

# Using AI Applications to Enhance EFL Learners' Writing Skills: Empirical Evidence from “Monica”

Muhammed Parviz\* 

*Assistant Professor of TEFL, Imam Ali University, Tehran, Iran*

Masoud Azizi 

*Assistant Professor of TEFL, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran*

**Received:** October 11, 2025; **Revised:** November 26, 2025; **Accepted:** December 09, 2025

## Abstract

Advances in generative artificial intelligence (AI), particularly large language models, offer new possibilities for language learning. This study, grounded in sociocultural theory, examined the impact of an AI assistant (*Monica*) on English as a Foreign Language (EFL) students' writing development. A total of 120 male students (aged 26–36, A2 proficiency) participated in a four-month in-service program. The participants were randomly assigned to either an experimental group, which engaged in AI-mediated collaborative writing, or a control group, which collaborated face-to-face. A sequential explanatory mixed-methods design was employed, integrating pretest–posttest measures, questionnaires and semi-structured interviews. Quantitative analyses (Split-Plot ANOVA) showed that the AI-assisted group significantly outperformed the control group in Organization, Language Use, and Mechanics, but not in Content or Vocabulary. Importantly, students' perceptions shifted dramatically: While only 23% initially held positive expectations, 90% reported positive experiences post-intervention, highlighting the role of affective factors in technology adoption. Qualitative data confirmed cognitive, metacognitive, and motivational benefits of AI interaction, but also revealed challenges including overreliance on AI, uncritical acceptance of feedback, difficulty integrating new vocabulary, and cognitive overload when faced with multiple corrections. Sociocultural concerns, such as academic integrity, face-saving practices, and peer judgment, further influenced student engagement. From a theoretical standpoint, the study contributes to sociocultural scholarship by evidencing the ways in which student–AI interaction shapes and mediates writing development. Pedagogically, it calls for intentional scaffolding to foster critical and reflective practices, reframing AI not as a functional tool but as a meaningful collaborator in the learning process.

**Keywords:** AI; Monica; EFL Writing; Sociocultural Theory; Split-Plot ANOVA; Analytic Scoring Rubric

---

\*Corresponding author's email: mohammad.parviz60@gmail.com

## INTRODUCTION

Artificial intelligence (AI) is increasingly being integrated into education, providing new ways for students to engage with content and strengthen their academic skills. A particularly promising application is in language learning, where AI-powered writing tools such as grammar checkers, automated writing assistants, paraphrasing and text generators are enhancing teaching English as a Foreign Language (EFL) instruction (Jmaiel et al., 2025; Marzuki et al., 2023). Driven by natural language processing and machine learning algorithms, these tools analyze multiple aspects of written composition, from basic grammar and vocabulary usage to more complex elements like syntax, content, organization, and structural coherence (Song & Song, 2023). By comparing students' texts to large corpora of proficient writing, they deliver formative feedback that not only improves linguistic accuracy but also supports metacognitive engagement with writing strategies (Nazari et al., 2021). These affordances are particularly salient in EFL contexts, where students often struggle with limited linguistic resources, writing anxiety, and the cognitive burden of producing coherent and well-structured texts (Gayed et al., 2022; Nazari et al., 2021). Immediate, tailored feedback helps students navigate these challenges and redirect attention toward higher-order writing skills such as idea development and organization. Empirical evidence suggests that such support enhances autonomy, mitigates anxiety, and fosters more effective and confident writing practices (Ebadi & Amini, 2022; Marzuki et al., 2023; Tran, 2025).

Despite its potential, the integration of AI writing tools into classroom practice remains uneven globally. Research, to date, has predominately examined technologically advanced contexts, where reliable infrastructure, institutional support, and professional training have facilitated adoption. In contrast, under-resourced settings often face structural and sociopolitical constraints that restrict systemic integration. Students in these environments may depend on outdated or limited-functionality applications, while unreliable internet connectivity further impedes sustained engagement

(Moorhouse & Wong, 2025). This challenge is particularly pronounced in Global South countries such as Iran, where AI adoption in education remains nascent despite growing interest in digital innovation. Iran represents a compelling context for examining AI-assisted writing for three reasons. First, EFL instruction is traditionally shaped by teacher-centered pedagogies, non-interactiveness of students, cultural norms of authority, exam-oriented learning and large class sizes. This raises questions about whether AI tools can mediate shifts toward more learner-centered approaches. Nevertheless, private language institutes offer a different instructional environment. In these settings, teaching tends to be more student-centered, with smaller class sizes that allow for greater interaction. Instructors take on facilitative roles, encouraging active participation from learners in communicative and interactive tasks.

Second, Iranian EFL students exhibit diverse proficiency levels, cognitive skills, and prior exposure to digital tools, offering opportunities to investigate differentiated AI interventions. Third, persistent resource constraints-including limited access to technologies and uneven teacher training-provide a realistic setting to assess the feasibility and pedagogical adaptability of AI-assisted writing. Examining this context can thus illuminate both the opportunities and limitations of generative AI (GenAI) in under-resourced EFL environments and contribute to equitable, context-sensitive approaches to language instruction.

To address this gap, the present study investigated the pedagogical role of *Monica*, an AI-powered writing assistant, in fostering Iranian EFL students' writing proficiency. Grounded in sociocultural theory (SCT), the research examines whether AI-mediated instruction can enhance writing skills more effectively than traditional face-to-face collaborative methods. By analyzing student interaction with AI tools in authentic classroom settings, this study seeks to generate empirical insights into how GenAI can be pedagogically and contextually adapted to support writing development in under-resourced environments. This study contributes to the field in three key ways: (a) by broadening the geographical scope of AI-assisted writing

research, (b) by providing empirical evidence on an alternative AI writing tool (*Monica*), and (c) by examining AI-student interactions through the lens of SCT, specifically within students' zones of proximal development (ZPD). Accordingly, the study pursues the following research questions:

- (1) To what extent does AI-mediated writing instruction using *Monica* enhance EFL students' writing proficiency compared to traditional face-to-face collaborative methods?
- (2) What are the key benefits, challenges and limitations of integrating *Monica* into EFL writing instruction from the students' perspective?

## LITERATURE REVIEW

The integration of technology into second language (L2) writing instruction has evolved through successive phases shaped by pedagogical paradigms and technological affordances. Early research in the 2000s established the efficacy of computer-assisted writing tools such as electronic feedback (Tuzi, 2004). With the advent of Web 2.0 technologies, the focus shifted toward collaboration and social interaction. Studies explored wiki-mediated writing (Rahimi & Fathi, 2021), social media-mediated instruction (Lee, 2020) and meta-analytic findings confirmed strong effects for socially mediated writing (Seyyedrezaei et al., 2022). However, much of this research remained descriptive, offering limited theoretical insight into the mechanisms of technology-supported collaboration. A third phase has emerged with AI technologies, representing a paradigm shift in how writing support is conceptualized and delivered. Early applications such as *Grammarly* (Koltovskaia, 2020) emphasized automated feedback for grammar, syntax, and paraphrasing. More recent GenAI tools, particularly ChatGPT, extend this support to higher-order concerns including discourse organization, lexical refinement, and rhetorical enhancement (Song & Song, 2023). Beyond improving textual quality, these tools have been shown to promote learner autonomy (Tran, 2025) and expanded writing practices beyond traditional classroom boundaries.

Empirical evidence further highlights the pedagogical value of GenAI while also pointing to critical limitations. Guo et al. (2022) found that chatbot-based scaffolding positively supported students' writing development. Similarly, Yan (2023) explored the use of ChatGPT in a one-week L2 writing practicum, identifying three key aspects of its role: (1) clear affordances for enhancing writing efficiency through automated workflows, (2) significant potential for broader pedagogical applications, and (3) concerns among participants regarding academic honesty and educational equity. While Yan's qualitative analysis of learner behaviors and reflections provides valuable preliminary insights into student perceptions, the study's short-term design and reliance on self-reported data limit generalizability. These limitations underscore the need for longitudinal, mixed-methods research to fully assess AI's educational impact. Complementing these findings, Liu et al. (2024) compared argumentative compositions written with and without AI assistance and found that AI-supported texts demonstrated enhanced structural organization, linguistic sophistication, coherence, and lexical diversity. However, critical limitations were identified, particularly in the depth of critical analysis and originality, highlighting the importance of integrating AI tools through pedagogically sound frameworks that balance efficiency and higher-order thinking. Similarly, Lin and Crosthwaite (2024) compared teacher-provided written corrective feedback (WCF) with that generated by ChatGPT. While teachers offered a balance of direct and indirect feedback on both local and global issues, ChatGPT mainly produced metalinguistic comments or reformulations. Although grammatically accurate, its feedback was often inconsistent, redundant, and overly grammar-focused. Importantly, their analysis centered on the product of WCF rather than learner uptake, leaving unresolved questions about reliability, pedagogical effectiveness, and ethical implications.

Across the broader literature, four consistent limitations can be identified. First, a geographical bias favors technologically advanced contexts, leaving underrepresented regions underexamined. Second, research has primarily focused on mainstream AI tools such as ChatGPT, with limited

attention given to alternative AI-assisted writing applications. Third, few studies employ sociocultural or other theoretical frameworks to systematically compare AI-mediated and traditional collaborative writing, thereby missing opportunities to theorize human-AI interaction. Fourth, emerging educational contexts such as Iran remain underexplored, despite differences in infrastructure, cultural norms, and implementation challenges (Parviz, 2024). Collectively, these gaps underscore the need for contextually sensitive, theoretically informed, and geographically diverse investigations of AI-assisted writing in EFL education.

Within Iran, emerging studies suggest both the potential of AI tools and persistent research limitations. Borna et al. (2024) reported improvements using *Grammarly* and *ProWritingAid*; however, reliance on post-test quantitative data precluded analysis of learner engagement, interaction, and discrete writing subskills. Similarly, Teimourtash (2024) compared Synthetic and Analytic AI tools, yet conceptual ambiguity and absence of process data obscured which affordances contributed to observed improvements. Ghafouri et al. (2024) also implemented a ten-week ChatGPT protocol and found gains in learner performance and teacher self-efficacy; nevertheless, the small sample (12 teachers, 48 learners) and purely quantitative design limited generalizability and interpretive depth, leaving contextual implementation challenges unexamined. Fathi and Rahimi (2024) employed a Vygotskian sociocultural framework using microgenetic analysis and think-aloud protocols, demonstrating cognitive and affective benefits; however, the limited sample ( $n = 14$ ) and short duration reduced external validity, and emerging concerns such as over-reliance on AI or erroneous feedback (Parviz, 2024) were not addressed. Finally, Fereidouni and Farahian (2024) found that combining AI with teacher mediation produced the strongest outcomes, yet the absence of a guiding theoretical framework weakened explanatory power.

