Zare et al. (2025), in a controlled pre-/post-design, found higher task motivation and perceived writing competence in the AI-integrated group. Their qualitative data, however, also revealed student concerns about overreliance on AI feedback and its potential to undermine independent writing capability, a tension that qualifies the overall positive picture. Mekheimer (2025) demonstrated that proficiency moderated outcomes, with higher-proficiency students showing greater revision depth from AI feedback than lower-proficiency peers.

The specific dimensions of formal writing received considerably less attention than general accuracy in previously reviewed studies on AI-assisted writing. Tuong and Tran (2025) found stronger AI effects on structural organization than on tone or register. Cui et al. (2025) showed that performance gains were most evident when students used tools for revision rather than generation. Against this broadly positive picture, Lo et al. (2025) found that surface corrections increased while deep argumentative revisions remained limited, suggesting AI feedback reinforced editing behavior over higher-order development. Marzuki et al. (2023) similarly reported that EFL teachers were concerned about authenticity and the displacement of compositional reasoning. Of direct relevance to the present study's Omani setting, Neiroukh et al. (2024) found that structured feedback frameworks enhanced learner autonomy when students were explicitly trained to evaluate and selectively apply external feedback. This principle directly informed the pedagogical design adopted in the present study.

Beyond writing quality, a substantial strand of research addressed the motivational and self-regulatory dimensions of AI tool use. A meta-analysis by Ren et al. (2026), across 34 studies, reported a moderate positive effect of AI on learner self-efficacy ($d = 0.52$), which was strongest when feedback was timely and specific. On the other hand, Lim et al. (2026) found that motivational benefits were more pronounced under instructor-guided integration than autonomous use. This optimism was qualified by Erliani et al. (2025), who documented that increased writing confidence did not always transfer to unaided tasks. They termed this phenomenon *borrowed efficacy*, noting that students who improved with AI support often struggled when the tool was removed. Similarly, Apriani et al. (2024) found that lower-proficiency students benefited most from structured guidance on evaluating AI-generated suggestions rather than accepting them uncritically.

Wang et al. (2024) identified metacognitive monitoring as the key mediating variable between tool use and improvement, underscoring that the quality of engagement with AI feedback mattered more than exposure alone. These effective and motivational findings were further complicated by evidence from the Gulf region. For instance, Khattak and Mathew (2024) found that engaging post-foundation Omani students through the MS Teams private chat application can significantly impact their formal writing skills.

Similarly, Al Zakwani and Mathew (2025), in an Omani university context, found high perceived usefulness of ChatGPT alongside reduced motivation for independent writing practice. This dependency dynamic is directly pertinent to the present study's concern with durable formal writing competence. It should be noted, however, that Khattak and Mathew's (2024) study investigated the MS Teams messaging application as a communication tool for enhancing students' formal writing skills, whereas Al Zakwani and Mathew's (2025) study employed ChatGPT as a general language learning resource rather than a specialized instrument for formal writing. Nevertheless, both studies, situated within the same academic context, exhibit limitations in their direct transferability to the constructs of the present study.

The evidence consistently suggests that AI writing tools yield better results when incorporated into purposeful teaching frameworks than when used simply as standalone correction tools. Nhan et al. (2025) argued that tools were most effective within process-writing cycles involving iterative drafting, AI feedback, critical evaluation, and revision. Ekizoğlu and Demir (2025) found that scaffolded exposure, starting with teacher-guided evaluation of AI feedback before students used it independently, was essential. Chen and Gong (2025) demonstrated that combining AI and instructor feedback produced better outcomes than either alone, while Teng and Huang (2025) advocated for a dialogic model in which students and instructors collaboratively applied AI feedback against learning objectives. Zamorano (2025) similarly reported that structured generative AI integration improved organization and lexical range when guided by clear instructional goals.

Within this framework, QuillBot is particularly relevant. It is one of the most widely used AI-assisted writing tools among EFL learners and university students. As discussed earlier, it works on student-produced text rather than generating text independently; it is well-suited to supporting revision and editing processes central to formal writing development. Several studies have highlighted the potential benefits of QuillBot in language learning contexts. Aljuaid (2024) argues that revision-oriented tools such as QuillBot encourage learners to engage with their own writing rather than relying on machine-generated outputs. Similarly, as cited earlier, Ozfidan et al. (2024) found that Saudi undergraduates used paraphrasing tools mainly for sentence-level revision and academic vocabulary improvement, aligning closely with the writing dimensions targeted in the present curriculum.

Academic integrity, however, added complexity. Bui and Tong (2025) found that students valued AI writing tools while acknowledging misuse risks in assessed work. Giray et al. (2025) warned that overreliance on AI-detection tools as a policing mechanism could erode the trust needed for productive AI integration. Mohammed (2024) found that EFL students struggled to distinguish appropriate from inappropriate AI revisions in academic register. This is a particular challenge where register precision is essential to communicative competence.

Taken together, the literature confirms that AI writing tools yield measurable positive outcomes when pedagogically scaffolded. However, it also reveals that formal writing proficiency as a specific outcome, paraphrasing-tool functionality as a distinct category, and Gulf higher education as an institutional setting remain substantially underexplored. Empirical evidence on QuillBot's effects on formal writing proficiency in controlled instructional designs remains limited, and it is this gap that the present study is designed to address.

3. Research Methodology

3.1 Research Design

This study employed a quasi-experimental pre-test–post-test mixed-methods design with a non-equivalent control group to investigate the impact of QuillBot as an AI-assisted writing tool on students' formal writing proficiency in a Technical Writing course (UNEN1203). Although the course is titled Technical Writing, its curriculum encompasses a range of formal professional communication tasks, including formal email writing, which constitutes a major assessed component of the MTE. The present study focused specifically on this formal email writing component, which requires adherence to formal register, organizational conventions, and professional tone – features characteristic of formal rather than technical writing genres.

The independent variable was group membership (QuillBot-assisted instruction vs. conventional instruction) and the dependent variable was formal writing performance as measured by the MTE rubric score (out of 10 points). The quantitative strand compared pre-test and post-test writing scores between groups using ANCOVA, while the supplementary qualitative strand drew on semi-structured interviews to explore students' perceptions of AI tool use. This research design was adopted because intact classroom groups were used rather than randomized individual assignments, reflecting authentic instructional conditions at UTAS-A.

Consistent with gaps identified in the literature regarding controlled studies on paraphrasing tools such as QuillBot in formal writing contexts, this design allowed measurable comparison of writing performance before and after the instructional intervention. Two course sections were purposefully selected. Both groups completed a pre-test and post-test (MTE writing component), enabling quantitative comparison of learning outcomes. Table 1 summarizes the instructional conditions for experimental and control groups.

Table 1: Instructional conditions for experimental and control groups

Group	Section	Instructional approach	Description
Experimental group	Section 18	AI-integrated instruction	Students received writing instructions that incorporated QuillBot (AI-assisted tool). The tool was used to support drafting, revising, and improving formal academic writing under guided instructional conditions.
Control group	Section 16	Traditional instruction	Students received traditional writing instruction aligned with the Technical Writing curriculum. Instruction focused on teacher-led explanations, textbook-based exercises, and manual feedback, with no use of AI-assisted writing tools.

3.2 Participants

The participants in this study were post-foundation students enrolled in Technical Writing (UNEN1203) under the General Requirement Programme at UTAS-A during the Spring semester of the academic year 2025–2026. Two active class sections with 37 students each, taught by the same instructor – the principal investigator of this study – were purposively selected. Section 18 functioned as the experimental group and received instruction integrating QuillBot as an AI-assisted writing tool, whereas Section 16 served as the control group and received standard instruction without the use of AI tools.

Participants in both sections had previously completed foundation-level English language courses and had begun engaging in formal academic and professional writing tasks characteristic of the diploma program. The selection of these two sections taught by the same instructor made it feasible for the investigation, as it restricted the instruction variance. In addition, it was ensured that both groups were comparable in terms of student demographics and overall English proficiency levels. Participation in the study adhered to institutional ethical guidelines, and students were informed that their coursework scores would be used anonymously for research purposes without affecting their academic standing.

Of the 74 students originally enrolled across the two sections, 16 did not complete the pre-test (i.e., 5 from the control group, Section 16; 11 from the experimental group, Section 18) due to being absent on the day of administration. All the students were present for the post-test. Because pre-test scores were required as the covariate in the ANCOVA models, students without a pre-test score were excluded from the analyses, yielding a final analytic sample of 58 students (32 control, 26 experimental). As the missing data resulted from pre-test absence prior to any treatment exposure, excluding these cases is unlikely to introduce a treatment-related attrition bias.

3.3 Research Instruments

3.3.1 Pre-test and post-test (midterm writing component)

The primary instrument used in this study was the formal email writing component of the MTE, which was administered as both a pre-test and a post-test. The pre-test (Appendix 1) was administered during Week 2 of the course, while the post-test (Appendix 2) was administered in Week 7 following the instructional intervention. The writing task assessed students' ability to produce formal written emails based on various scenarios in accordance with the learning objectives of the Technical Writing course.

