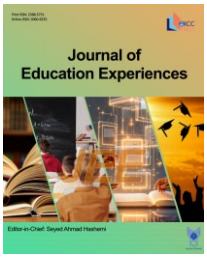

J. Edu. Exp. 2026; 09(01): 54-70

Journal of Education Experiences (JEE)

https://doi.org/10.57647/jee.2026.0901.05

Research Article

Investigating the Impact of AI-Assisted Learning-Oriented Assessment on the Development of Academic Writing Skills Among Iranian EFL Learners: A Mixed Methods Survey

Hassan Alizadeh Mahmoud Alilo[1,*], Yasaman Behravesh[1], Shahla Shahitat[2]

[1]*Department of English Language, Ta.C., Islamic Azad University, Tabriz, Iran*
[2]*Department of English Language and Literature, Institute for Higher Education, ACECR, Khouzestan, Iran*

[*]**Corresponding authors:** hassan.alizademahmoudalilo@iau.ac.ir

Article History:

Received:
18 July 2025

Revised:
28 January 2026

Accepted:
16 February 2026

Published in Issue:
30 June 2026

Abstract

This mixed-methods study investigated the impact of AI-assisted learning-oriented assessment (AI-LOA) on the academic writing skills of Iranian EFL learners. Using a quasi-experimental design, the study compared AI-LOA, teacher-assisted LOA (T-LOA), and traditional instruction across three groups: two experimental (AI-LOA and T-LOA) and one control group. Participants were 60 lower-intermediate EFL learners (aged 15–24) from two institutes in Tabriz, Iran. The English proficiency of participants, who were native speakers of Azari, was measured using the Oxford Placement Test. They completed pre- and post-tests assessed using analytic rubrics and engaged in semi-structured interviews throughout the eight-session intervention. While the experimental groups underwent instruction through LOA, AI- and teacher-mediated, the control group received traditional instruction. Statistical analysis revealed that both experimental groups demonstrated significantly better academic writing performance than the control group, with no significant difference between the AI-LOA and T-LOA conditions. Thematic analysis of interview data identified that AI feedback enhanced learners' independence, motivation, and revision habits, while participants expressed concerns regarding trust in AI responses and potential ethical issues. The findings suggest that structured AI integration within LOA frameworks can be as effective as traditional teacher feedback while encouraging greater learner autonomy. These results have important implications for EFL writing instruction, highlighting the need for guided AI implementation, digital literacy training, and clear ethical guidelines when incorporating AI tools in assessment practices.

Keywords: Academic Writing; Artificial Intelligence; EFL Learners; Learning-Oriented Assessment

Cite this article: AlizadehMahmoudAlilo, H., Behravesh, Y. & Shahitat, Sh. (2026). Investigating the Impact of AI-Assisted Learning-Oriented Assessment on the Development of Academic Writing Skills Among Iranian EFL Learners: A Mixed Methods Survey. *Journal of Education Experiences, 9* (1), 54-70. http://doi.org/10.57647/jee.2026.0901.05

1. Introduction

The incorporation of artificial intelligence (AI) in language education has influenced swiftly how writing skills are taught and evaluated. For EFL learners, particularly those from countries where English is not the main language, academic writing remains rigorous, considering it is hard to comprehend, has rigid guidelines, and they lack the opportunity to see authentic instances of it.

There has been an extensive amount of research on how to instruct writing; however, there still exists a significant gap in our understanding of how to use technology, especially AI-assisted systems, to improve writing outcomes when they are used with LOA. In digital learning environments following the pandemic, there is a heightened requirement to develop instructional frameworks that help students advance and make use of technology. As a domain-specific literacy skill, academic writing requires students to perform tasks including drafting, arranging, evaluating, and revising texts (Mudawy & Mousa, 2017; Singh, 2019). In EFL settings, though, teaching often focuses on surface-level correctness, like grammar, vocabulary, and structure, instead of higher-order skills like argumentation and coherence (Pineteh, 2014). Carless (2007) came up with LOA, which is a promising change because it includes formative feedback, self-assessment, and chances to revise right in the lessons. LOA has worked well in writing classes (Boud, 2000; Zeng et al., 2018), but it can't be used on a large scale because of time, money, and the amount of work teachers have to do, especially in large EFL classes.