Taken together, prior studies indicate that AI tools can enhance L2 writing performance among Iranian EFL learners, particularly when paired with human mediation. Nevertheless, methodological and conceptual

limitations remain: small, non-representative samples; limited qualitative and process-oriented data; underdeveloped theoretical frameworks; and insufficient focus on discrete writing subskills. These gaps highlight the need for mixed-method, theory-driven research capturing both outcomes and learning processes. Table 1 presents an overview of AI-assisted writing studies in EFL contexts, showing key findings, contexts, and limitations by Global and Iranian research.

**Table 1.** Empirical Studies on AI-Assisted Writing in EFL Contexts: Global and Iranian Research

**A. Global Studies**

Study	Context / Participants	AI Tool(s)	Focus / Design	Key Findings	Key Limitations	Identified Gaps
Guo et al. (2022)	EFL learners	Chatbot-based system	Experimental	AI scaffolding positively supported writing development	Limited analysis of learning processes and long-term impact	Need for longitudinal, process-oriented research
Yan (2023)	L2 learners; 1-week practicum	ChatGPT	Qualitative	Improved efficiency; pedagogical potential; concerns over academic honesty and equity	Short duration; self-reported data	Need for mixed-methods and sustained classroom studies
Liu et al. (2024)	EFL learners	AI writing assistant	Comparative	Enhanced organization, coherence, linguistic sophistication, lexical diversity	Limited gains in critical depth and originality	Need to examine higher-order writing and critical thinking
Lin & Crosthwaite (2024)	EFL context	ChatGPT vs. teacher WCF	Product-based comparison	Teachers balanced global/local feedback; AI feedback grammar-heavy and redundant	Learner uptake and retention not examined	Need to study feedback uptake, retention, and pedagogical effectiveness

**B. Iranian Studies**

Study	Context / Participants	AI Tool(s)	Focus / Design	Key Findings	Key Limitations	Identified Gaps
Borna et al. (2024)	Iranian EFL learners	Grammarly, ProWriting Aid	Quantitative	Post-test improvements in writing performance	No process or qualitative data	Need to examine learner engagement and writing subskills
Teimourta sh (2024)	Iranian EFL learners	Synthetic vs. Analytic AI	Comparative	Overall writing improvement observed	Conceptual ambiguity; no process data	Need for theory-driven analysis of AI affordances
Ghafouri et al. (2024)	Iran; 12 teachers, 48 learners	ChatGPT	Quantitative intervention	Gains in learner performance and teacher self-efficacy	Small sample; lack of qualitative data	Need for larger-scale, mixed-methods research
Fathi & Rahimi (2024)	Iran; n = 14	AI-assisted tool	Sociocultural, microgenetic	Cognitive and affective benefits; ZPD-based mediation	Small sample; short duration	Need for scalable sociocultural investigations
Fereidouni & Farahian (2024)	Iranian EFL learners	AI + teacher mediation	Experimental	Combined AI–teacher mediation most effective	Weak theoretical grounding	Need for explicit theoretical framework

**PURPOSE OF THE STUDY**

Addressing these gaps, the present study examines *Monica*, an AI-powered writing assistant, in Iranian EFL classrooms. Grounded in Vygotsky's (1978) SCT, the study conceptualizes writing development as a socially mediated

process facilitated by cultural tools. Writing tasks which involve inherently dialogic processes such as planning, drafting, and revising offer an ideal context for examining the internalization of knowledge through interaction with GenAI tools. *Monica* provides immediate, individualized scaffolding aligned with learners' ZPD, functioning as both a mediator and interactive partner. In contemporary learning environments, such AI applications are increasingly regarded as legitimate "more knowledgeable others" (e.g., teachers or peers) that extend learner capabilities (Song & Song, 2023; Vygotsky, 1978). Within this role, *Monica* operates as a writing coach, guiding L2 writers through dialogic exchanges in which feedback is internalized and gradually transformed into greater autonomy in independent writing.

*Monica* operates across three pedagogically significant dimensions: facilitating low-anxiety linguistic interaction, providing adaptive feedback on linguistic, strategic, and cultural aspects of writing, and fostering engagement through contextually relevant, interest-driven dialogues (Song & Song, 2023). These affordances position *Monica* not merely as a tool but as an interactive participant in writing development, contributing to hybrid human-AI social-cognitive spaces. However, challenges such as over-reliance, inaccurate feedback, and ethical concerns persist (Parviz, 2024). Framing *Monica* within a sociocultural lens allows for a nuanced understanding of how students negotiate cognitive, affective, and contextual dimensions of AI-mediated writing. This study thus offers context-sensitive insights into the mechanisms by which AI can support, extend, and reshape L2 writing instruction.

## METHOD

### Research Design

This study employed a sequential explanatory mixed-methods approach to investigate the comparative effectiveness of AI-mediated versus traditional face-to-face language instruction. The quantitative component comprised a

quasi-experimental, mixed within-between design with random assignment to treatment and control conditions. The subsequent qualitative phase investigated students' experiences and perceptions through semi-structured interviews, thereby contextualizing and extending the quantitative results.

## Participants

The study was conducted during the summer semester of 2024 at a major university-affiliated language institute in Tehran, Iran, which offers intensive in-service English programs focusing on productive skills. A total of 120 male EFL students participated. Their age ranged from 24 to 36 years ( $M = 28.2$ ,  $SD = 3.65$ ). They were all active military personnel enrolled in a four-month intensive English language training program as part of their professional development. All had completed a minimum of eight years of formal English instruction within Iran's national education system, and held undergraduate degrees in humanities-related disciplines. To minimize variability in academic orientation, the sample was restricted to graduates of management ( $n = 42$ ), history ( $n = 28$ ), political science ( $n = 29$ ), and physical education ( $n = 21$ ). However, as a number of participants in each group ceased to participate in the program, the study ended with 57 participants in the treatment group and 54 participants in the control group.

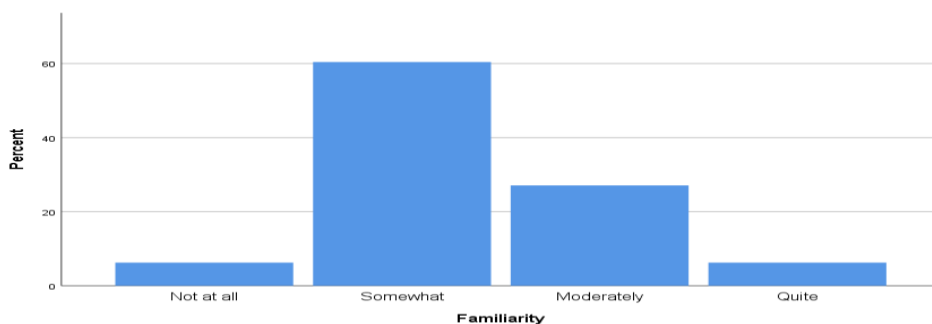
The participants were recruited through convenience sampling and were randomly distributed into six intact classes upon entry to the program. These classes were subsequently randomly assigned to either the treatment or control condition. Instruction was implemented over a continuous four-month period, during which the participants attended face-to-face classes five days per week, from Saturday to Wednesday, between 7:30 a.m. and 1:30 p.m. This schedule amounted to approximately six hours of daily instruction and reflects a highly intensive language learning environment. Such intensity is characteristic of in-service military language programs and was intended to maximize sustained exposure to English and opportunities for communicative practice.

Across the six classes, instruction was standardized in terms of contact hours, curricular sequencing, instructional materials, assessment procedures, and learning objectives. All groups followed the same institutionally prescribed syllabus, which was aligned with the program's stated emphasis on the development of productive skills, particularly speaking and writing. Instructors adhered to a common course outline and pacing guide to ensure parallel progression through course content.

The sole systematic difference between the treatment and control groups concerned the pedagogical intervention under investigation. Aside from this instructional manipulation, no variations were introduced with respect to instructional time, classroom activities, teacher workload, or assessment frequency. This design was intended to isolate the effects of the target pedagogical approach while minimizing the influence of extraneous variables, thereby strengthening the internal validity of the study.

Baseline proficiency was assessed using the Oxford Placement Text (OPT; Allan, 2004). Although the instructional focus of the classrooms was on productive skills, the OPT was employed solely to ensure a common baseline of overall language proficiency across the participants. As a standardized and widely validated measure of general L2 proficiency, the OPT provides a reliable estimate of learners' global language competence, which underpins performance in productive skills. The test was not used as an outcome measure but rather as a control variable to confirm group homogeneity prior to instruction.

Scores ranged from 18 to 29, corresponding to the A2 level of the Common European Framework of Reference for Languages (A1 to C2). An independent samples *t*-test confirmed no statistically significant difference between the groups ( $t_{(109)}=-1.06, p=.71$ ), thereby verifying their homogeneity at the outset. An a priori power analysis was conducted with G\*Power (Faul et al., 2007), indicating that a sample size of 52 for each group would be sufficient to detect medium effect sizes ( $f = 0.40$ ) with 0.80 power in the chosen design. The actual sample size ( $n = 111$ ) therefore provided strong statistical power.



**Figure 1.** Treatment group’s familiarity with AI tools

Within the treatment group, a survey of AI familiarity revealed limited prior exposure: 6.3% reported no familiarity, 60.4% were somewhat familiar, 27.1% moderately familiar, and 6.3% quite familiar (See Figure 1). None had previously used AI tools for writing, providing a consistent baseline for evaluating the intervention.

## Instrumentation

The instructional intervention targeted three core writing genres: narrative, descriptive, and instructional/procedural. These “elemental genres” are pedagogically relevant for A2-level students, providing foundational scaffolds for progression toward more complex academic writing (Hyland, 2007). Their selection reflects both linguistic priorities and sociocultural considerations, enabling students to recount personal experiences, describe familiar contexts, and engage in culturally embedded practices (Hyland, 2007) while supporting key writing competencies such as grammatical accuracy, vocabulary control, organization, and coherence.

Narrative writing tasks focused on recounting personal experiences and past events (e.g., *My Weekend*), encouraging the use of past simple tense and chronological markers (e.g., *first, then, after that*) to support temporal sequencing and narrative fluency. Descriptive writing tasks (e.g., *My Best Friend*) required students to depict people, places, or objects using present

simple tense and fundamental adjectives, fostering observational precision and cultural relevance. The instructional/procedural writing component guided students in creating simple how-to guides (e.g., *Making Tea/Hungarian Cabbage and Noodles*) using imperative verbs and sequencing words to reflect everyday practices.

All prompts were adapted from the *Top-Notch Series* and *Complete Assessment Package* (Pearson Education, 2016) with visual supports incorporated to enhance comprehension and engagement, particularly for procedural tasks. Instructors adhered to standardized lesson scripts including explicit instruction, modeling, guided practice, and gradual release toward independent writing to maintain instructional fidelity.