Specifically, the assessment evaluated students' use of formal register, organizational coherence, grammatical and sentence-level accuracy, and adherence to genre conventions relevant to technical and formal email writing. Student performance was evaluated using a standardized departmental rubric that measured task achievement, organization, vocabulary use, grammatical accuracy, and compliance with genre conventions. This rubric is a university-wide instrument developed centrally by ELT specialists at the UTAS head office rather than specifically for this study. Prior to system-wide implementation, the rubric underwent a piloting phase, and it was periodically revised based on feedback from teachers and course coordinators across all UTAS branches.

No formal psychometric validation data (e.g., construct validity indices or reliability coefficients) for the rubric were available to the researchers at the time of this study. The full rubric is provided in Appendix 3. The resulting scores constituted the primary quantitative data used to measure changes in students' formal writing proficiency. All writing samples across both groups were scored by the course instructor, who also served as the principal investigator and sole instructor for both sections. No second rater independently assessed any of the writing samples, and the scoring was not conducted blind to group allocation, as the instructor was aware of which section had produced each script. These conditions introduce the possibility of scoring bias, which cannot be ruled out and is discussed further in the limitations section.

3.3.2 QuillBot as an intervention instrument

For the experimental group (Section 18), the instructional intervention incorporated QuillBot as the AI-assisted writing tool. QuillBot was used to assist students with paraphrasing, improving grammar, and adjusting tone during writing tasks.

QuillBot was selected as the primary AI-assisted writing tool because it aligns with the existing literature, which identifies it as an AI-based paraphrasing and corrective feedback tool. It has also been shown to be an effective tool for revision-based writing instruction (Ozfidan et al., 2024). Importantly, students in the experimental group were explicitly guided to use AI tools critically by evaluating and reflecting on AI-generated suggestions rather than accepting them without scrutiny. This instructional approach was intended to promote learner autonomy and minimize overdependence on AI-generated output.

Teacher-guided instruction was retained alongside QuillBot use for two reasons. First, QuillBot provides surface-level corrections and paraphrase suggestions but does not explain the underlying linguistic rules. This makes teacher mediation necessary for developing metalinguistic awareness rather than passive reliance on AI output. Second, as stated in the literature review, AI writing tools yield stronger outcomes when embedded within teacher-directed pedagogical frameworks rather than used independently (Chen & Gong, 2025; Nhan et al., 2025; Teng & Huang, 2025). The guided instructional design therefore aimed to ensure that students engaged critically with AI suggestions rather than accepting them uncritically.

3.3.3 Students' perception through short semi-structured interviews

Although the primary focus of the study was quantitative, brief semi-structured interviews were conducted with the five most active students (Appendix 4) to determine their perceptions regarding the usefulness of AI-assisted writing tools. These reflections served as supplementary qualitative data and supported research objectives related to pedagogical usefulness and learners' experiences with AI integration.

Semi-structured individual interviews were employed as the qualitative data collection method to capture students' reflections on their experience with AI-assisted writing tools. In practice, students who were less engaged with the AI tools during the intervention also showed correspondingly little interest in volunteering for an interview when invited. Additionally, because the researcher and participants did not share a first language (Arabic), articulating nuanced reflections in English was expected to be more demanding for lower-proficiency students, which may have further discouraged participation among this group.

As a result, the five students who agreed to be interviewed were purposively drawn from among the most active users of AI tools in the experimental group, who had also expressed willingness and relative confidence in discussing their experience in English. Interview data were analyzed using thematic analysis following the six-phase approach outlined by Braun and Clarke (2006). The six steps are familiarization with the data, initial code generation, theme development, theme review, theme definition and naming, and final write-up.

3.3.4 Observation and instructional logs

To complement the quantitative data, the instructor maintained two types of contextual records throughout the intervention period. Observation notes documented in-class student behavior, including observable engagement with AI tools, peer interaction during drafting and revision tasks, and classroom dynamics related to AI use. Instructional logs recorded the pedagogical decisions made during each session, including the tasks assigned, the AI tool functions demonstrated, and any adaptations made to the planned activities. Some in-class writing output from the experimental group was saved for comparison purposes (see Appendix 5). These records were maintained as a contextual account of the instructional process rather than as a formally coded qualitative data source.

3.4 Procedure

During the second week of the course, both the experimental (Section 18) and control groups (Section 16) took a pre-test based on a standardized old MTE writing task (Appendix 1) under identical testing conditions. Week 2 was selected as the baseline administration point for two practical reasons. First, section enrollment is not finalized until shortly before the course begins, making it impossible to confirm group composition earlier. Second, Week 1 functions as an orientation period, in which students are introduced to the course syllabus, assessment requirements, and timeline. Because Week 1 typically begins mid-week and includes limited contact time, no substantive instructional content is delivered during this period. The beginning of Week 2, therefore, represented the earliest practical point at which a baseline measure of students' formal writing proficiency could be obtained under stable section assignments and before the start of substantive instruction.

The pre-test scores served as baseline measures of students' formal writing proficiency levels prior to instructional intervention. Over the following five weeks of instruction (Weeks 2 through 6), these two groups experienced distinct instructional approaches for scenario-based writing of formal emails. The experimental group in Section 18 was introduced to QuillBot (the AI-assisted writing tool) through guided demonstrations. The instructor kept students engaged in AI-supported drafting and revision tasks during the class. He also guided them on evaluating AI suggestions, maintaining academic integrity, and revising tone, clarity, and text structure.

During in-class writing activities, students were required to complete an initial draft independently before using QuillBot to revise and improve subsequent drafts. This scaffolded approach, in which AI assistance was restricted to the revision stage rather than the drafting stage, was adopted in line with principles of guided AI integration in writing instruction (Almashour et al., 2025; Begmatova & Saydazimova, 2026; Mhamdi, 2026). They received focused instructions on how to achieve the overall task in the forthcoming exam by using all the prompts given in the scenario question: to use appropriate cohesive devices for organization; to correct their grammar mistakes and spellings in their first draft by taking guidance from the AI tools; and to paraphrase and maintain the right tone and register in their drafts using QuillBot to correct the structural form of the formal email by looking at the AI-generated model answers.

In contrast, the control group in Section 16 followed the same course activities on formal emails without the use of AI tools. Instruction in this group relied on traditional pedagogical practices, including direct teacher feedback, peer review activities, and manual drafting and revision processes. Toward the end of the instructional period, both groups completed the MTE writing component as a post-test (Appendix 2) under identical conditions. Instruction concluded at the end of Week 6, and the post-test was administered at the beginning of Week 7, immediately following the closing of instruction, with no intervening delay beyond the standard weekend break. The post-test, therefore, functioned as an immediate measure of writing performance following the intervention rather than

a delayed measure of retention. The post-test scores were used to assess changes in students' formal writing proficiency and to determine the impact of AI-assisted instruction over the duration of the course.

4. Data Analysis

All statistical analyses were conducted using IBM SPSS Statistics, Version 27 (IBM Corp., 2020). To examine whether the AI-assisted feedback intervention produced a statistically significant difference in the writing performance of EFL students, a one-way analysis of covariance (ANCOVA) was employed as the primary analytical approach for both total writing scores and each of the five writing categories (Task Achievement, Organization, Grammar, Lexical Resource, and Structural Elements).

ANCOVA was selected because it adjusts post-test scores for pre-existing differences on the covariate while preserving the original scale of measurement, making it particularly appropriate for quasi-experimental pre-test−post-test designs where random assignment is not possible (Field, 2013). In each ANCOVA model, the post-test score served as the dependent variable, the corresponding pre-test score served as the covariate to control pre-existing differences in performance, and group membership (control vs. experimental) served as the fixed factor. Effect sizes were reported as partial eta squared (partial η^2), interpreted using the benchmarks of .01 (small), .06 (medium), and .14 (large) proposed by Cohen (1988).

Prior to conducting the main analyses, a series of assumption checks were performed, including tests of pre-test equivalence between groups (independent samples *t*-test), homogeneity of regression slopes (Group x pre-test interaction term), linearity (scatterplot inspection), homogeneity of error variance (Levene's test), and normality of residuals (Shapiro-Wilk test). Where the normality assumption appeared violated for total scores, a Mann-Whitney U test on gain scores (post-test minus pre-test) was conducted as a non-parametric robustness check. Gain scores were used for the non-parametric check to directly assess improvement while mitigating normality concerns in ANCOVA residuals, and their use was supported by the pre-test equivalence of the two groups.

Where Levene's test indicated a violation of the homogeneity of error variance assumption at the category level, Welch's *F* test and a Mann-Whitney U test on gain scores were conducted as robustness checks. To control Type, I error inflation across the five simultaneous category-level comparisons, a Bonferroni-corrected significance threshold of .010 (0.05 / 5) was applied to control the family-wise error rate for the five category-level tests (Dunn, 1961). Because the total score and the five category scores represent two conceptually distinct analytical questions, namely the overall intervention effect versus category-specific effects, the Bonferroni correction was applied only within the family of five category-level comparisons. The total score analysis was treated as a single a priori confirmatory test of the study's primary hypothesis and was therefore not subject to multiple-comparison correction.