Recent advances in AI, especially natural language processing models like ChatGPT, have made it possible to give personalized, timely feedback and encourage independent learning in ways that have never been possible before. Researchers have started to look into how AI can help with L2 writing (Kovalenko & Baranivska, 2024). They have found that students can benefit from automated feedback in areas like fixing grammar and improving vocabulary. But the theoretical connection between AI tools and LOA principles is still not well understood. Also, there is not a lot of real-world evidence about how AI affects the development of academic writing skills, especially when it comes to metacognitive engagement and formative feedback cycles. Some studies praise AI as a tool for building things (Jafarnia et al., 2023), while others warn that it could lead to too much reliance and cheating in school (Farrokhnia et al., 2024; Kasneci et al., 2023).

In the Iranian EFL context, the lack of access to personalized instruction and the students' fear of writing for school make the need for new solutions even greater. Even though many people have access to mobile phones and the internet, digital tools are not yet fully used in formative assessment. This study fills a gap that needs to be filled right now by looking into how AI-LOA can help Iranian EFL students improve their academic writing. The study looks at two main factors: AI-LOA, which is the planned use of AI feedback tools like ChatGPT in writing tasks that help students learn; and academic writing development, which is measured by how well students do on genre-based writing tasks that follow a standardized rubric.

The purpose of this research is to investigate both learning outcomes and learner perceptions by using AI tools in an LOA framework. In theory, it builds on LOA models by observing if they function under AI mediation. In practice, it has implications for language teachers who want to find a balance between new technology and good teaching. What makes this study different is that it focuses on two things: testing how well AI-assisted LOA improves writing and getting students' personal experiences through a mixed-methods design. In this way, the study adds to the growing discussions about how to use AI in EFL instruction in a responsible and educationally sound way.

2. Literature Review

In recent years, the intersection of cutting-edge educational technologies and second-language teaching has become a hot topic. The use of AI in language teaching and testing is changing very quickly. LOA, which combines assessment with learning, has also become more popular, especially for encouraging students to be independent and improve their writing. Academic writing is a very important skill for EFL students in Iran who want to do well in school and at work. But Iranian EFL students often have a lot of trouble with writing (for example, making sure their language is correct, staying organized, and staying motivated) and do not get much help with it (Asaad & Shabdin, 2021; Mallahi, 2024). This section looks at recent research on (1) using AI to teach and test language, (2) assessment that focuses on learning, and (3) improving academic writing in EFL settings. It brings together research, debates, and methods, and it focuses on theoretical lenses that help us understand these areas, like Vygotsky's sociocultural theory, formative assessment theory, and self-regulated learning (SRL).

2.1. AI in Language Assessment and Instruction

AI has made new tools for teaching and testing language. Some examples are automated essay scoring (AES) engines, intelligent tutoring systems, AI writing assistants (like Grammarly or ChatGPT), and adaptive learning platforms. AI-based assessment can quickly handle a lot of written work, which could make it more objective, consistent, and scalable than traditional human grading. For instance, Al-Abbas et al. (2023) say that AI can reach unprecedented levels of objectivity, scalability, and personalization in language testing. These kinds of systems look at things like grammar, vocabulary use, and organization, and then they give scores or feedback automatically. AI writing tools can also give personalized suggestions right away, which

10.57647/jee.2026.0901.05

can help learners find and fix mistakes and make their writing more coherent (Chen, 2023; Zhao, 2022; Liu et al., 2021). Song and Song (2023) say that AI-powered writing apps for mobile devices offer automated feedback on various aspects of writing, such as grammar and sentence structure, which helps people get better quickly. In theory, this kind of feedback is in line with formative assessment because it helps with revision (Zhang & Zou, 2023; Loncar et al., 2023).