Four primary instruments were employed. First, the OPT was administered to assess baseline proficiency. Second, writing proficiency was assessed through parallel pre-test and post-test tasks. In the pre-test, the participants responded to three writing textbook-based prompts (e.g., *My Typical Day*, *My Perfect Weekend*) and produced 80–100-word compositions per prompt within a 30-minute time limit. The post-test included comparable prompts under identical conditions. Third, *Monica*<sup>1</sup>, as an AI-based language tool, served as the independent variable for the treatment group. *Monica* was selected for its unrestricted availability in Iran, unlike platforms such as ChatGPT, which face geopolitical limitations. Its integration of natural language processing and knowledge management features, along with cross-platform compatibility (desktop and mobile), ensured equitable access and pedagogical consistency. The participants used *Monica* to revise drafts focusing on grammar, vocabulary, organization, and style. Usage was standardized through a two-hour training workshop and monitored via teacher logs, classroom observations, and student self-reports. Instructors completed structured session logs after each class, documenting lesson activities, AI usage (for treatment classes), student engagement, and any deviations from the lesson plan. Classroom Observations: Research team members also

---

<sup>1</sup> <https://monica.im/home>

conducted periodic observations to ensure adherence to the protocol, providing corrective feedback where necessary. Moreover, after each session, students completed a short anonymous checklist reporting their engagement, perceived clarity of instruction, technical issues with AI (if any), and difficulties encountered during drafting or revision.

Across the intervention, teachers noted occasional technical delays with the AI tool, minor difficulties in students adjusting to iterative AI feedback, and variations in prompt interpretation. Students generally reported positive engagement, but some expressed temporary uncertainty about how to selectively apply AI suggestions without over-relying on them. These issues were addressed during class discussions and incorporated into the teachers' logs to inform session adjustments, ensuring consistency across classes.

Finally, students' subjective experiences were captured using a researcher-developed, quantitative questionnaire grounded in established educational technology and AI-in-education frameworks (Tran, 2025). The instrument was adapted conceptually from prior framework-based studies rather than directly adopted from an existing validated scale (Lin & Crosthwaite, 2024; Liu et al., 2024; Song & Song, 2023; Yan, 2023)

The questionnaire comprised 21 predominantly closed-ended Likert-scale items, organized into four domains: demographic/background (five items), pre-intervention assessment (six items), during-intervention monitoring (five items), and post-intervention evaluation (five items). Pre-intervention items measured AI familiarity, writing confidence, and expectations; during-intervention items captured interaction patterns and perceived benefits; post-intervention items assessed perceived skill gains, future adoption intentions, and anxiety reduction. One multiple-choice item in both the pre- and post-intervention phases elicited changes in specific writing challenges.

Content validity was addressed through expert alignment with established educational technology constructs and through a pilot study with five EFL students, which informed item clarity, wording, and cultural

appropriateness. Internal consistency reliability was evaluated using Cronbach's alpha for the Likert-scale items, yielding acceptable reliability coefficients across domains ( $\alpha \geq .70$ ). A uniform 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) was used to enhance response consistency and comparability across phases. Given its quantitative design, the questionnaire was not intended as an open-ended qualitative instrument; rather, it served to capture structured self-reported perceptions before, during, and after the intervention.

### **Data Collection Procedure**

This study spanned a 16-week summer semester and was structured into three sequential phases: pre-intervention (induction), implementation, and post-intervention. The pre-intervention induction phase was conducted during the first instructional week and formed part of the 16 scheduled class sessions, during which the participants were oriented to the study procedures and completed baseline measures. The implementation phase occupied the subsequent instructional weeks, and the post-intervention phase was conducted during the final week of the semester. Writing instruction was integrated into a general English course to preserve ecological validity.

### **Phase 1: Pre-Intervention (Weeks 1–2)**

The participants completed the OPT and the baseline questionnaire. In the second week, the participants in the treatment condition attended a mandatory two-hour interactive workshop on the AI-assisted writing tool, *Monica*. This workshop was designed to introduce the platform's core functionalities and provide structured instruction on effective prompt engineering strategies. It moved beyond basic instruction to emphasize strategic prompting, the critical appraisal of AI suggestions, and the metacognitive integration of feedback. This involved training the participants to pause, reflect on how AI's suggestions aligned with their own writing intentions, and consciously

regulate their revision strategies, thereby equipping them to leverage AI as a collaborative cognitive tool rather than a passive editor.

## **Phase 2: Implementation (Weeks 3–14)**

The core implementation phase comprised twelve 80-minute writing-focused sessions distributed across three instructional modules over a twelve-week period. These sessions were embedded within the center's regular English course and aligned with the institutional syllabus, which is informed by commercially available materials (e.g., *Top Notch*), but were supplemented with genre-based writing tasks developed specifically for this study.

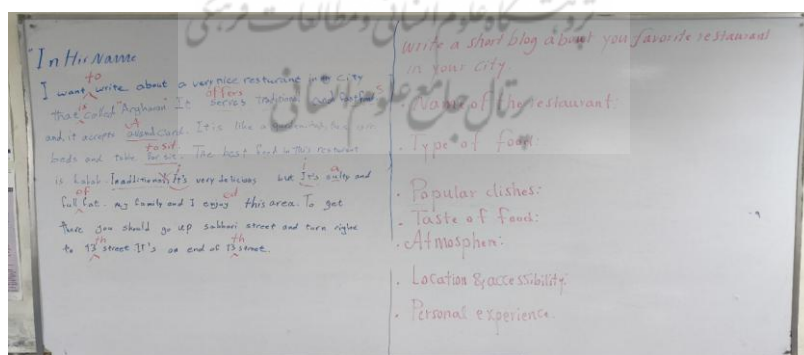
Instruction was not exclusively devoted to writing, as the broader course continued to address integrated language skills; however, the sessions reported here represent the writing component of the normal curriculum, delivered by the course instructor during regularly scheduled class time. The three modules targeted narrative (Weeks 3–6), descriptive (Weeks 7–10), and procedural writing (Weeks 11–14). This genre-based progression aimed to expose learners to diverse textual forms while developing transferable writing skills through scaffolded practice. To ensure methodological consistency, each session followed a structured pedagogical sequence.

All instructional sessions were conducted inside the classroom during regularly scheduled class time, and both experimental and control groups followed identical initial procedures. Each session began with teacher-led prewriting activities, including brainstorming tasks, genre-specific discussions, and explicit modeling of rhetorical and linguistic features. These activities were delivered whole-class and were identical for both groups, serving to activate prior knowledge, scaffold creative and critical thinking, and orient students to the communicative purposes and structural conventions of the target genre (Hyland, 2007). Following prewriting, the participants in both groups engaged in individual drafting, producing original texts under time constraints to approximate authentic writing conditions. Drafting was completed independently, without peer correction or collaborative dialogue

at this stage, to ensure that initial text production reflected individual competence.

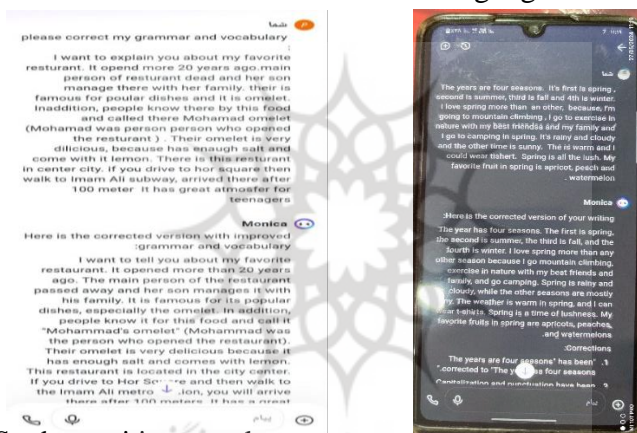
The revision stage constituted the sole point of procedural divergence between conditions. The participants in the experimental group used their personal digital devices (smartphones or laptops) to engage in iterative, AI-mediated revision with *Monica*. This process emphasized self-directed interaction with AI-generated feedback targeting grammar, vocabulary, cohesion, and genre-appropriate organization. In contrast, the participants in the control group completed guided self-revision using a standardized checklist focusing on clarity, grammatical accuracy, lexical appropriateness, and text organization. No AI tools were permitted in the control condition.

Peer feedback was implemented for both groups only after the revision stage, ensuring parity of collaborative exposure. These peer feedback sessions were conducted within condition-specific groups to prevent cross-contamination and focused on evaluating revised drafts rather than generating corrective input during drafting. In selected sessions, representative student texts from each condition were projected and discussed through teacher-guided whole-class feedback, reinforcing shared genre awareness and revision strategies. Across the intervention, each participant completed twelve major writing assignments, distributed evenly across the three instructional modules (see Figures 2 & 3.).



**Figure 2.** Example of teacher's written feedback.

Reflective practice was embedded into the instructional design. The instructor consistently modeled effective AI prompt formulation for the treatment group (e.g., *highlight grammatical errors in my writing; suggest ways to improve vocabulary in this paragraph*), thereby cultivating participants' AI literacy and metacognitive awareness of tool use. Simultaneously, individualized teacher feedback was provided to all participants, ensuring equitable instructional support across conditions. This iterative cycle – from drafting to AI-mediated or checklist-driven revision, followed by peer reflection and teacher feedback – was intended to foster not only the development of genre-specific writing competencies but also a deeper understanding of the affordances and limitations of AI-mediated language learning.



**Figure 3.** Student writing sample

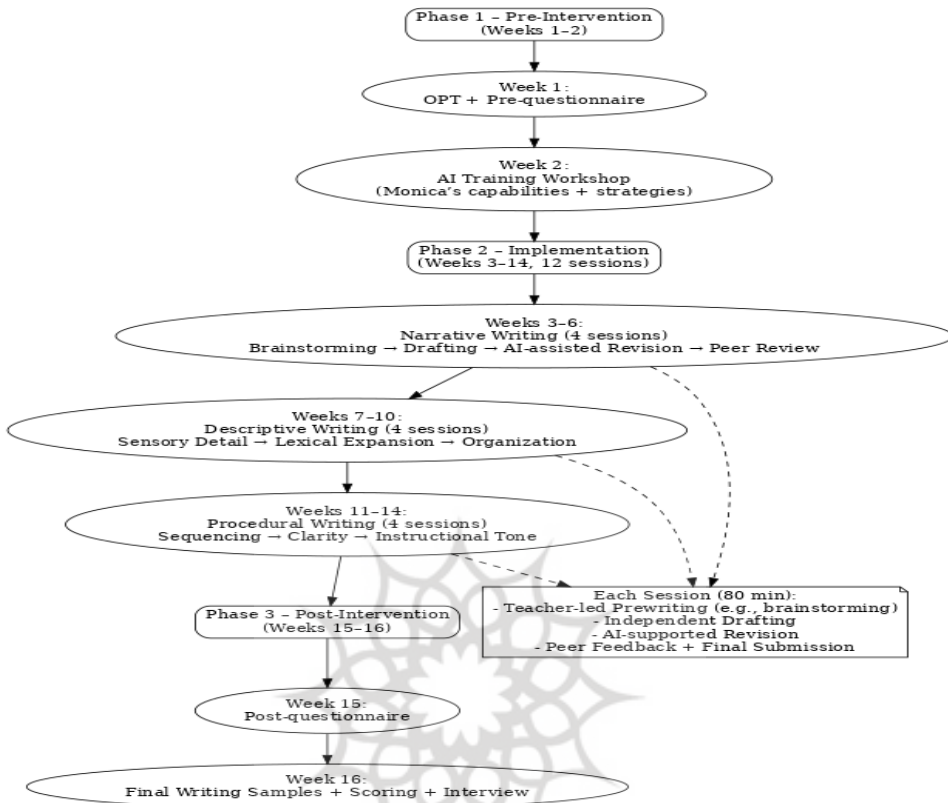
In addition, to ensure instructional consistency across classes within each condition, several measures were implemented. All three treatment classes received identical AI-supported writing instruction, using the same prompts, scaffolding activities, and access to the AI tool (*Monica*), with teachers following a standardized lesson plan and session-by-session protocol. Similarly, the three control classes followed the same writing tasks and pre-designed self-revision procedures, guided by a common checklist, ensuring that the only systematic difference between groups was the presence or absence of AI-mediated feedback. In addition, teachers participated in brief orientation sessions prior to the intervention to calibrate instructional

delivery, clarify procedural expectations, and align grading or monitoring practices. Regular supervision and session observations were conducted to verify adherence to the protocol and minimize inter-class variability within each condition.