A post hoc sensitivity power analysis was conducted using G*Power 3.1.9.7 (Faul et al., 2009) to determine the minimum effect size detectable given the achieved sample size, since no a priori power analysis had been conducted prior to data collection. With N = 58, one covariate, two groups, and power fixed at .80, the analysis indicated that the design was sensitive to effects of partial $\eta^2 \geq .123$ at the conventional alpha level of .05, and to effects of partial $\eta^2 \geq .176$ at the Bonferroni-corrected threshold of a = .010 applied to the category-level comparisons. The latter exceeds Cohen's (1988) benchmark for a large effect, indicating that the corrected category-level analyses were sensitive only to very large effects.

4.1 Quantitative Findings: Writing Improvement

Before the main analysis, statistical assumption checks were conducted. The homogeneity of regression slopes assumption was satisfied (group x pre-test interaction) (F[1, 54] = 2.055, p = .158). All other key assumptions were found to be sufficiently met, except for residual normality for total scores (Shapiro-Wilk W (58) = 0.937, p = .005) and homogeneity of error variance for Organization (Levene's p = .043) and Structural Elements (Levene's p < .001). These violations prompted the use of non-parametric and robust alternative tests, as described in the Data Analysis section. Full assumption check results are reported in Appendix 6.

4.1.1 Total writing score analyses

a) Descriptive statistics

Table 2 presents the pre-test and post-test means and standard deviations and the ANCOVA-adjusted post-test means for both groups on total writing scores.

Table 2: Total writing score descriptive statistics and adjusted means by group

Group	*n*	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	Adjusted *M* (*SE*)
Control	32	6.7266 (1.34158)	7.0547 (1.20940)	7.052 (0.193)
Experimental	26	6.6635 (1.35820)	6.8558 (0.91699)	6.859 (0.214)

Note. Adjusted means are estimated marginal means from ANCOVA with pre-test scores as the covariate, evaluated at the grand mean of pre-test scores (M = 6.6983). SE = standard error. 95% CI: Control [6.666, 7.438]; Experimental [6.431, 7.288].

4.1.2 Primary analysis: ANCOVA

A one-way ANCOVA was conducted with post-test total writing score as the dependent variable, group as the fixed factor, and pre-test total writing score as the covariate. After controlling pre-test performance, the main effect of group did not appear statistically significant (F[1, 55] = 0.447, p = .507, partial η^2 = .008). The effect size appeared negligible relative to the conventional benchmarks proposed by Cohen (1988). The covariate did not reach statistical significance either (F[1, 55] = 0.939, p = .337, partial η^2 = .017). The overall model accounted for only 2.5% of the variance in post-test scores (R^2 = .025, adjusted R^2 = -.010). The adjusted post-test means were closely aligned across groups (Control: M = 7.052, SE = 0.193; Experimental: M = 6.859, SE = 0.214), with a difference of only 0.193 points on the 10-point scale. Table 3 presents the full ANCOVA summary.

Table 3: ANCOVA summary for total writing post-test scores

Source	*SS*	df	*MS*	*F*	*p* / η^2p
Corrected model	1.681	2	0.841	0.709	.497 / .025
Pre-test (covariate)	1.114	1	1.114	0.939	.337 / .017
Group	0.530	1	0.530	0.447	.507 / .008
Error	65.250	55	1.186		
Total	2881.000	58			
Corrected total	66.931	57			

Note. SS = Type III sum of squares. η^2p = partial eta squared. R^2 = .025 (adjusted R^2 = -.010).

a) Robustness check: Mann-Whitney U test

In view of the residual normality violation, a Mann-Whitney U test was conducted on total writing gain scores as a non-parametric robustness check. Table 4 presents the gain score descriptive statistics by group.

Table 4: Total writing gain score descriptive statistics by group

Group	*n*	*M*	*SD*	Mdn	Mean rank
Control	32	0.3281	1.54429	0.7500	29.84
Experimental	26	0.1923	1.70925	0.5000	29.08

Note. M = mean gain score. SD = standard deviation. Mdn = median gain score. Mean rank = mean rank from Mann-Whitney U analysis.

The Mann-Whitney U test indicated no statistically significant difference in gain scores between groups (U = 405.000, Z = -0.173, p = .863). The mean ranks were nearly identical (Control: 29.84; Experimental: 29.08). Table 5 presents a convergence summary of both total score analyses.

Table 5: Convergence of ANCOVA and Mann-Whitney U results for total writing scores

Analysis	Test statistic	*p*	Decision
ANCOVA	F (1, 55) = 0.447	.507	Non-significant
Mann-Whitney U	U = 405.000, Z = -0.173	.863	Non-significant

Note. ANCOVA controlled for pre-test total writing scores as a covariate. Mann-Whitney U test was conducted on gain scores. Wilcoxon W = 756.000.

Both analyses converged on a non-significant result, suggesting that the ANCOVA finding was unlikely to be an artefact of the violation of the residual normality assumption. The AI-assisted feedback intervention did not appear to produce a statistically significant effect on overall EFL writing performance compared to conventional instruction under either parametric or non-parametric analyses.

4.1.3 Writing category analyses

Table 6 presents the post-test descriptive statistics and ANCOVA-adjusted means for each writing category by group.

Table 6: Post-test descriptive statistics and adjusted means by writing category and group

Category	Group	*n*	Post-test *M* (*SD*)	Adj *M* (*SE*)	95% CI
Task Achievement (4 pts)	Control	32	2.8203 (0.64519)	2.816 (0.099)	[2.618, 3.015]
	Experimental	26	2.8769 (0.42994)	2.882 (0.110)	[2.661, 3.102]
Organization (2 pts)	Control	32	1.4609 (0.24683)	1.462 (0.043)	[1.376, 1.549]
	Experimental	26	1.3077 (0.23778)	1.306 (0.048)	[1.210, 1.402]
Grammar (1.5 pts)	Control	32	0.9063 (0.19828)	0.906 (0.034)	[0.838, 0.973]
	Experimental	26	0.8846 (0.17650)	0.885 (0.037)	[0.811, 0.960]
Lexical Resource (1.5 pts)	Control	32	0.9063 (0.19828)	0.906 (0.032)	[0.841, 0.971]
	Experimental	26	0.8750 (0.16202)	0.875 (0.036)	[0.803, 0.947]
Structural Elements (1 pt)	Control	32	0.9609 (0.11197)	0.961 (0.027)	[0.907, 1.015]
	Experimental	26	0.8750 (0.19039)	0.875 (0.030)	[0.815, 0.935]

Note. Adj M = ANCOVA-adjusted mean (estimated marginal mean evaluated at the grand mean of the corresponding pre-test covariate). SE = standard error. 95% CI = confidence interval for the adjusted mean. Points in parentheses indicate the maximum score for each category.

Table 7 presents the ANCOVA group effects for all five categories.

a) Task Achievement

After controlling for pre-test Task Achievement scores, the group effect did not appear statistically significant ($F[1, 55] = 0.195$, $p = .661$, partial $\eta^2 = .004$). The adjusted means were nearly identical across groups (Control: $M = 2.816$, $SE = 0.099$, 95% CI [2.618, 3.015]; Experimental: $M = 2.882$, $SE = 0.110$, 95% CI [2.661, 3.102]).

Table 7: ANCOVA group effects for each writing category

Category	Levene's *p*	*df*	*F*	*p*	η^2p	Decision
Task Achievement	.101	(1, 55)	0.195	.661	.004	Non-significant
Organization	.043*	(1, 55)	5.826	.019	.096	$p < .05$; fails Bonferroni
Grammar	.539	(1, 55)	0.161	.690	.003	Non-significant
Lexical Resource	.367	(1, 55)	0.410	.524	.007	Non-significant
Structural Elements	< .001*	(1, 55)	4.494	.039	.076	See robustness check

*Note. F values reported for the group effect (df = 1, 55 for all categories). η^2p = partial eta squared. Levene's p = significance of Levene's test for equality of error variances. * Indicates Levene's test was violated. Bonferroni-corrected alpha = .010.*

b) Organization

After controlling for pre-test Organization scores, the group effect reached statistical significance at the uncorrected alpha level (F[1, 55] = 5.826, p = .019, partial η^2 = .096), which may be interpreted as a medium effect by conventional benchmarks (Cohen, 1988). Given Levene's test violation (p = .043), Welch's F test and the Mann-Whitney U test on gain scores were conducted as robustness checks (see Table 8). All three tests converged on a statistically significant group difference favoring the control group (Control: Adj M = 1.462, SE = 0.043, 95% CI [1.376, 1.549]; Experimental: Adj M = 1.306, SE = 0.048, 95% CI [1.210, 1.402]). However, after applying the Bonferroni correction, the result did not survive at the adjusted threshold of .010 (p = .019 > .010). The finding should therefore be interpreted with caution as a trend rather than a confirmed effect.

c) Grammar

After controlling for pre-test Grammar scores, the group effect did not appear statistically significant (F[1, 55] = 0.161, p = .690, partial η^2 = .003). The adjusted means were closely aligned (Control: M = 0.906, SE = 0.034, 95% CI [0.838, 0.973]; Experimental: M = 0.885, SE = 0.037, 95% CI [0.811, 0.960]).

d) Lexical Resource

After controlling for pre-test Lexical Resource scores, the group effect did not appear statistically significant (F[1, 55] = 0.410, p = .524, partial η^2 = .007). The adjusted means were closely aligned (Control: M = 0.906, SE = 0.032, 95% CI [0.841, 0.971]; Experimental: M = 0.875, SE = 0.036, 95% CI [0.803, 0.947]).

e) Structural Elements

After controlling for pre-test Structural Elements scores, the ANCOVA yielded a statistically significant group effect at the uncorrected alpha level (F[1, 55] = 4.494, p = .039, partial η^2 = .076). Given a statistically significant Levene's test violation (p < .001), robustness checks were prioritized (see Table 8).