There may be benefits based on empirical studies. For example, a mixed-methods study in Bangladesh found that an AI-assisted assessment framework lowered anxiety about foreign languages and increased motivation, even though writing scores were not significantly different from those of traditional assessments (Biju et al., 2024). The group that used AI said they were more interested in learning English and had better attitudes about it. This fits with the idea that AI can make the learning environment more supportive. Song and Song (2023) also looked at work that showed that learners who took AI courses often did better than control groups in terms of achievement and engagement. Despite these benefits, researchers also warn about the current limitations. There are differences between ratings given by humans and AI. In a recent test, ChatGPT gave intermediate EFL essays much lower scores than human raters. This suggests that AI still has not reached sufficient proficiency for high-stakes scoring without further work (Al-Abbas et al., 2023).

There are also worries about algorithmic bias, a lack of openness, and the danger of relying too much on AI (Al-Abbas et al., 2023; Uyar & Büyükahıska, 2025). For instance, human raters gave higher overall scores than ChatGPT on different types of essays, which suggests that AI might not recognize some strengths as well as it should. There is also a lot of debate about the teacher's role. Some people say that AI feedback should add to but not replace teacher feedback (Uyar & Büyükahıska, 2025) and that students need help understanding automated feedback (Sharifzadeh et al., 2025). Overall, the literature shows that AI has both potential and problems when it comes to language assessment. It can help learning by giving quick, personalized feedback (Hwang et al., 2023; Hsiao & Chang, 2023), but there are still questions about its validity and how well it fits into education. The studies use different methods, from systematic reviews (Al-Abbas et al., 2023) to controlled experiments (Uyar & Büyükahıska, 2025; Biju et al., 2024). There is not much research on AI-assisted writing in Iranian EFL settings; most of it is general or comes from other countries. In the future, researchers should test AI tools with Iranian students, look at how cultural factors affect how people accept technology, and look into how using AI in the classroom changes the way things work and how much freedom students have.

2.2. Learning-Oriented Assessment

LOA is a way of thinking that puts a lot of emphasis on how important assessment is for learning (Carless, 2007). LOA is different from just giving tests because it sees assessment tasks as learning activities and gets students involved through self- and peer-assessment. Feedback is used as feedforward for future work. People were worried that traditional testing only led to surface learning, so LOA was created to encourage deeper involvement and metacognition (Carless, 2007). Carless (2007) found three main parts, for example: using assessment tasks as chances to learn, having students evaluate work, and making sure feedback is forward-looking to help students get better. In real life, LOA often looks like repeating cycles of writing, getting feedback, thinking about it, and making changes (William & Thompson, 2007).

Regular formative assessments, self-checklists, peer review workshops, and group projects are all examples of LOA techniques used in language classes. The focus is on making the criteria for judging writing clear and getting students involved in the process. Studies show that LOA can help people write better: Er and Farhady (2023) did a 12-week quasi-experiment with Iranian EFL undergraduates. They found that a writing curriculum based on LOA (with ongoing low-stakes assessments and feedback) led to higher post-test essay scores than a control group. They said that students who were in the LOA condition (for example, doing self- and peer-assessment tasks) made significant improvements in their argumentative writing. Estaji and Safari (2023) also found that Iranian intermediate learners who learned to write using LOA principles (regular diagnostic feedback plus self- and peer-assessment) made a lot of progress in argumentative writing over 16 sessions. Students liked getting feedback from teachers and peers and talking about things in class with this method.

Comparative studies in LOA often look at where feedback comes from. Ghaneiarani et al. (2024) randomly gave Iranian EFL students either peer or teacher feedback during a 12-session writing course. The peer-feedback group had better writing (correct language, good content, and good organization) and remembered what they learned better by the end. This means that having peers review your work, which is a big part of LOA, might be just as good or even better than getting feedback from a teacher. The authors think that peer review in an LOA setting helps students become more independent and practice more.

Formative assessment theory and sociocultural theory are two of the ideas that LOA is based on. William and Thompson (2007) and Stiggings (2005) mention that formative assessment is all about using

tests to help students learn in the future. LOA combines formative and summative elements by making all tests help students learn. For instance, Er and Farhady (2023) indicate that LOA combines assessment, learning, and teaching into one process and encourages learners to learn on their own through feedback/feedforward cycles. SRL (Zimmerman, 2002) is all about students setting goals and judging their work. This fits well with LOA's focus on the learner.