Moreover, the intervention was delivered by two experienced EFL instructors affiliated with the language institute. Both instructors had a minimum of ten years of teaching experience in EFL contexts and were familiar with student-centered methodologies. Both had experience specifically teaching AI-based writing classes.

### **Phase 3: Post-Intervention (Weeks 15–16)**

The post-intervention phase, conducted during the final two weeks, was designed to capture the outcomes of the intervention through both quantitative and qualitative measures. In week fifteen, all participants completed a post-intervention questionnaire examining changes in writing confidence, skills, and attitudes toward AI integration. In the final week, final writing samples were collected for subsequent blind scoring. All compositions were evaluated using the standardized composition profile rubric by Jacobs et al. (1981) which assigns scores across five criteria: Content (25 points), Organization (25 points), Language Use (25 points), Vocabulary (15 points), and Mechanics (10 points) on a 100-point scale. Scores were proportionally converted to Iran's 0–20 grading system following Rahimi and Fathi (2021). Inter-rater reliability yielded an acceptable score of 0.84, indicating strong consistency between raters. In addition, the intra-rater reliability, based on a random selection of 25% of the participants writing samples resulted in an index of 0.93 in the case of the first rater and 0.89 in the case of the second rater. Complementary semi-structured interviews were also conducted with 15 participants using a protocol adapted from Song and Song (2023). The protocol included five open-ended questions about their experiences with AI-assisted instruction (See Appendix B.).



**Figure 4.** Visual overview of the study timeline

To facilitate clear communication, the interviews were conducted in Persian and audio-recorded. Each interview lasted about 35 minutes, starting with rapport-building questions and progressing to participants' perceptions of AI-assisted instruction, including its benefits, limitations, and challenges. A visual overview of the study timeline is shown in Figure 4 above.

## Data Analysis

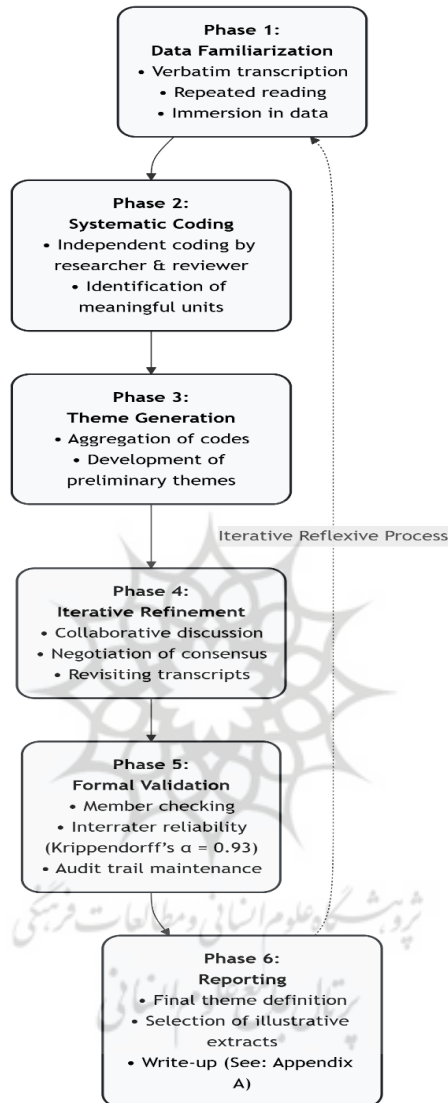
### Quantitative Phase

To ensure comparability between groups, a series of independent samples *t*-tests were run to check for potential pre-existing differences in English

proficiency test scores (i.e., the OPT) as well as the pretest writing scores. As no significant differences were observed, we proceeded to evaluate the impact of the intervention on participants' total writing score as well as the rubric subscales of Content, Organization, Language Use, Vocabulary, and Mechanics using a series of Split-plot Analysis of Variance (ANOVA), which examines both within-subjects (time) and between-subjects (group) effects simultaneously and probes the main effects and their interaction. Additionally, the participants' responses to the questionnaire items were analyzed using frequency counts to determine the distribution of selected options.

## **Qualitative Phase**

Qualitative data were collected through semi-structured interviews with 15 EFL students. To ensure confidentiality and comply with ethical standards, all participants were assigned pseudonyms (ST1, ST2, etc.). All interviews were transcribed verbatim, and the transcripts were reviewed repeatedly to develop a thorough understanding of participants' experiences, perceptions, and recurring themes. To ensure accuracy and credibility, a member-check procedure was implemented, whereby transcripts were returned to participants for verification. In addition, all research activities, including data collection and analysis, were systematically documented in an audit trail to enhance transparency and allow for independent verification.



**Figure 5.** The flowchart follows the six phases of thematic analysis as defined by Braun & Clarke (2021)

As Figure 5 above illustrates, data analysis followed the thematic analysis framework proposed by Braun and Clarke (2021). The process began with data familiarization, in which researchers immersed themselves in the

transcripts and written responses to gain comprehensive insight into the dataset. Meaningful units were then identified and independently coded by both the primary researcher and a qualitative research expert, ensuring rigorous and unbiased identification of patterns. Initial codes were subsequently grouped into preliminary themes, which were iteratively refined through discussion and negotiation to resolve divergent interpretations. Original transcripts were revisited multiple times throughout this process to confirm that the themes accurately represented participants' perspectives, thereby ensuring coherence and consistency. Validation procedures, including member checking, interrater reliability assessment, and audit trail documentation, were applied throughout the analysis. Interrater reliability was calculated using Krippendorff's (2004) alpha, which reached 0.93, exceeding the minimum acceptable threshold and demonstrating the robustness of the analytical process. Finalized themes were defined and reported alongside illustrative examples, providing a nuanced and comprehensive account of students' perspectives.

## **RESULTS**

### **Quantitative Phase**

As the descriptive statistics presented in Table 2 indicate, the two groups started the writing program almost at the same level of writing ability. While the control group's mean score at the pretest was 14.09 (SD = 1.92), the treatment group's score was found to be 14.33 (SD = 1.42). A similar pattern was observed for all the components examined in the writing scoring rubric (i.e., Content, Organization, Language Use, Vocabulary, and Mechanics) with the treatment group marginally outperforming the control group in all cases. As the distribution of the scores in all cases for both groups did not violate the normality assumption, a number of independent samples *t*-tests were run to ensure the comparability of the two groups at the study outset. The results revealed no significant differences between the two groups at the pretest in the writing total score and the five rubric components, all yielding

very small to moderate effect sizes (Cohens'  $d$ ),  $t(109)_{\text{Pretest}} = -0.75, p = 0.46$ , Cohen's  $d = 0.33$ ;  $t(109)_{\text{Content}} = -0.54, p = 0.59$ , Cohen's  $d = 0.10$ ;  $t(109)_{\text{Organization}} = -1.26, p = 0.21$ , Cohen's  $d = 0.24$ ;  $t(109)_{\text{Language Use}} = -1.55, p = 0.12$ , Cohen's  $d = 0.29$ ;  $t(109)_{\text{Vocabulary}} = 0.13, p = .90$ , Cohen's  $d = 0.02$ ;  $t(109)_{\text{Mechanics}} = -1.18, p = .24$ , Cohen's  $d = 0.23$ .

**Table 2.** Descriptive Statistics on Participants' Performance on the Pretest

	Group	N	Minimum	Maximum	Mean	Std. Deviation
Total Score	Control	54	10.60	17.20	14.09	1.92
	Treatment	57	11.40	16.60	14.33	1.42
Content	Control	54	12	21	17.28	2.48
	Treatment	57	14	21	17.51	2.01
Organization	Control	54	13	21	16.81	2.55
	Treatment	57	15	21	17.35	1.92
Language Use	Control	54	12	21	16.59	2.76
	Treatment	57	14	21	17.33	2.26
Vocabulary	Control	54	7	15	11.24	2.45
	Treatment	57	9	15	11.19	1.42
Mechanics	Control	54	6	9	7.12	1.06
	Treatment	57	6	9	7.33	.72

By the end of the program, at the posttest, both groups exhibited improvement both in the overall writing score and the five mentioned components, with the treatment group outperforming the control group in all cases. Unlike at the pretest, these posttest differences were substantial for some components. As indicated by Table 3, while the control group's overall writing score was to 15.17 (SD = 1.71), that of the treatment group improved more and reached 16.19 (SD = 1.33). The difference was also noticeable in the case of Organization (control = 18.09, treatment = 20.04) and Language Use (control = 18.17, treatment = 20.54), but not for Content (control = 18.41, treatment = 18.70), Vocabulary (control = 12.11, treatment = 12.23), and Mechanics (control = 7.96, treatment = 8.46).