Table 8: Robustness check results for Organization and Structural Elements

Category	Test	Statistic	*df*	*p*	Decision
Organization	Welch's F	5.758	(1, 54.345)	.020	Significant
	Mann-Whitney U	U = 288.500, Z = -2.065	—	.039	Significant
Structural Elements	Welch's F	4.135	(1, 38.569)	.049	Significant
	Mann-Whitney U	U = 331.500, Z = -1.455	—	.146	Non-significant

Note. Welch's F does not assume equal variances. Mann-Whitney U was conducted on gain scores (post-test minus pre-test). Mean ranks for Organization: Control = 33.48, Experimental = 24.60. Mean ranks for Structural Elements: Control = 32.14, Experimental = 26.25. Wilcoxon W: Organization = 639.500; Structural Elements = 682.500.

Welch's *F* test reached significance at the uncorrected level (F [1, 38.569] = 4.135, p = .049), but a Mann-Whitney U test on gain scores did not (U = 331.500, Z = -1.455, p = .146). In line with the pre-specified decision to follow the non-parametric result when Levene's test is significantly violated, the group difference for Structural Elements is not considered statistically significant. The ANCOVA result also did not survive the Bonferroni correction (p = .039 > .010).

f) Overall interpretation

Collectively, the results suggest that the AI-assisted feedback intervention did not produce a statistically significant effect on total EFL writing performance or on any individual writing category after applying the Bonferroni correction. For total scores, both the ANCOVA (F (1, 55) = 0.447, p = .507, partial η^2 = .008) and the non-parametric Mann-Whitney U test (U = 405.000, Z = -0.173, p = .863) converged on a non-significant result, providing some confidence in the robustness of this result.

At the category level, Organization showed a consistent pattern across all three tests conducted (ANCOVA: p = .019; Welch's F: p = .020; Mann-Whitney U: p = .039), with adjusted means indicating that the control group appeared to outperform the experimental group on this category (Control: Adj M = 1.462; Experimental: Adj M = 1.306). This direction is noteworthy in that it runs counter to the expected effect of the intervention, suggesting that conventional instruction may have been associated with relatively stronger Organization scores than AI-assisted feedback during the period studied. This counterintuitive trend warrants further investigation into future studies with larger samples.

Nevertheless, the pattern did not survive the Bonferroni correction applied to control the family-wise error rate across the five category comparisons (Dunn, 1961) and may therefore represent a trend rather than a confirmed effect. For Structural Elements, the ANCOVA and Welch's F test produced a marginal significance at the uncorrected level, but the Mann-Whitney U test did not, and the result similarly did not survive correction. Task Achievement, Grammar, and Lexical Resource showed no evidence of a significant group difference under any analysis.

It is worth noting that the narrow scoring ranges across several categories (Grammar and Lexical Resource: 0 to 1.5 points; Structural Elements: 0 to 1 point) may have constrained score variability and thereby attenuated statistical power, potentially limiting the capacity to detect small but substantively meaningful effects.

4.2 Key Insights from Students' Perceptions through Semi-Structured Interviews

Unlike the quantitative findings, the qualitative data analysis indicated a broadly positive impact of AI tools on students' formal writing development, particularly in terms of clarity, structure, accuracy, and learner awareness. Across responses, students described their pre-AI writing as weak, unclear, and poorly organized, with one student stating that writing "*before using artificial intelligence was weak or almost weak*" and another describing pre-intervention difficulty organizing ideas in "*a professional academic style*".

Post-use reflections emphasize noticeable improvement in coherence, paragraph organization, and professional tone, with one student reporting improvement in "*structuring paragraphs, connecting ideas, and writing in a more formal way*". AI tools played a central role in supporting grammatical accuracy and sentence construction by identifying recurring errors, with one student noting that "*the AI show[ed] me the error and tell[s] what to write instead*", and modeling clearer, more academic sentence forms by targeting specific categories such as "*tenses, articles, subject–verb agreement, and punctuation*", which helped students internalize rules rather than simply accept corrections.

In addition, AI tools enhanced vocabulary choice and register awareness, enabling students to distinguish between informal and formal language and to adopt more appropriate lexical choices for academic assignments and professional emails. A significant impact was also observed in students' revision practices: rather than limiting revision to spelling, students began engaging more deeply with clarity, structure, and organization, with one student explaining that: "*Now I pay more attention to grammar, sentence structure, clarity, and word choice*." This indicates a shift toward a more process-oriented approach to writing.

Importantly, repeated AI feedback increased students' awareness of their common writing errors, fostering metalinguistic awareness and encouraging greater self-monitoring, to the point that one student described becoming "*less dependent on correction and more able to write formal texts by myself*". While some students expressed cautious use of AI and a desire to maintain independent thinking, the overall data suggest that AI tools functioned not merely as corrective aids but as learning supports that scaffolded skills development, promoted confidence, and contributed to long-term improvement in formal writing ability.

5. Discussion

This study examined the impact of QuillBot-assisted instruction on the formal writing proficiency of Omani post-foundation EFL learners. The findings indicate that the use of QuillBot as an AI-assisted tool did not produce any statistically

significant improvement in overall EFL writing performance or any individual writing category after applying the Bonferroni correction. This finding aligns with a growing body of research suggesting that AI intervention does not consistently outperform traditional instruction, as found by Alnemrat et al. (2025) and Shi and Aryadoust (2024) in their respective studies. Although the quantitative analyses revealed only a small, statistically non-significant improvement, the qualitative findings demonstrate meaningful pedagogical value and developmental changes in learners' writing processes.

This finding aligns with that of Chen and Chen's (2025) study, which shows that GenAI-assisted instruction did not significantly improve students' writing performance except in syntactic complexity. However, the interview data indicated that this approach positively influenced students' motivation, engagement, and self-efficacy. This mixed pattern reflects broader research on AI-assisted writing, where deeper learning gains often emerge qualitatively through critical engagement, metacognitive reflection, and improved revision practices rather than immediate score increases.

Given this apparent tension between the null quantitative results and the positive qualitative findings, it is important to distinguish clearly between three separate outcome domains examined in this study: performance outcomes, linguistic outcomes, and affective outcomes. These domains were assessed through different instruments and should not be conflated when interpreting the findings. Performance outcomes refer to the composite writing scores derived from the standardized rubric, including Task Achievement and Structural Elements. Linguistic outcomes refer specifically to Grammar and Lexical Resource, which capture accuracy and range at the sentence level.

Affective outcomes refer to the interview data on learner confidence, anxiety, and motivation. The quantitative analysis addressed only the first two domains, with neither showing a statistically significant group difference. The qualitative interview data addressed the third domain and pointed to a favorable pattern. These are not equivalent forms of evidence, and a positive affective response should not be read as confirmation of measurable gains in writing performance. The two sets of findings are complementary, but they answer different questions and carry different evidential weight.

The qualitative interview data suggests that students engaged more critically with revision and reported greater confidence. However, the small sample of five interview participants limits the generalizability of these patterns, and the positive affective responses reported did not correspond to measurable gains in writing performance. These qualitative patterns suggest engagement and process change but should not be read as evidence of measurable pedagogical improvement. These findings should, therefore, be interpreted alongside the quantitative results, which showed no statistically significant group difference in any writing category.

These findings also suggest that AI intervention, providing instant feedback to students while they work on their drafts, can be considered a complementary tool rather than a full-fledged replacement for traditional instruction. This likely reflects the fact that the effectiveness of AI tools depends heavily on how they are integrated into teaching. While research shows that automated feedback can have a moderate impact on writing performance, results are inconsistent across contexts. Therefore, the lack of significant effects in this study aligns with the findings of Fleckenstein et al. (2023), suggesting that the effectiveness of AI interventions may be context-dependent rather than inherently ineffective.

The analyses of the different writing assessment categories also showed consistent non-significant patterns. However, among all findings reported in this study, the pattern observed for Organization stood out as the most consistent and arguably the most theoretically significant. Across three independent analytical approaches – the ANCOVA, Welch's *F* test, and the Mann-Whitney U test on gain scores – the control group outperformed the experimental group on this category. This pattern persisted, even though the homogeneity of variance assumption was violated and robustness checks were required. Although this finding was not significant after the Bonferroni correction and should be interpreted with caution, it showed a consistent pattern across different statistical analyses. The post hoc sensitivity analysis also suggested that the corrected threshold could detect only large effects. Therefore, this result may reflect a meaningful trend rather than random variation and deserve further theoretical consideration.

This direction is counterintuitive on its face, since AI-assisted feedback tools are often expected to support higher-order writing skills such as coherence and organization alongside surface-level accuracy. However, the finding becomes more interpretable when considered against the specific functional design of QuillBot as a corrective tool operating on student-produced text, as distinguished by Aljuaid (2024) from generative tools that produce text on behalf of the student. QuillBot's core functions – paraphrasing, grammar correction, and tone adjustment – operate primarily at the sentence and word level.

The tool has no mechanism for evaluating or restructuring paragraph-level cohesion, logical sequencing, or argumentative structure, which are the actual constructs measured by the Organization category in this study's rubric. By contrast, the control group received direct teacher feedback and peer review, both of which are pedagogical practices that naturally and explicitly target discourse-level organization. This is because human readers evaluating a draft for clarity inherently attend to how ideas are sequenced and connected, rather than merely determining whether individual sentences are grammatically correct.

This interpretation is consistent with Tran (2025), who similarly found that AI feedback tends to concentrate on surface-level features rather than global discourse-level aspects. It also helps reconcile the apparent tension with Marzuki et al. (2023), who reported that AI tools can support organizational improvement, but specifically when used in combination with teacher feedback rather than as a standalone intervention. In the present study, while teacher guidance was present

throughout the experimental condition, the explicit instructional emphasis described in Section 3.4 centered on AI-supported paraphrasing, grammar correction, and tone, rather than on direct teacher feedback targeting paragraph-level structure. The control group, which did not use AI tools, received comparatively more direct, structure-focused teacher attention because no AI tool assumed part of the feedback role.

Taken together, this pattern suggests a specific and falsifiable hypothesis for future research. AI corrective tools such as QuillBot may inadvertently narrow the locus of feedback toward sentence-level features because they are effective and readily available for that purpose. As a result, they may displace, rather than supplement, the kind of discourse-level feedback typically provided through direct teacher engagement. If accurate, this would imply that the integration of AI tools into writing instruction requires deliberate pedagogical design to preserve, rather than crowd out, teacher attention to higher-order organizational skills. This represents the most theoretical generative finding of the present study and warrants direct investigation in future research.

The qualitative data adds a further layer of complexity to this picture. While the quantitative results indicated that the control group outperformed the experimental group on Organization, at least one interviewed student reported a subjective sense of improved paragraph structuring and idea connection following AI tool use. This divergence may reflect a genuine gap between perceived and measured organizational ability, a pattern noted elsewhere in AI-assisted writing research (Lo et al., 2025), in which students may feel more capable of organizing their writing without this translating into measurable gains in the dimensions assessed by structured rubrics. Alternatively, it may reflect the unrepresentative nature of the interview sample, whose generally positive orientation toward the tool may extend to overestimating its effect on aspects of writing it was not specifically designed to support.

The absence of significant differences across the Grammar, Lexical Resource, and Task Achievement categories contrast with the literature, which emphasizes the strengths of AI tools in improving linguistic accuracy. For instance, Elmotri et al. (2025) reported that automated feedback is particularly effective in enhancing grammar, vocabulary, and sentence-level accuracy through immediate corrective feedback. However, empirical results are not always consistent; as Benali's (2021) study indicates, improvements in these areas can vary depending on learner proficiency, engagement, and feedback quality. One possible explanation lies in the short duration or limited intensity of the intervention, which may not have been sufficient to produce measurable gains.

Additionally, ceiling effects or limited variability in scores may have reduced the ability to detect differences. This suggests that observable improvements in linguistic features may require longer exposure or iterative feedback cycles. The absence of a statistically significant difference in post-test scores suggests that the five-week intervention was insufficient to produce measurable gains. Writing development in EFL contexts typically unfolds gradually and requires repeated

cycles of drafting, feedback, and revision. Students reported perceived improvements in coherence, vocabulary choice, grammatical accuracy, and awareness of formal register, suggesting that AI tool use may have supported the development of formal writing skills at the sentence level.

Students showed positive tendencies toward using QuillBot, reporting reduced anxiety, increased confidence, and a clearer understanding of writing expectations. These affective benefits align with evidence showing that AI tools can support learner autonomy and reduce cognitive load when implemented with guidance. However, affective benefit and writing improvement are distinct constructs, and the positive student perceptions reported here do not constitute evidence of measurable writing gains. Erliani et al. (2025) documented a similar dissociation, which they termed borrowed efficacy: learners' confidence gains during AI-supported writing did not transfer to unaided tasks.

This possibility cannot be ruled out in the present study, since writing performance without AI support was not separately assessed after the intervention. Furthermore, some learners were also unsure why QuillBot made certain suggestions, showing the need for teacher guidance. The uncertainty some students expressed about the rationale behind specific QuillBot suggestions indicates that positive affect toward a tool does not necessarily reflect critical engagement with its output. This distinction between passive and critical AI use carries direct implications for instructional design, which are discussed in the following sections.

6. Pedagogical Implications

The findings of this experimental study suggest that QuillBot is most effective when used as a guided revision tool rather than a substitute for writing instruction. Although no statistically significant improvement was observed in students' overall writing performance, the interview data indicated that students perceived benefits from using QuillBot during the drafting and revising stages, particularly for improving grammatical accuracy, vocabulary use, and sentence clarity.

Therefore, EFL instructors should integrate QuillBot into process-based writing activities and provide explicit training on how to evaluate and selectively apply its suggestions. Encouraging students to reflect on AI-generated revisions can strengthen critical thinking and editing skills while reducing overreliance on automation. A balanced approach combining QuillBot feedback, teacher guidance, and peer review can enhance EFL students' academic writing development and maintain the integrity of learning outcomes at UTAS-A.

7. Conclusion

Although the QuillBot-assisted intervention did not yield statistically significant quantitative improvements, it led to notable qualitative gains. Students developed improved revision practices, stronger critical engagement with language, and greater awareness of formal writing conventions. These outcomes align with evidence that AI tools enhance writing development most effectively when they

promote higher-order thinking, metacognitive monitoring, and reflective learning. Overall, the study supports the conclusion that AI writing tools can meaningfully enhance EFL learners' writing development when integrated into teacher-mediated, critically oriented, and process-driven pedagogical frameworks.

Overall, the findings support a modest and qualified conclusion. QuillBot-assisted instruction did not produce measurable short-term gains in formal writing performance within the five-week study period. However, it was associated with self-reported improvements in learner confidence and engagement, as well as a plausible, theoretically grounded explanation for the unexpected findings related to Organization. These two strands of evidence should be considered separately rather than combined into a single narrative of pedagogical success. Accordingly, broader claims about QuillBot's instructional value should await confirmation through longer intervention periods, larger samples, and research designs that directly assess unaided writing performance.

8. Limitations

Several limitations should be acknowledged to contextualize the findings of the present study. First, the five-week intervention period may have been insufficient to produce significant improvements in complex writing competencies. Second, this experimental study was conducted at a single institution, which limits the generalizability of the findings. Third, the study focused primarily on one AI tool, QuillBot, meaning that the findings may not be generalizable to other AI-assisted tools or platforms.

Additionally, the exclusion of 16 students who were absent for the pre-test (5 from the control group, 11 from the experimental group) reduced the final analytic sample from 74 to 58. Although this missingness occurred prior to treatment exposure and is therefore unlikely to reflect a treatment-related pattern, no retained data was available on the excluded students' characteristics or engagement. Consequently, a formal comparison between completers and non-completers was not possible. Furthermore, the disproportionate exclusion rate between the two sections (5 vs. 11) should be interpreted with appropriate caution.

Relatedly, as no a priori power analysis was conducted before data collection, a post hoc sensitivity analysis was performed to estimate the smallest effect size that the study could reliably detect given its achieved sample size. The analysis indicated that the study, particularly after applying the Bonferroni correction, was sensitive to detect only relatively large effects. Consequently, some non-significant or borderline category-level findings may reflect limited statistical power rather than the true absence of an effect.

Further limitations relate to the scoring of the writing assessments. The same individual served as the course instructor for both sections, designed the instructional intervention, and was the sole scorer of all pre-test and post-test writing samples. No inter-rater reliability check was conducted, and the writing

samples were not scored blind to group membership. This overlap of roles introduces the possibility of conscious or unconscious scoring bias, particularly given the instructor's awareness of each participant's group allocation and personal investment in the success of the intervention. The use of standardized rubric and consistent scoring criteria was intended to minimize subjectivity.

Moreover, although the writing assessment rubric is a university-wide instrument developed by ELT specialists and subject to institutional piloting and periodic revision based on feedback from teachers and coordinators across multiple branches, no formal validation evidence (e.g., inter-rater reliability coefficients or construct validity data) was available for this specific rubric. Its institutional development and ongoing refinement provide some assurance of its practical appropriateness. However, the absence of documented psychometric evidence means that the precision of the rubric could not be independently verified within the scope of this study.