**Table 3.** Descriptive Statistics on Participants' Performance on the Posttest

	Group	N	Minimum	Maximum	Mean	Std. Deviation
Total Score	Control	54	12.00	17.60	15.17	1.71
	Treatment	57	13.00	20.00	16.19	1.33
Content	Control	54	10	22	18.41	2.86
	Treatment	57	15	25	18.70	1.97
Organization	Control	54	15	22	18.09	2.48
	Treatment	57	15	25	20.04	2.36
Language Use	Control	54	12	21	18.17	2.68
	Treatment	57	15	25	20.54	2.17
Vocabulary	Control	54	9	15	12.11	1.76
	Treatment	57	9	15	12.23	1.65
Mechanics	Control	54	7	9	7.96	.78
	Treatment	57	6	9	8.46	.76

Since the distribution of scores in the posttest was found to be normal, a number of split-plot ANOVAs were run to examine both the effect of *Time* and *Group* as well as their interaction. In the case of the total score, the Levene's tests of equality of error variance were not found to be significant ( $p_{pretest} = 0.22$ ,  $p_{posttest} = 0.10$ ). In addition, the Box's test of equality of covariance matrices was found non-significant,  $p = 0.08$ . The results of the split-plot ANOVA indicated a statistically significant interaction between *Time* and *Group*, Wilks' Lambda = 0.86,  $F(1, 109) = 17.15$ ,  $p < 0.001$ ,  $\eta^2 = .14$ , which is a large effect size according to Cohen's (1988) benchmarks to Eta squared measure of effect size (i.e.,  $\eta^2 = 0.01 = \text{small}$ ,  $0.06 = \text{medium}$ ,  $0.14 = \text{large}$ ). This indicates that the change in the participants' scores over time has been different for the two groups. There was also a substantial effect for *Time*, Wilks' Lambda = 0.33,  $F(1, 109) = 224.05$ ,  $p < 0.001$ ,  $\eta^2 = 0.67$ . This further indicates that, on average, the scores have increased over time from pretest to posttest for both groups. More importantly, the main effect for *Group*, comparing the effect of the mode of presentation, was found statistically significant,  $F(1, 109) = 4.97$ ,  $p = 0.03$ ,  $\eta^2 = .06$ , indicating that while both groups had significantly improved over time from pretest to

posttest as the result of the instruction they received, the treatment group making additional use of AI could significantly outperform the control group over time. This finding was confirmed by a simple effect post-hoc analysis at posttest,  $t(109)_{\text{Posttest}} = -3.62, p < 0.001$ , Cohen's  $d = 0.67$ , indicating a large effect size.

Regarding the subscale of Content in the writing scoring rubric, the Levene's tests of equality of error variance were not found to be significant ( $p_{\text{pretest}} = 0.64, p_{\text{posttest}} = 0.08$ ). In addition, the Box's test of equality of covariance matrices was found to be larger than .001,  $p = 0.03$ . The results of the main analysis showed no significant interaction between *Time* and *Group*, Wilks' Lambda = 0.99,  $F(1, 109) = 0.13, p = 0.72, \eta^2 = 0.001$ . On the other hand, there was a substantial effect for *Time*, Wilks' Lambda = 0.38,  $F(1, 109) = 175.38, p < 0.001, \eta^2 = 0.62$ , indicating that the scores for both groups significantly changed over time from pretest to posttest. However, the main effect for *Group*, comparing the effect of the instruction, was not found statistically significant,  $F(1, 109) = 0.36, p = 0.55, \eta^2 = 0.003$  indicating no advantage for the treatment group using AI as part of their writing instruction. This was also confirmed by the simple effect post-hoc analysis conducted at the posttest,  $t(109)_{\text{Posttest}} = -0.63, p = .53$ , Cohen's  $d = 0.12$ , which is a quite small effect size.

In the case of the Organization, the Levene's tests of equality of error variance were not found to be significant ( $p_{\text{pretest}} = 0.20, p_{\text{posttest}} = 0.23$ ). In addition, the Box's test of equality of covariance matrices was found non-significant,  $p = 0.02$ . The results of the split-plot ANOVA indicated a statistically significant interaction between *Time* and *Group*, Wilks' Lambda = 0.88,  $F(1, 109) = 14.90, p < 0.001, \eta^2 = 0.12$ . There was also a substantial effect for *Time*, Wilks' Lambda = 0.48,  $F(1, 109) = 118.28, p < 0.001, \eta^2 = .52$ . The main effect for *Group*, too, was found statistically significant,  $F(1, 109) = 9.39, p = 0.003, \eta^2 = 0.08$ , indicating that while both groups had significantly improved over time from pretest to posttest in their scores in the organization of the texts they wrote, the treatment group making additional use of AI could significantly outperform the control group. This was also

confirmed by the simple effect post-hoc analysis conducted at the posttest,  $t(109)_{\text{Posttest}} = -4.23$ ,  $p < 0.001$ , Cohen's  $d = 0.81$ .

Regarding the Language Use component, a similar pattern of results was observed as that of the Organization. The Levene's tests of equality of error variance were not found to be significant ( $p_{\text{pretest}} = 0.68$ ,  $p_{\text{posttest}} = 0.12$ ). In addition, the Box's test of equality of covariance matrices was found non-significant,  $p = 0.09$ . A statistically significant interaction was found between *Time* and *Group*, Wilks' Lambda = 0.86,  $F(1, 109) = 17.80$ ,  $p < 0.001$ ,  $\eta^2 = 0.14$ . There was also a substantial effect for *Time*, Wilks' Lambda = 0.42,  $F(1, 109) = 151.18$ ,  $p < 0.001$ ,  $\eta^2 = 0.58$ . The main effect for *Group*, too, was found statistically significant,  $F(1, 109) = 13.26$ ,  $p < 0.001$ ,  $\eta^2 = 0.11$ , indicating that while both groups had significantly improved over time, the treatment group could significantly outperform the control group due to the type of instruction it received, also confirmed by the simple effect post-hoc analysis conducted at the posttest,  $t(109)_{\text{Posttest}} = -5.12$ ,  $p < 0.001$ , Cohen's  $d = 0.97$ , a very large effect size.

In the case of the changes in the participants' scores in the Vocabulary component, a different pattern of results was observed. The Levene's tests of equality of error variance were not found to be significant ( $p_{\text{pretest}} = 0.23$ ,  $p_{\text{posttest}} = 0.52$ ). In addition, the Box's test of equality of covariance matrices was found to be larger than 0.001,  $p = 0.06$ . However, while the interaction between *Time* and *Group* was not statistically significant, Wilks' Lambda = 0.99,  $F(1, 109) = 0.27$ ,  $p = 0.61$ , there was a substantial effect for *Time*, Wilks' Lambda = 0.78,  $F(1, 109) = 35.63$ ,  $p < 0.001$ ,  $\eta^2 = 0.25$ . However, the effect of the *Group*, comparing the two groups' performance over time from pretest to posttest was not found statistically significant,  $F(1, 109) = 0.01$ ,  $p = 0.91$ , indicating that while both groups could improve over time as the result of the instructions they received, the AI could not help the treatment group outperform the control group. This was also confirmed by the simple effect post-hoc analysis conducted at the posttest,  $t(109)_{\text{Posttest}} = -0.36$ ,  $p = 0.72$ , Cohen's  $d = 0.07$ , which is a small effect size

Finally, regarding the two groups' performance in the Mechanics of their writings, the Levene's tests of equality of error variance were not found to be significant ( $p_{pretest} = 0.12$ ,  $p_{posttest} = 0.69$ ). In addition, the Box's test of equality of covariance matrices was found non-significant,  $p = 0.06$ . The results of the split-plot ANOVA indicated no significant interaction between Time and Group, Wilks' Lambda = 0.98,  $F(1, 109) = 2.65$ ,  $p = 0.11$ . There was, however, a substantial effect for Time, Wilks' Lambda = 0.47,  $F(1, 109) = 120.92$ ,  $p < 0.001$ ,  $\eta^2 = 0.53$ . The main effect for Group, was also found statistically significant,  $F(1, 109) = 6.99$ ,  $p = 0.01$ ,  $\eta^2 = 0.06$ , indicating that while both groups had significantly improved over time, the treatment group could significantly outperform the control group due to the type of instruction it received. This was also confirmed by the simple effect post-hoc analysis conducted at the posttest,  $t(109)_{Posttest} = -3.39$ ,  $p = 0.001$ , Cohen's  $d = 0.65$ , which is a quite large effect size.

## Survey Results

Before the instruction began, the treatment group was asked for their attitude toward their writing abilities, its importance, and the use of AI in the writing process. Almost 80% of the participants agreed or strongly agreed that improving their English writing skills was their priority, though only 47.4% of them expressed high confidence in their writing abilities. Regarding their attitude toward using AI in the writing process, only 23% of them expressed having positive expectations about using *Monica* as an AI assistant in their writing process. More importantly, 51% of them expressed their concerns about the potential negative effects such AI tools might have in this regard. Table 4 summarizes the treatment group participants' responses to these items.

**Table 4.** Treatment Groups' Perceptions of their L2 Writing Abilities and AI Use before the Instruction Began

	Strongly Disagree		Disagree		No Idea		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%	f	%
1 Improving my English writing skills is my top priority.	-	-	5	8.8	7	12.3	29	50.9	16	28.1
2 I have confidence in my current writing abilities in English.	5	8.8	12	21.1	13	22.8	20	35.1	7	12.3
3 I have positive expectation about using Monica as an AI assistant.	4	7	29	50.9	11	19.3	11	19.3	2	3.5
4 I am concerned about the potential negative effects of using Monica.	-	-	18	31.6	10	17.5	29	50.9	-	-

A few weeks after the instruction began, when the participants had practiced using *Monica* as an assistant in their writing process, they were asked about their experience with *Monica* so far. As evident in Table 5, more than 90% of the participants agreed or strongly agreed that their experience with *Monica* had been positive so far, they had been able to use more efficient strategies for communicating with *Monica* in their writing process, and they could easily collaborate with it in order to complete their writing assignments.

Interestingly enough, about 95% of them expressed an increase in their self-confidence in writing in English which has substantially improved in comparison with their responses before the instruction began. Moreover, while more than half the participants had expressed their concerns regarding the negative effects of using AI in the writing process, after a few weeks of working with *Monica*, 96.5% of them agreed or strongly agreed that the advantages of using *Monica* as an AI writing assistant outweigh its disadvantages.

**Table 5.** Treatment Groups' Perceptions of their Experience of Using AI a few Weeks after the Instruction Began

	Strongly Disagree		Disagree		No Idea		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%	f	%
1 My experience with <i>Monica</i> has been positive so far.	-	-	-	-	4	7	32	56.1	21	36.8
2 I have adopted more efficient strategies for communicating with <i>Monica</i> during the writing process.	-	-	-	-	3	5.3	35	61.4	19	33.3
3 I can easily collaborate with <i>Monica</i> to complete my writing assignments.	-	-	-	-	5	8.8	39	68.4	13	22.8
4 My confidence in writing in English has increased since I began using <i>Monica</i> .	-	-	-	-	3	5.3	41	71.9	13	22.8
5 The advantages of using <i>Monica</i> as an AI writing assistant outweigh its disadvantages.	-	-	-	-	2	3.5	44	77.2	11	19.3

Finally, after the instruction was over, the participants were asked about the extent to which they believed *Monica* could help them with their writing skills, its effect on the amount of anxiety they felt about writing in English, and whether they would recommend it to other learners. 93% of the participants believed (agreed or strongly agreed) that *Monica* had significantly helped improve their writing. Also, 93% of them believed that *Monica* could help them overcome their challenges in writing in English.