Further limitations relate to the composition of the interview sample. The five interview participants were purposively selected from among the most active users of the AI tool. This was partly because lower-engagement students showed little interest in volunteering when invited and partly because language constraints may have further discouraged participation among less proficient students. Although this reflects a recruitment constraint rather than a deliberate exclusion, the resulting sample was not representative of the experimental group. Consequently, the predominantly positive qualitative findings should be interpreted as reflecting the experiences of highly engaged AI tools rather than those of the experimental group as a whole.

Finally, although weekly observation notes and instructional logs were maintained throughout the intervention, these records were not subjected to formal qualitative coding or systematic analysis. Instead, they served as a contextual record of the instructional process and are not presented as a formally analyzed data source.

9. Recommendations

Based on these findings, the following recommendations are proposed for practice and future research. From a pedagogical perspective, AI tools should be integrated into structured drafting and revision cycles, with students first completing an independent draft before consulting AI-generated feedback. Educators should emphasize critical evaluation of AI outputs to foster metacognitive awareness and higher-order thinking skills. They should also provide explicit guidance to help learners interpret and question AI suggestions. In addition, incorporating reflective writing logs or revision commentaries may encourage students to become more aware of their revision decisions and writing development.

For future research, longer-term intervention studies are recommended to capture development in complex writing constructs. Researchers should include multiple institutions or diverse learner populations to enhance the generalizability of findings. Different AI tools should be compared to understanding how design

features influence learning. Finally, there is a need to examine how varying levels of critical engagement with AI mediate writing outcomes, thereby strengthening the evidence on the role of critical engagement in developing students' evaluative and creative writing skills.

10. Acknowledgments

The authors used Microsoft 365 Copilot primarily for language editing and grammatical refinement during the preparation of this manuscript. The content remains an accurate representation of the author's work and intellectual contributions.

11. References

Abd Rahim, S., & Gregory, N. A. (2025). AI writing tools as catalysts for L2 writing development: A quantitative investigation of students' experiences. *International Journal of Academic Research in Business and Social Sciences*, *15*(9), 350–364. https://doi.org/10.6007/IJARBSS/v15-i9/26490

Ahmadi, A., Samsudin, D., & Ismail, S. F. S. (2025). Writing and artificial intelligence (AI): A systematic literature review of studies (2000–2024). *International Journal of Education*, *18*(1), 117–124. https://vm36.upi.edu/index.php/ije/article/view/76636

Aljuaid, H. (2024). The impact of artificial intelligence tools on academic writing instruction in higher education: A systematic review. *Arab World English Journal*, *Special Issue on ChatGPT*, 26–55. https://doi.org/10.24093/awej/ChatGPT.2

Almashour, M., Aldamen, H. A. K., & Jarrah, M. (2025). Algorithmic feedback and multilingual identity: Translanguaging practices in Jordanian EFL academic writing. *Social Sciences & Humanities Open, 12*, Article 102016. https://doi.org/10.1016/j.ssaho.2025.102016

Alnemrat, A., Aldamen, H., Almashour, M., Al-Deaibes, M., & AlSharefeen, R. (2025). AI vs. teacher feedback on EFL argumentative writing: A quantitative study. *Frontiers in Education, 10*, Article 1614673. https://doi.org/10.3389/feduc.2025.1614673

Al Zakwani, M., & Mathew, B. P. (2025). Exploring student insights on ChatGPT as a resource for learning English and its influence on learner independence. *Forum for Linguistic Studies, 7*(2), 813–824. https://doi.org/10.30564/fls.v7i2.7955

Amani, N., & Bisriyah, M. (2025). University students' perceptions of AI-assisted writing tools in supporting self-regulated writing practices. *Indonesian Journal of English Language Teaching and Applied Linguistics, 10*(1), 91–107. https://doi.org/10.21093/ijeltal.v10i1.1942

Apriani, E., Cardoso, L., Obaid, A. J., Muthmainnah, Wijayanti, E., Esmianti, F., & Supardan, D. (2024). Impact of AI-powered chatbots on EFL students' writing skills, self-efficacy, and self-regulation: A mixed-methods study. *Global Educational Research Review*, *1*(2), 57–72. https://doi.org/10.71380/GERR-08-2024-8

Begmatova, K., & Saydazimova, I. (2026). Enhancing learners' academic writing skills: A comparative analysis of traditional and AI-assisted instruction approaches. *Applied Corpus Linguistics, 6*(1), Article 100176. https://doi.org/10.1016/j.acorp.2025.100176

Benali, A. (2021). The impact of using automated writing feedback in ESL/EFL classroom contexts. *English Language Teaching, 14*(12), 189–195. https://doi.org/10.5539/elt.v14n12p189

Binu, P. M. (2022). The effect of online interaction via Microsoft Teams private chat on enhancing the communicative competence of introverted students. *Arab World English Journal*, *13*(4), 106–114. https://doi.org/10.24093/awej/vol13no4.8

Binu, P. M. (2024). Affordances and challenges of integrating artificial intelligence into English language education: A critical analysis. *English Scholarship Beyond Borders, 10*(1), 34–51. https://www.scopus.com/inward/record.url?eid=2-s2.0-85197370526&partnerID=MN8TOARS

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bui, T. T. U., & Tong, T. V. A. (2025). The impact of AI writing tools on academic integrity: Unveiling English-majored students' perceptions and practical solutions. *Asia CALL Online Journal*, *16*(1), 83–110. https://www.researchgate.net/publication/388445282

Chen, C., & Gong, Y. (2025). The role of AI-assisted learning in academic writing: A mixed-methods study on Chinese as a second language students. *Education Sciences*, *15*(2), Article 141. https://doi.org/10.3390/educsci15020141

Chen, Y., & Chen, J. (2025). Generative AI in EFL writing instruction for lower-intermediate learners: Do perceived affective gains translate to improved performance? *Language Teaching Research.* https://doi.org/10.1177/13621688251397724

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cui, Y., He, W., Du, X., Zeng, M., & Liu, D. (2025). The impact of AI writing assistants on academic writing performance: From the perspective of subjectivity. *International Journal of Distance Education Technologies*, *23*(1), 1–28. https://doi.org/10.4018/IJDET.391326

Deep, P. D., & Chen, Y. (2025). The role of AI in academic writing: Impacts on writing skills, critical thinking, and integrity in higher education. *Societies*, *15*(9), Article 247. https://doi.org/10.3390/soc15090247

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association, 56*(293), 52–64. https://doi.org/10.1080/01621459.1961.10482090

Ekizoğlu, M., & Demir, A. N. (2025). The role of AI assisted writing feedback in developing secondary students' writing skills. *Discover Education*, *4*, Article 454. https://doi.org/10.1007/s44217-025-00919-3

Elmotri, B., Harizi, R., Boujlida, A., Elyasa, Y. M., Garrouri, S., Amri, F., Malik, F. H., & Al-humari, M. A. (2025). The impact of AI-generated feedback explicitness. *Arab World English Journal, 16*(1), 384–402. https://doi.org/10.24093/awej/vol16no1.24

Erliani, N. N. P., Eryansyah, & Amrullah. (2025). The impact of AI tools on EFL students' self-efficacy in academic writing: A qualitative study. *Edukasi: Jurnal Pendidikan dan Pengajaran*, *12*(1), 6–17. https://doi.org/10.19109/s75mbm13

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). SAGE Publications.

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence, 6*, Article 1162454. https://doi.org/10.3389/frai.2023.1162454

Giray, L., Sevnarayan, K., & Ranjbaran Madiseh, F. (2025). Beyond policing: AI writing detection tools, trust, academic integrity, and their implications for college

writing. *Internet Reference Services Quarterly, 29*(1), 83-116. https://doi.org/10.1080/10875301.2024.2437174

IBM Corp. (2020). *IBM SPSS Statistics for Windows* (Version 27.0) [Computer software].

Khattak, Z. I. & Mathew, B. P. (2024). Exploring the effects of Microsoft Teams Messaging app on post foundation students' writing skills: A socio-constructivist analysis. *English Language Teaching, 17*(2), 51–57. https://doi.org/10.5539/elt.v17n2p51

Lim, J. H., Yunus, M. M., & Wong, W. L. (2026). The impact of artificial intelligence writing tools on learners' motivation in English writing: A systematic review. *International Journal of Learning, Teaching and Educational Research*, *25*(1), 695–713. https://doi.org/10.26803/ijlter.25.1.33

Lo, N., Wong, A., & Chan, S. (2025). The impact of generative AI on essay revisions and student engagement. *Computers and Education Open*, *9*, Article 100249. https://doi.org/10.1016/j.caeo.2025.100249

Luo, Y., Xia, Y., & Lu, X. (2025). A systematic review of the role of artificial intelligence in second language writing education. *Digital Studies in Language and Literature*, *2*(2), 302–325. https://doi.org/10.1515/dsll-2025-0001

Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, *10*(2), Article 2236469. https://doi.org/10.1080/2331186X.2023.2236469

Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. *Discover Education*, *4*, Article 170. https://doi.org/10.1007/s44217-025-00602-7

Mhamdi, N. A. (2026). AI-mediated feedback in English learning: Moroccan graduate students' accounts of motivation, autonomy, and trust conditions. *Computers and Education: Open, 10,* Article 100356. https://doi.org/10.1016/j.caeo.2026.100356

Mohammed, G. M. S. (2024). Exploring the Saudi EFL learners' perceptions and challenges in AI-assisted post-editing tools for writing. *Pakistan Journal of Life and Social Sciences*, *22*(2), 13802-13814. https://doi.org/10.57239/PJLSS-2024S-22.2.00991

Neiroukh, S., Mathew, B. P., & Ghorbanpoor, M. (2024). Investigating the impact of a comprehensive writing feedback guide on enhancing learner autonomy. *Journal for Foreign Languages*, *16*(1), 337–363. https://doi.org/10.4312/vestnik.16.337-363

Nhan, L. K., Hoa, N. T. M., & Quang, L. V. N. (2025). Leveraging AI for writing instruction in EFL classrooms: Opportunities and challenges. *Educational Process: International Journal*, *15*, e2025158. https://doi.org/10.22521/edupij.2025.15.158

Ozfidan, B., El-Dakhs, D. A. S., & Alsalim, L. A. (2024). The use of AI tools in English academic writing by Saudi undergraduates. *Contemporary Educational Technology*, *16*(4), ep527. https://doi.org/10.30935/cedtech/15013

Ren, L., Stephens, J. M., & Lee, K. (2026). The impact of AI on learners' self-efficacy: A meta-analysis. *Behavioral Sciences*, *16*(1), Article 158. https://doi.org/10.3390/bs16010158

Selim, A. S. M. (2024). The transformative impact of AI-powered tools on academic writing: Perspectives of EFL university students. *International Journal of English Linguistics*, *14*(1), 14–29. https://doi.org/10.5539/ijel.v14n1p14

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, *36*(2), 187–209. https://doi.org/10.1017/S0958344023000265

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*, Article 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Syafei, M., & Nuraeningsih. (2025). Leveraging artificial intelligence in writing: ELT students' perspectives and experiences. *International Journal of Learning, Teaching and Educational Research*, *24*(5), 416–430. https://doi.org/10.26803/ijlter.24.5.22

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

Teng, M. F., & Huang, J. (2025). Incorporating ChatGPT for EFL writing and its effects on writing engagement. *International Journal of Computer-Assisted Language Learning and Teaching*, *15*(1), 1–21. https://doi.org/10.4018/IJCALLT.367874

Tran, T. T. T. (2025). Enhancing EFL writing revision practices: The impact of AI- and teacher-generated feedback and their sequences. *Education Sciences*, *15*(2), Article 232. https://doi.org/10.3390/educsci15020232

Tuong, T.-P.-L., & Tran, T. T. (2025). The differential impact of AI tools among EFL university learners: A process writing approach. *International Journal of Learning, Teaching and Educational Research*, *24*(5), 452–471. https://doi.org/10.26803/ijlter.24.5.24

Wang, C., Li, Z., & Bonk, C. (2024). Understanding self-directed learning in AI-assisted writing: A mixed methods study of postsecondary learners. *Computers and Education: Artificial Intelligence*, *6*, Article 100247. https://doi.org/10.1016/j.caeai.2024.100247

Zamorano, C. (2025). Enhancing academic writing in English language education through generative AI integration. *Research Studies in English Language Teaching and Learning*, *3*(3), 424–447. https://doi.org/10.62583/rseltl.v3i3.87

Zare, J., Al-Issa, A., & Ranjbaran Madiseh, F. (2025). Interacting with ChatGPT in essay writing: A study of L2 learners' task motivation. *ReCALL*, *37*(3), 385–402. https://doi.org/10.1017/S0958344025000035

Appendix 1: Pre-Test

جامعة التقنية والعلوم التطبيقية
University of Technology and Applied Sciences
المصنعة
Al-Mussanah

PREPARATORY STUDIES CENTER
Fall Semester, AY 2025-2026
Technical Writing (UNEN1203)

Student name: _____ ID: _____ Sec: _____

Sending Emails

Read the following scenarios and write an email using all the functions given.

A | Imagine you are an Alumni from UTAS-A. You have recently seen a poster on Instagram inviting you for an Alumni meet. You would like to know more details about the gathering as they are not mentioned in the invitation. You must write an email to the Student Support Officer, UTAS-A, Mr. Mazin Al Subhi, mazins@utasa.edu.om, and Alumni office, alumni.utasa@edu.om. Send a blind carbon copy to your personal email. Send an email that includes the following points:

- **enquire: location**
- **request: any family member allowed**
- **ask: cultural programs**
- **inform: interest in addressing the audience**
- **attach: CV**
- **enquire: program sheet**

_____

_____

_____

Appendix 2: Post-Test

University of Technology and Applied Sciences
Preparatory Studies Centre
Technical Writing
Mid-Term Question Paper
Spring Semester, 2025-2026

Question 1: Sending Emails **10 Marks**

Instructions: Read the following scenario and write an email using all the given functions.

Imagine you are Thuraya Al-Baluchi, the International Relations Coordinator at Sultan Qaboos University in Muscat. Write an email to all students in the university at all@squ.edu.om about the student exchange program. Send a carbon copy to Maryam Al-Siyabi, the Academic Registrar, at ms@squ.edu.om. Additionally, send a blind carbon copy to Dr. Badr Al-Kindi, the Dean, at bk@squ.edu.om.

Write an email that includes the following points:

- Announce: student exchange program / Vniversity of London
- Arrange: workshop / exchange program (give time and date)
- Attach: exchange program application forms / Word document
- Request: complete application forms / deadline 25th April
- Complain: incomplete application forms / last year
- Inform: office hours for questions / Mondays

Appendix 3: Standardized Departmental Rubric

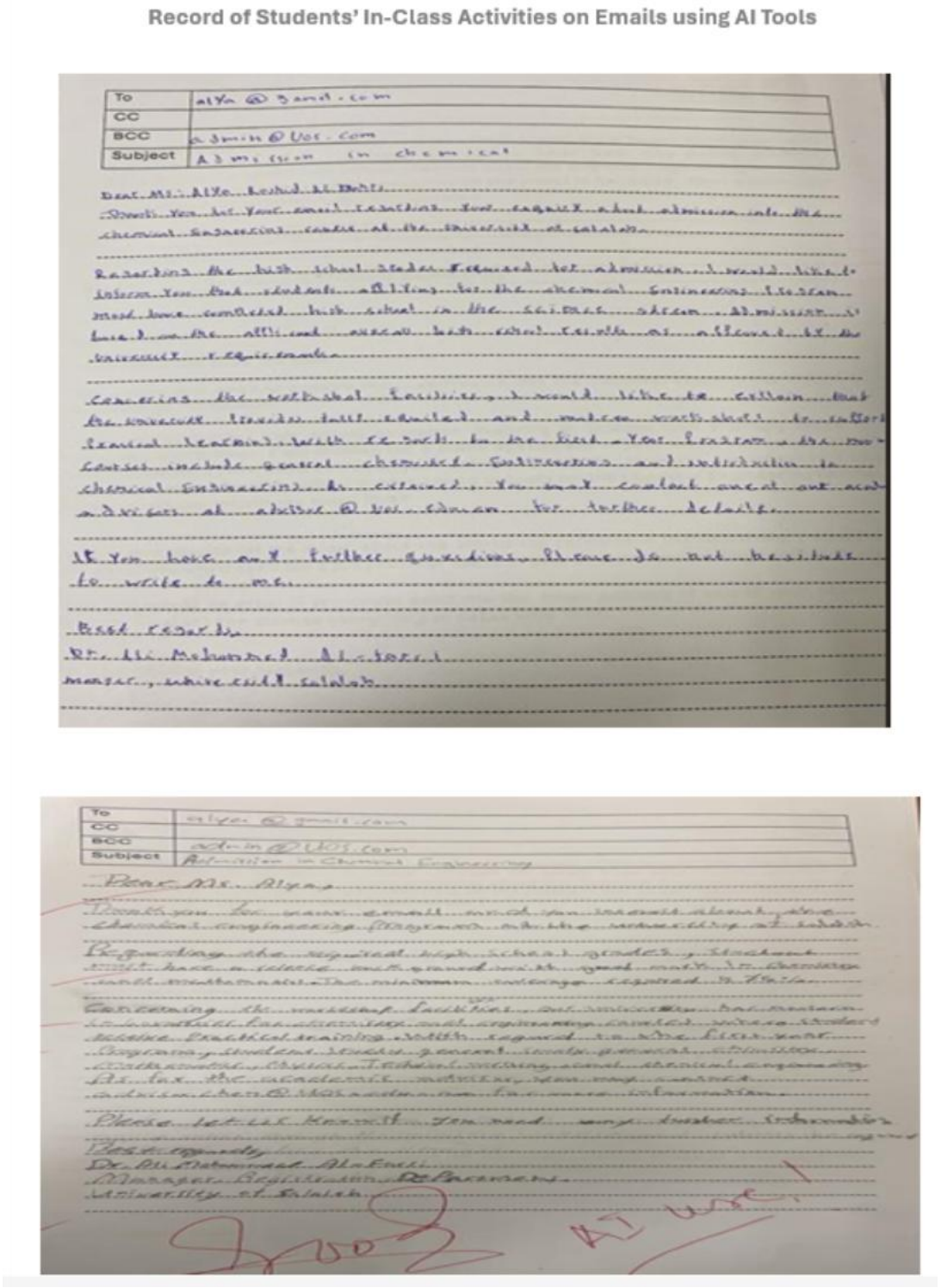
Email Marking Rubric					
Task Achievement **4 Marks**	Responds to **less than 25%** of the requirements; minimal comprehension. Very few or no logical connections between ideas. **1 Mark**	Responds to **25%–50%** of the requirements; **limited** comprehension; some logical connections; weak overall structure. **2 Marks**	Responds to **50%–75%** of the requirements; **moderate** comprehension; strong logical connections; clear and mostly consistent structure. **3 Marks**	Responds to **100%** of the requirements; **excellent** comprehension; very strong logical connections; clear, consistent, and well-organized structure. **4 marks**	**4**
Organization **2 Marks**	No cohesion in presenting the points of the email **0.5 Marks**		Cohesion is not always clear in presenting the points or sequence/transition is not easy to follow. **1 Mark**	Clear cohesion and transition between and within the points. **2 Marks**	**2**
Grammar **1.5 Marks**	**Frequent** grammatical errors. **0.5 Marks**		**Occasional** grammatical errors. **1 Mark**	**Almost no** grammatical errors. **1.5 Marks**	**1.5**
Vocabulary **1.5 Marks**	Word choice, spelling and/or punctuation are **sometimes inaccurate**. **0.5 Marks**		Word choice, spelling and/or punctuation are **mostly accurate**. **1 Mark**	Word choice, spelling and/or punctuation are **consistently accurate**. **1.5 Marks**	**1.5**
Structural Elements address boxes subject line salutation signing off signature **1 Mark**	Missing or incorrect **3 or more** elements. **0 Marks**		Missing or incorrect **1 or 2** elements. **0.5 Marks**	**All** elements are correctly included. **1 mark**	**1**
Total					**10**