**Table 6.** Treatment Groups' Perceptions of the Impact of *Monica* on their Writing Abilities after the Instruction

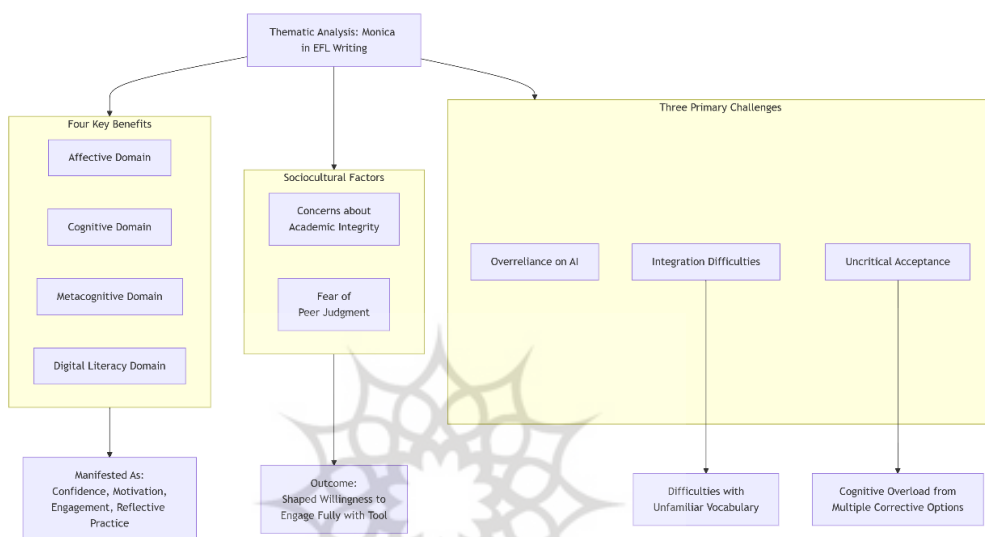
	Strongly Disagree		Disagree		No Idea		Agree		Strongly Agree	
	f	%	f	%	f	%	f	%	f	%
1 <i>Monica</i> has significantly helped improve my writing skills.	-	-	-	-	4	7	40	70.2	13	22.8
2 <i>Monica</i> has assisted me in adopting new strategies to improve my writing skills.	-	-	-	-	20	35.1	32	56.1	5	8.8
3 With the help of <i>Monica</i> , my challenges with writing in English have diminished.	-	-	-	-	4	7	16	28.1	37	64.9
4 Utilizing <i>Monica</i> has reduced my anxiety about writing in English.	-	-	16	28.1	6	10.5	31	54.4	4	7
5 In the future, I plan to use <i>Monica</i> or similar tools for learning English.	-	-	-	-	2	3.5	37	64.9	18	31.6
6 I will recommend <i>Monica</i> to other language learners as a valuable resource for enhancing their writing skills.	-	-	-	-	5	8.8	43	75.4	9	15.8

Regarding the effect of using *Monica* on their anxiety about writing in English, more than 60% of the participants reported the positive impact of *Monica* on reducing their experienced anxiety in writing in English. As a result of their positive experiences with *Monica*, 96.5% of them expressed their willingness to use *Monica* or similar AI tools in the future for learning English. Moreover, more than 90% stated that they would recommend *Monica* to other language learners as a valuable resource for enhancing their writing skills (See Table 6.).

## Qualitative Phase

Thematic analysis of EFL students' interviews revealed four key benefits, three primary challenges, and several sociocultural considerations associated

with the integration of *Monica* into writing instruction. Benefits spanned *affective*, *cognitive*, *metacognitive*, and *digital literacy* domains, reflecting increased confidence, motivation, engagement, and reflective practice.



**Figure 6.** Summary of the most salient themes

Challenges included overreliance on AI, difficulties integrating unfamiliar vocabulary, and uncritical acceptance of feedback, as well as cognitive overload when multiple corrective options were presented. Sociocultural considerations, particularly concerns regarding academic integrity and peer judgment, also shaped students' willingness to engage fully with the tool. A summary of the most salient themes is presented in Figure 6 above.

## DISCUSSION

The present study examined the effect of using *Monica* on EFL students' writing development. Specifically, it aimed to determine the extent to which

AI-mediated writing instruction with *Monica* enhanced students' writing proficiency compared to face-to-face collaborative methods. In addition, the study sought to explore the key benefits, challenges and limitations associated with integrating this tool into EFL writing instruction.

*To what extent does AI-mediated writing instruction using Monica enhance EFL students' writing proficiency compared to traditional face-to-face collaborative methods?*

To address the first research question, both the AI-assisted group and the face-to-face collaborative group demonstrated measurable improvement from pretest to posttest. However, the AI-assisted group achieved significantly greater gains, suggesting that *Monica* served as an effective mediational artifact- an essential concept within SCT- that supported learners' progression from assisted to more autonomous performance within their ZPD. Statistical analyses also revealed that students in the AI-assisted group significantly outperformed their peers in three critical components of writing: Organization, Language Use, and Mechanics. These lower-order writing skills are foundational to effective academic communication because they directly affect clarity, coherence, and overall readability (Marzuki et al., 2023). The intervention produced large effect sizes in Language Use (Cohen's  $d = 0.97$ ) and Organization ( $d = 0.81$ ), and a moderate effect in Mechanics ( $d = 0.65$ ). These results underscore the substantial role of *Monica* in scaffolding students' engagement with form-focused aspects of writing, enabling them to internalize linguistic patterns and structures that may otherwise remain beyond their independent control.

One explanation for these findings lies in the capacity of educational technologies to support multiple dimensions of EFL writing, particularly at the micro-skill level. By offering adaptive tools and resources, AI technologies can enhance accuracy and foster meaningful language development (Mahapatra, 2024; Seyyedrezaei et al., 2022; Yan, 2023). Prior studies suggest that technology-enhanced writing instruction increases engagement and metacognitive awareness, thereby motivating sustained effort and deeper investment in writing tasks (Fathi & Rahimi, 2024; Jmaiel

et al., 2025; Tran, 2025). Together, these pedagogical affordances help explain the strong improvements observed in students' performance within technology-integrated contexts.

Another contributing factor is *Monica*'s ability to deliver immediate, individualized, and form-focused feedback. This functionality is supported by core principles of second language acquisition, as it directly promotes error noticing (Schmidt, 1990), provides the data necessary for guided revision (Bitchener & Ferris, 2012), and ultimately fosters self-regulated learning (Jmaiel et al., 2025; Tran, 2025; Zimmerman, 2002). Because lower-order writing skills are largely rule-governed and pattern-based, they are especially amenable to AI-driven interventions that provide the consistent, focused feedback shown to be effective. Within a SCT framework, this function positions *Monica* as a powerful mediational artifact within the students' ZPD, directly facilitating their progression from other-regulated to independent performance. Finally, the low-anxiety, high-autonomy environment it fostered encouraged the iterative practice essential for skill development, a finding consistent with a growing body of recent research on AI in EFL writing (Fathi & Rahimi, 2024; Mahapatra, 2024; Yan, 2023).

Among the three components assessed, the most substantial gains were observed in Organization—the “how” of writing—which pertains to the logical sequencing and coherent structuring of ideas (Marzuki et al., 2023). This is a critical area of development, as strong organization is essential to ensure that readers can follow arguments, narratives, or explanations with clarity. Applications such as *Monica* appear particularly effective in assisting students with cohesion and textual flow by providing real-time, actionable suggestions for restructuring sentences and paragraphs. This functionality also addresses common challenges in L2 writing such as achieving cohesion and textual flow, thereby corroborating the findings of prior research on AI's efficacy in improving textual coherence (Jmaiel et al., 2025; Song & Song, 2023; Tran, 2025). Recently, Mahapatra (2024) proposed examining whether low-proficiency students could benefit from AI in content organization; the

present findings indicate that A2-level students do benefit, particularly in structuring their writing.

This study further observed that, beyond measurable linguistic gains, AI interaction prompted a dramatic positive shift in students' perceptions, confidence, and motivation—a finding consistent with recent literature (Ghafouri et al., 2024; Mahapatra, 2024; Ryan & Deci, 2017). This shift is critically important, as learners with positive perceptions exhibit stronger behavioral intentions to adopt and integrate technology deeply (Chang et al., 2022). Our results demonstrated this principle in action: the transformation from initial skepticism (with only 23% of participants holding positive expectations) to widespread acceptance (90% reporting positive experiences) created the conditions for deeper engagement. As students began to engage positively with GenAI, they moved beyond simple use and started to merge its capabilities with their own cognitive processes, forming the foundations of human-AI intelligence. This transformation was primarily facilitated by the AI's role as a mediational tool within a SCT framework. A marked increase in writing confidence, reported by 95% of the AI group, was the key driver. The AI's immediate, non-judgmental feedback created a low-anxiety environment that encouraged the risk-taking and persistence essential for developing self-regulation and agency (Mahapatra, 2024; Tran, 2025; Vygotsky, 1978; Zimmerman, 2002). Central to this confidence-building was the tool's scaffolding capability, which broke down complex writing tasks into manageable steps, enabling students to progress through their ZPD toward independent performance. The immediacy of this support enhanced metacognitive awareness and sustained engagement (Jmaiel et al., 2025; Song & Song, 2023). Ultimately, this process of successfully accomplishing tasks with calibrated support boosted intrinsic motivation by directly fostering feelings of competence and autonomy (Ryan & Deci, 2017).

Despite these clear benefits, the study identified important limitations of AI-mediated instruction. Most notably, there was no significant improvement in content development—the “what” of writing—which involves generating and articulating ideas, arguments, and perspectives (Marzuki et al.,

2023). High-quality content requires originality, depth, and critical integration of ideas and perspectives which extend beyond rule-based correction (Molina et al., 2021). Therefore, these higher-order cognitive tasks appear to exceed the current capabilities of AI scaffolding tools. This finding presents a crucial nuance to the broader positive outcomes reported in previous studies (Gayed et al., 2022; Marzuki et al., 2023) and aligns with SCT's assertion that complex, ideational development is best facilitated through dialogic human mediation, a function AI cannot replicate. Furthermore, since content development is shaped by individual factors such as learners' proficiency, cultural background, and personal experiences (Wong & Mak, 2019), the continued importance of personalized teacher guidance is underscored to foster the deep conceptual engagement that AI cannot yet provide.

Another area of limited improvement was vocabulary acquisition. The negligible gain observed (Cohen's  $d=0.07$ ) aligns with theoretical models that posit lexical development as an incremental, context-dependent process requiring sustained exposure to authentic input (Nation, 2019; Schmitt, 2008). Unlike the rule-based domains (e.g., Mechanics), vocabulary acquisition demands the complex mapping of form to meaning, the recognition of nuanced connotations, and the ability to deploy words appropriately across diverse contexts (Richards, 2015; Teng, 2019). While *Monica* offered synonym suggestions and flagged lexical errors, these features did not translate into measurable vocabulary growth. Effective lexical acquisition depends on deep processing, frequent encounters, and meaningful use-conditions better met through rich input and interaction than through short-term AI feedback (Richards, 2015).