Appendix 4: Semi-Structured Interview Guide

Student Perceptions of AI Tools Effectiveness for Formal Writing Skills **الذكاء أدوات فاعلية حول الطلبة تصوّرات الرسمية الكتابة مهارات تحسين في الاصطناعي**

1. *How would you describe your formal writing ability before using AI tools compared to after using them in this course?* **بعد بقدرتك مقارنة الاصطناعي الذكاء أدوات استخدام قبل الرسمية الكتابة على قدرتك تصف كيف المقرر؟ هذا في استخدامها**
2. *In what ways, if any, did AI tools help you improve your grammar and sentence structure in formal writing?* *الاصطناعي الذكاء أدوات استخدام نتيجة تحسّنت أنها شعرتم التي الرسمية كتابتكم في المحددة الجوانب ما (الوضوح أو الجملة، بناء القواعد، مثل)؟*
3. *How did using AI tools affect your vocabulary choice and level of formality in academic or professional writing tasks?* **في الرسمية ومستوى للمفردات اختياركم على الاصطناعي الذكاء أدوات استخدام أثّر كيف الرسمية؟ الرسائل كتابة مثل المهنية، أو الأكاديمية الكتابة**
4. *How did AI tools influence the way you revised and edited your writing drafts?* **أدوات أثّرت طريقة بأي تسليمها؟ قبل الكتابة مسودات وتحرير مراجعة في أسلوبكم في الاصطناعي الذكاء**
5. *Did using AI tools make you more aware of your common writing errors? Why or why not?* **ساعدكم هل أمكن إن أمثلة مع ذلك توضيح يرجى كتاباتكم؟ في الشائعة بالأخطاء وعيكم زيادة على الاصطناعي الذكاء أدوات استخدام**
6. *Do you feel that AI tools helped you learn how to write better, or did they mainly help you correct mistakes? Please explain.* **اقتصرت أم أفضل، بشكل الكتابة كيفية تعلّم على الاصطناعي الذكاء أدوات ساعدتكم هل برأيكم، ولماذا؟ فقط؟ الأخطاء تصحيح على قائدتها**

Appendix 5: In-Class Writing Output, Experimental Group



Appendix 6: Full Assumption Check Results

Statistical Assumption Checks

This appendix reports the full results of the statistical assumption checks conducted prior to the main ANCOVA analyses. Checks were performed for pre-test equivalence between groups, homogeneity of regression slopes, linearity of the covariate-outcome relationship, homogeneity of error variance (Levene's test) for total scores and all five writing categories, and normality of residuals and gain scores (Shapiro-Wilk test). Where assumptions were violated, appropriate corrective analyses were conducted in the main results section.

Pre-test Equivalence. An independent samples t-test confirmed that the control group (M = 6.7266, SD = 1.34158, n = 32) and the experimental group (M = 6.6635, SD = 1.35820, n = 26) were statistically equivalent on total writing scores prior to the intervention, t(56) = 0.177, p = .860, Cohen's d = 0.047. Equal variances were assumed (Levene's F(1, 56) = 0.037, p = .847). Table A1 presents these findings.

Table A1: Pre-test total writing score equivalence between groups

Group	n	M	SD	t(56)	p
Control	32	6.7266	1.34158	.177	.860
Experimental	26	6.6635	1.35820		

Note. Cohen's d = 0.047. Equal variances assumed (Levene's F(1, 56) = 0.037, p = .847).

Homogeneity of Regression Slopes. A preliminary ANCOVA model incorporating a Group x Pre-test interaction term was estimated for total writing scores. The interaction was non-significant, F(1, 54) = 2.055, p = .158, indicating that the regression slopes were reasonably consistent across the two groups. This assumption was considered satisfied.

Linearity. A scatterplot of pre-test against post-test total scores, disaggregated by group, indicated a weak but positive linear trend in both groups. The narrow overall score range likely attenuated the strength of the trend. No non-linear pattern was apparent, and the assumption was considered marginally satisfied.

Homogeneity of Error Variance and Normality. Table A2 presents the full results of Levene's test for all five categories and total scores, as well as Shapiro-Wilk results for ANCOVA residuals and gain score distributions. Levene's test was violated for Organization (p = .043) and Structural Elements (p < .001). Residual normality was violated for total scores (W(58) = 0.937, p = .005), with the distribution exhibiting negative skewness (skewness = -1.109, SE = 0.314) and positive excess kurtosis (kurtosis = 2.030, SE = 0.618). Gain score distributions were approximately normal for both groups (Control: W(32) = 0.952, p = .166; Experimental: W(26) = 0.927, p = .068). Given that the ANCOVA F test has been argued to show reasonable robustness to moderate violations of normality when group sizes are adequate (Glass et al., 1972; Harwell et al., 1992), the main analyses proceeded, with non-parametric and robust alternatives applied where relevant.

Table A2: Summary of statistical assumption check results

Variable	Assumption	Test / Method	Result	Decision
Total Score	Homogeneity of Regression Slopes	Group x Pre-test interaction	$F(1, 54) = 2.055$, $p = .158$	Satisfied
Total Score	Linearity	Scatterplot inspection	Weak positive trend	Marginally satisfied
Total Score	Homogeneity of Error Variance	Levene's test	$F(1, 56) = 0.870$, $p = .355$	Satisfied
Task Achievement	Homogeneity of Error Variance	Levene's test	$F(1, 56) = 2.777$, $p = .101$	Satisfied
Organization	Homogeneity of Error Variance	Levene's test	$F(1, 56) = 4.308$, $p = .043$	Violated — Welch's F and Mann-Whitney U conducted
Grammar	Homogeneity of Error Variance	Levene's test	$F(1, 56) = 0.382$, $p = .539$	Satisfied
Lexical Resource	Homogeneity of Error Variance	Levene's test	$F(1, 56) = 0.829$, $p = .367$	Satisfied
Structural Elements	Homogeneity of Error Variance	Levene's test	$F(1, 56) = 16.702$, $p < .001$	Violated — Welch's F and Mann-Whitney U conducted
Total Score Residuals	Normality of Residuals	Shapiro-Wilk	$W(58) = 0.937$, $p = .005$	Violated — Mann-Whitney U conducted as robustness check
Total Gain Score (Control)	Normality of Gain Scores	Shapiro-Wilk	$W(32) = 0.952$, $p = .166$	Satisfied
Total Gain Score (Experimental)	Normality of Gain Scores	Shapiro-Wilk	$W(26) = 0.927$, $p = .068$	Satisfied

Note. Levene's test assesses homogeneity of error variance across groups. Shapiro-Wilk tests normality of the specified distribution. W = Shapiro-Wilk statistics. F = Levene's or interaction F statistic as applicable.