Interestingly, this finding diverges from recent studies reporting positive AI effects on vocabulary development (e.g., Gayed et al., 2022; Jmaiel et al., 2025; Marzuki et al., 2023; Song & Song, 2023). However, such discrepancies may stem from differences in intervention duration, AI tool features, or distinct student profiles. The relatively short treatment period in this study may not have allowed sufficient time for vocabulary growth to

emerge, reinforcing the view that lexical learning is gradual and long-term. Taken together, the findings highlight a clear distinction between surface-level skills (Organization, Language Use, Mechanics) and deep-level skills (Content, Vocabulary). AI-generated feedback tends to be immediate and rule-focused, making it effective for lower-order concerns, while teacher feedback is often more dialogic and meaning-oriented, targeting content development, argument coherence, and stylistic nuance (Tran, 2025). This helps explain why the AI-assisted group achieved substantial improvements in structural areas, whereas higher-order skills remained relatively unchanged.

*What are the key benefits, challenges and limitations of integrating Monica into EFL writing instruction?*

Qualitative data revealed a layered account of how *Monica* shaped writing development. While participants frequently highlighted affective, cognitive, and metacognitive benefits, their reflections also exposed drawbacks and persistent challenges that complicate AI's role in EFL writing instruction.

In addition to the commonly reported benefits of AI integration in language learning—such as 24/7 availability, tireless engagement, facilitation of collaboration, provision of authentic materials, and reduced anxiety in the Iranian context (Fereidouni & Farahian, 2024; Parviz, 2024)—this study revealed further advantages. Participants emphasized that unlike human teachers, whose performance can be affected by knowledge limitations, time constraints, mood fluctuations, concentration lapses, and emotional burdens, AI systems (e.g., *Monica*) operate without such psychological or affective constraints. This was illustrated by one student who noted:

*“My teacher is very good, but sometimes he has a bad day or is very busy with other students. Sometimes he doesn't know an answer. But my AI friend always has time for me. AI is always calm and knows many answers.” (ST5)*

This quality positioned AI not merely as a supplementary tool but also as a potential mentor-like figure, consistently available to support students

without judgment or fatigue. Another participant reinforced this mentor-like role, stating: *“I am sometimes scared to ask my teacher a question... But with this (Monica), I can ask anything. He is never tired and never says my question is bad. It is like having a patient mentor only for me....”* (ST9)

A recurring theme in the findings was the affective relief and motivational boost students experienced when interacting with *Monica*. Participants described the AI as a “silent partner” offering non-judgmental support and mitigating performance-related anxiety. One student remarked: *“I didn’t feel embarrassed making mistakes-it’s not like a teacher watching me”* (ST10). This affective reassurance not only reduced stress but also encouraged risk-taking and experimentation, aligning with Zimmerman’s (2002) assertion that confidence serves as a catalyst for self-regulated learning. The findings also resonate with recent research indicating that AI tools can foster students’ willingness to engage in language use despite making errors (Song & Song, 2023). Beyond reducing anxiety, several participants reported increased motivation and confidence in their writing. For instance, ST2 noted: *“I feel more confident writing now because Monica shows me what I can fix right away.”* Similarly, ST6 explained: *“It’s less stressful than waiting for teacher feedback. I can try, get help, and correct errors myself.”* These observations reinforce the notion that AI can function as an affective regulator in language learning, supporting evidence that such tools contribute to emotional regulation within educational contexts. Furthermore, the immediacy and constructiveness of AI-generated feedback appeared to reduce stress and enhance enjoyment, thereby deepening student’s emotional engagement with the feedback process (Teng, 2024).

At the cognitive level, *Monica* often acted as a peer-like mediator, helping students identify gaps in coherence and logical progression. As ST4 observed: *“Monica showed me where my ideas didn’t connect, like a classmate pointing out mistakes.”* Such scaffolding echoes Vygotsky’s (1978) concept of the ZPD and is supported by research highlighting AI’s role as a supplementary feedback provider (Mekheimer, 2025; Tran, 2025). Within the SCT framework, *Monica* thus functioned as a mediational tool, extending

learners' capacities to perform beyond their current levels of independent ability (Vygotsky, 1978). These findings align with Mekheimer's (2025) observation that AI-enhanced writing feedback strengthens EFL learners' revision strategies. Interactions with *Monica* also promoted metacognitive reflection. The act of formulating prompts required students to monitor and evaluate their writing more carefully. ST7 explained: "*I had to think carefully about what to ask Monica... it made me notice problems I wouldn't have seen alone.*" Similarly, ST6 highlighted the value of iterative drafting: "*I wrote several drafts, and each time Monica helped me make my sentences clearer and my ideas flow better.*" These practices resonate with the findings of recent studies (e.g., Tran, 2025), which emphasize AI's potential to scaffold reflective practice and recursive drafting. Importantly, such qualitative insights are consistent with the quantitative improvements reported under Research Question One, where AI-supported learners demonstrated significant gains in organization, language use, and mechanics. Finally, several students reported gains in digital literacy and learner autonomy. ST8 reflected: "*When I asked more specific things, I got better help. I learned to 'talk' to it better.*" Others leveraged *Monica* for real-time bilingual support, switching between first language (L1) and L2 to clarify idiomatic expressions and cultural meanings. For example, ST10 described: "*I use it like a smart dictionary. I type a word in Persian and ask for the English meaning and an example sentence.*" Such practices illustrate how AI can bridge L1–L2 gaps and expand students' ability to self-direct practice beyond the classroom. Nonetheless, while these affordances enhanced learner agency, they also introduced pedagogical challenges that warrant critical consideration.

In addition to common technical and ethical concerns—such as privacy issues, security concerns, slow internet connectivity, inadequate infrastructure, high traffic, and occasional inaccuracies in the Iranian context (Parviz, 2024)—several pedagogical and sociocultural limitations emerged from student feedback. A prominent concern pertained to lexical development, particularly the accessibility and usability of AI-generated vocabulary. Although *Monica* frequently suggested sophisticated lexical

items, students often struggled to integrate them effectively into their writing. As participants noted: “*Sometimes Monica uses words I don’t know, so I don’t use them*” (ST11; ST13). This highlights a misalignment between AI-generated input and learners’ productive vocabulary capacity, reinforcing the need for graded, recyclable vocabulary exposure rather than isolated advanced terms, as emphasized by Nation (2019) and Schmitt (2008).

The integration of *Monica* in writing instruction also revealed significant cognitive and metacognitive challenges, primarily related to its potential to foster passive learning behaviors. A notable tendency was students’ overreliance on the tool, using it as a shortcut for obtaining answers rather than as a facilitator of learning. One student admitted: “*I use it to answer the questions, not to learn a lesson*” (ST14). This unreflective reliance mirrors documented behaviors with solution manuals in the Iranian educational context, as captured in a participant’s analogy: “*When I use the app, it feels like checking the solution manual... I get the answer, but I don’t always understand how to write it myself. I make the same mistake again.*” In this sense, AI feedback, like solution manuals, can produce superficially correct outputs that are pedagogically misaligned, underscoring the necessity of teacher mediation. This often led to a “click-and-accept” syndrome, characterized by uncritical acceptance of AI suggestions. One student reflected: “*Sometimes I just accept Monica’s suggestions without thinking, and I’m not sure if it’s always correct*” (ST12). Such passive compliance circumvents crucial metacognitive stages of reflection and error analysis, thereby impeding deeper learning. This finding aligns with Schmitt’s (2008) argument that shallow processing—even when resulting in correct answers—fails to support long-term acquisition. Conversely, some students experienced cognitive overload due to an overabundance of algorithmic options. A participant noted, “*It (Monica) gave me five ways to write one sentence-I didn’t know which to choose*” (ST12). *Monica*’s feedback strategy, which involves providing multiple corrective alternatives for each error, does not appear to account for learners’ proficiency or the nature of the errors (Lin & Crosthwaite, 2024). This illustrates that an overabundance of options can be

as counterproductive as insufficient support, imposing cognitive burdens that hinder decision-making. Collectively, these insights underscore a key instructional design principle: maximizing feedback quantity does not necessarily enhance learning, reinforcing the idea that “more is not always better.”

Finally, sociocultural barriers also influenced AI adoption, particularly concerns regarding legitimacy and potential stigma. Participants expressed apprehension that even using *Monica* for grammar support could be perceived as cheating by peers or instructors. One student explained: “*Some friends think using AI is cheating, even if I just use it to check grammar*” (ST15; ST14). These perceptions reflect broader Iranian educational norms that place high value on academic integrity while exhibiting skepticism toward non-traditional practices. Such concerns align with research indicating that Iranian EFL students often lack a nuanced understanding of plagiarism, which can lead to inadvertent academic misconduct (Mohseni et al., 2023). Moreover, cultural emphasis on face-saving and peer judgment intensified anxiety around tool use, potentially discouraging appropriate engagement with AI-supported learning.

## CONCLUSION AND IMPLICATIONS

This study provides empirical evidence that AI scaffolding, as exemplified by tools like *Monica*, enhances EFL writing proficiency, particularly in organization, language use, and mechanics, with large effect sizes confirming its efficacy. Beyond quantitative gains, the research illuminated the underlying mechanisms of this improvement: by providing immediate, non-judgmental feedback, AI fosters a low-anxiety environment that encourages iterative practice. Notably, even low-proficiency students demonstrated strategic agency and metacognitive development by selectively using AI suggestions, refining prompts, and engaging in recursive drafting-processes that foster reflection and skill internalization. However, the findings reveal a more nuanced picture. While AI excels as a mediational tool within the ZPD

for foundational skills, its limitations in promoting vocabulary depth, original content, and critical thinking underscore the irreplaceable role of the teacher. Challenges such as lexical misalignment, the “click-and-accept” syndrome, and sociocultural concerns about legitimacy further complicate its integration. Therefore, this study advocates for a complementary pedagogical model where AI and human instructors function synergistically. GenAI is optimally deployed for developing lower-order writing skills and building affective-motivational support, while teachers focus on guiding higher-order reasoning, ensuring deep learning, and navigating ethical dimensions. For practitioners, this necessitates designing instruction that promotes reflective AI use, manages cognitive load, and includes explicit integrity training. Future research could investigate the long-term effects of AI assistance on lexical development, explore strategies for integrating AI into collaborative peer settings, and examine cross-cultural variations in adoption. Such work is vital for moving toward nuanced, sustainable, and equitable frameworks for AI integration in language education.

From a pedagogical standpoint, the findings indicate that generative AI tools such as *Monica* can be meaningfully integrated into EFL writing instruction when their use is purposeful, scaffolded, and pedagogically guided. AI appears to be most effective in supporting lower-order writing processes, iterative revision, and affective-motivational engagement, particularly for lower-proficiency learners who benefit from individualized, low-anxiety feedback environments. However, successful integration depends on instructional designs that foster reflective and strategic AI use rather than uncritical reliance on automated suggestions. Teachers therefore play a pivotal role in regulating cognitive load, addressing challenges such as lexical misalignment and the “click-and-accept” tendency, and providing explicit guidance on academic integrity and ethical AI use. In light of these considerations, a complementary instructional model is recommended, wherein AI supports drafting and revision processes while teachers retain primary responsibility for cultivating higher-order reasoning, genre awareness, and critical evaluation skills.

Beyond pedagogical practice, this study contributes to research on second language writing and educational technology by offering empirical support for AI-assisted feedback as a mediational tool within a sociocultural framework. The results suggest that AI scaffolding can operate within learners' ZPD by facilitating lower-order writing skills—such as organization, language use, and mechanics—while simultaneously lowering affective barriers through immediate, non-judgmental feedback. Notably, evidence of selective uptake, recursive drafting, and prompt refinement among lower-proficiency learners indicates that AI-mediated environments can promote strategic agency and metacognitive engagement rather than passive dependence.

At the same time, the limited effects of AI assistance on lexical depth, originality, and higher-order reasoning underscore the constraints of automated scaffolding and reaffirm theoretical distinctions between surface-level accuracy and deeper cognitive–linguistic development. These findings support models of human–AI complementarity rather than substitution, highlighting the continued centrality of teacher mediation in fostering critical thinking, content development, and ethical awareness in L2 writing instruction.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### ORCID

Muhammad Parviz



<http://orcid.org/0000-0002-1449-1651>

Masoud Azizi



<http://orcid.org/0000-0001-9054-1131>

### References

- Allan, D. (2004). *Oxford placement test*. Oxford University Press.
- Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.

- Borna, P., Mohammadi, R., & Nia, R. K. (2024). Investigating the effect of AI writing assistance tools on Iranian intermediate EFL learners' writing performance: A comparative study of ProWritingAid and Grammarly. *Research in English Language Pedagogy*, 12(3), 478–504.
- Braun, V. and Clarke, V. (2021). *Thematic Analysis: A Practical Guide*. Sage, London.
- Chang, Y., Lee, S., Wong, S. F., & Jeong, S.-P. (2022). AI-powered learning application use and gratification: An integrative model. *Information Technology & People*, 35(7), 2115–2139. <https://doi.org/10.1108/ITP-03-2022-0048>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Publishers.
- Ebadi, S., & Amini, A. (2022). Examining the roles of social presence and human-likeness on Iranian EFL learners' motivation using artificial intelligence technology: A case of CSIEC chatbot. *Interactive Learning Environments*, 32(2), 655–673. <https://doi.org/10.1080/14695845.2022.2084947>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03192966>
- Fathi, J., & Rahimi, M. (2024). Utilising artificial intelligence-enhanced writing mediation to develop academic writing skills in EFL learners: A qualitative study. *Computer Assisted Language Learning*, 37(1), 2–40. <https://doi.org/10.1080/09588236.2024.1678901>
- Fereidouni, P. & Farahian, M. (2024). Is ChatGPT a cure for all? Demystifying the impact of using ChatGPT on EFL learners' writing skill. *Applied Linguistics Inquiry*, 2(1), 89–103. <https://doi.org/10.1080/23273798.2024.2345678>
- Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, 3, 100055. <https://doi.org/10.1016/j.caeai.2022.100055>
- Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners'

- writing skill. *Language Teaching Research*, 13621688241239764. <https://doi.org/10.1075/lr.23.001gh>
- Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing*, 54, 100666. <https://doi.org/10.1016/j.asw.2022.100666>
- Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16(3), 148–164. <https://doi.org/10.1016/j.jslw.2007.04.002>
- Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley.
- Jmaiel, H. A., Abukhait, R. O., Mohamed, A. M., Shaaban, T. S., Al-khresheh, M. H., & AL-Qadri, A. H. (2025). The role of ChatGPT in enhancing EFL students' ESP writing skills: An experimental study of gender and major differences. *Discover Education*, 4(1), 1–19.
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, 100450. <https://doi.org/10.1016/j.asw.2020.100450>
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3), 411–433. <https://doi.org/10.1111/j.1468-4446.2004.00027.x>
- Lee, L. (2020). An exploratory study of using personal blogs for L2 writing in fully online language courses. In B. Zou & M. Thomas (Eds.), *Recent developments in technology-enhanced and computer-assisted language learning* (pp. 145–163). Information Science Reference.
- Lin, S., & Crosthwaite, P. (2024). The grass is not always greener: Teacher vs. GPT-assisted written corrective feedback. *System*, 127, 103529. <https://doi.org/10.1016/j.system.2024.103529>
- Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education*, 211, 104977. <https://doi.org/10.1016/j.compedu.2024.104977>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(1), 1–18. <https://doi.org/10.1186/s40594-024-00457-2>

- 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*, 1–17.
- Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: a study on proficiency, revision frequency and writing quality. *Discover Education*, *4*(1), 1–20.
- Mohseni, F., Navidinia, H., & Chahkandi, F. (2024). Unveiling plagiarism practices in Iranian English language students' theses. *Applied Linguistics Inquiry*, *2*(1), 104–113.
- Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake news" is not simply false information: A concept explication and taxonomy of online content. *American Behavioral Scientist*, *65*(2), 180–212. <https://doi.org/10.1177/01650254211020288>
- Moorhouse, B. L., & Wong, K. M. (2025). *Generative artificial intelligence and language teaching*. Cambridge University Press.
- Nation, P. (2019). The different aspects of vocabulary knowledge. In M. C. R. L. Nation & P. C. L. Nation (Eds.). *The Routledge handbook of vocabulary studies* (pp. 15–29). Routledge.
- Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence powered digital writing assistant in higher education: randomized controlled trial. *Heliyon*, *7*, e07014. <https://doi.org/10.1016/j.heliyon.2021.e07014>
- Parviz, M. (2024). "The double-edged sword:" AI integration in English language education from the perspectives of Iranian EFL instructors. *Complutense Journal of English Studies*, *32*, 1–20.
- Pearson Education. (2016). *Top-Notch Series and Complete Assessment Package*. Pearson.
- Rahimi, M., & Fathi, J. (2021). Exploring the impact of wiki-mediated collaborative writing on EFL students' writing performance, writing self-regulation, and writing self-efficacy: A mixed methods study. *Computer Assisted Language Learning*, *35*(9), 2627–2674. <https://doi.org/10.1080/01446650.2021.1978822>
- Ryan, R. M. & Deci, E. L. (2017). *Self-determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. The Guilford Press. <https://doi.org/10.1037/a0039062>

- Richards, J. C. (2015). *Key Issues in Language Teaching*. Cambridge University Press.
- Seyyedrezaei, M. S., Amiryousefi, M., Gimeno-Sanz, A., & Tavakoli, M. (2022). A meta-analysis of the relative effectiveness of technology-enhanced language learning on ESL/EFL writing performance: Retrospect and prospect. *Computer Assisted Language Learning*, 37(7), 1771–1805.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363. <https://doi.org/10.1075/ltr.12.3.02sch>
- 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, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Teimourdash, M. (2024). On the plausibility of integrating synthetic vs. analytic artificial intelligence (ai)-powered academic writing tasks into Iranian EFL classrooms: State-of-the-art. *International Journal of Practical and Pedagogical Issues in English Education*, 2(4), 54–75. <https://doi.org/10.1080/23273798.2024.2345678>
- Teng, F. (2019). The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading. *The Language Learning Journal*, 47(2), 145–158. <https://doi.org/10.1080/01427363.2019.1558945>
- Teng, M. F. (2024). “ChatGPT is the companion, not enemies”: EFL learners’ perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- 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), 2–22. <https://doi.org/10.3390/educsci15020002>
- Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Composition*, 21(2), 217–235. <https://doi.org/10.1016/j.compcom.2003.11.001>

- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wong, K. M., & Mak, P. (2019). Self-assessment in the primary L2 writing classroom. *The Canadian Modern Language Review*, 75(2), 183–196. <https://doi.org/10.1080/00043751.2019.1587645>
- Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28(11), 13943–13967. <https://doi.org/10.1007/s10639-023-10558-9>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70. <https://doi.org/10.1080/00405980209608497>



### Appendix A

#### Themes and Codes: Benefits, Challenges, and Limitations of Integrating Monica in EFL Writing Instruction

Theme	Category	Collapsed Codes	Representative Evidence
<b>Benefits of AI Integration</b>	Affective Regulation	Non-judgmental support; anxiety reduction; emotional consistency; increased confidence	Students described Monica as “always calm,” “never tired,” and less threatening than teachers, reducing embarrassment and encouraging risk-taking (ST5, ST9, ST10).
	Cognitive Scaffolding	Coherence support; peer-like mediation; immediate feedback	AI helped identify logical gaps and organizational problems, functioning as a peer mediator that supported revision (ST4, ST2).
	Metacognitive Engagement	Prompt monitoring; recursive drafting; reflective revision	Formulating prompts and revising iteratively encouraged self-monitoring and awareness of writing problems (ST6, ST7).
	Learner Autonomy & Digital Literacy	Strategic AI use; L1–L2 mediation; self-directed learning	Learners refined prompts, used bilingual support, and extended practice beyond classroom constraints (ST8, ST10).
<b>Challenges and Limitations</b>	Lexical Misalignment	Inaccessible vocabulary; limited uptake of advanced lexis	AI-generated vocabulary often exceeded learners’ productive capacity, leading to avoidance rather than acquisition (ST11, ST13).
	Cognitive & Metacognitive Risks	Overreliance on AI; click-and-accept behavior; superficial learning	Students reported using AI as an answer provider, accepting suggestions uncritically without deep processing (ST12, ST14).
	Cognitive Load	Feedback overload; decision paralysis	Multiple corrective options created confusion and hindered decision-making (ST12).
	Sociocultural Constraints	Perceived cheating; fear of stigma; integrity anxiety	AI use was sometimes viewed as illegitimate or dishonest, discouraging appropriate engagement (ST14, ST15).

*Codes were inductively derived from participant narratives and collapsed into broader analytic categories aligned with affective, cognitive, metacognitive, and sociocultural dimensions of AI-supported EFL writing.*

**Appendix B: Interview Questions**

1. Can you describe your overall experience with AI-assisted writing instruction using *Monica*?
2. In what specific ways did you notice improvements in your writing skills while using *Monica*?
3. How did the integration of AI in writing instruction affect your motivation and engagement in writing tasks?
4. Can you share any advantages or benefits you experienced with AI-assisted writing instruction?
5. Were there any challenges or limitations you encountered when using *Monica* for writing instruction?