LOA is also based on sociocultural theory (Vygotsky, 1978). Vygotsky believed that social interaction helps people develop higher cognitive functions. LOA practices like peer review and group reflection are in line with this idea. Ghaneiarani et al. (2024) make a clear link between peer feedback and Vygotskian ideas. They say that working with peers on assessment tasks creates a social environment that helps language development. This view says that peer reviewers are more knowledgeable than each other in their Zone of Proximal Development (ZPD), which helps them improve their writing. This means that LOA can be thought of as a way to mediate social interactions during the assessment process (Carless, 2007).

There are still important debates, even though the results were good. Researchers do not all agree on the best way to give feedback in LOA. Some people say that teacher feedback is too one-sided, which could make students less motivated or miss their strengths (Lee, 2019). Some people say that peer feedback is not always accurate if the peers do not know what they're talking about (Topping, 2009). Ghaneiarani et al. (2024) point out that there is no agreement: their study favored peer feedback, but they also point out that different studies use different methods and settings. When it comes to methods, LOA research often uses a mix of questionnaires about attitudes, writing tests before and after, and qualitative feedback (Estaji & Safari, 2023; Ghaneiarani et al., 2024). There aren't many long-term studies that look at how LOA affects writing development over time, and there is not much focus on student factors like motivation and cultural background.

2.3. Academic Writing Development in EFL Contexts

Academic writing in an EFL setting is a difficult skill that is affected by language, thought, and emotion. Students need to learn how to use grammar, register, discourse structure, and critical thinking (Asaad & Shabdin, 2021). Studies have consistently shown that EFL students have trouble with language accuracy (like morphology and syntax) and organizational coherence (like paragraphing and argumentation). Anggraeni et al. (2025) say that some of the most common problems are having a small vocabulary, not being able to connect ideas, and not being able to organize them.

Psychological factors also matter. For example, students often feel anxious about writing, do not believe in themselves, and do not have enough confidence (Mallahi, 2024). In a study of Iranian students, Mallahi (2024) found that a lack of self-efficacy is the cause of many writing problems, like being afraid of getting a bad grade.

So, to help EFL students improve their writing, we need to look at these cognitive and emotional aspects. Iranian EFL students face these global problems as well as problems that are specific to their situation. Aliakbari et al. (2025) say that high-stakes tests and traditional grammar-based teaching in Iran can make it hard to practice real writing.

Changes to the curriculum have put more emphasis on communication skills, but in practice, testing still focuses on exams. Studies in Iran show that there is a need for better ways to teach writing. For example, Esfandiari et al. (2022) used narrative interviews to show that Iranian postgraduate writers want more help and feedback with their academic writing. Ghaneiarani et al. (2024) say that LOA is still new in Iran and needs to be changed to fit the needs of the people there. Because of these cultural differences, new technologies like AI or LOA may need to be put in context (for example, how easily students can get to technology or what teachers believe).

Writing development is all about getting feedback. According to Rostampour and Akbarpour (2025), good feedback from teachers, classmates, or technology helps students find mistakes and fix them. Feedback in LOA frameworks is ongoing and dialogic, not just one-time corrections. Researchers and teachers are starting to agree that feedback needs to be clear, useful, and given promptly (Hyland & Hyland, 2006). AI tools that give automated feedback open up a new way to learn. Early evidence shows that tools like Grammarly or ChatGPT can help with learning by pointing out mistakes and suggesting ways to improve. But students still need to learn how to think critically about AI feedback and not rely on it.

In a nutshell, research on developing EFL writing skills shows that we need support that combines language skills, cognitive strategies, and emotional factors. Recent trends focus on digital tools and giving students more freedom, but there are still gaps. For example, not many studies have looked at how AI-assisted writing tools might work with or against traditional teaching in Iranian EFL settings. In the same way, LOA strategies seem to help with writing, but we need to do more research on how social and cultural factors (like attitudes and institutional constraints) affect their effectiveness. Studies that look at AI use in LOA settings together could be especially useful because the combined effects are not well understood.

10.57647/jee.2026.0901.05

2.4. Theoretical Perspectives

There are a number of theoretical frameworks that help explain the findings above. People often use Vygotsky's sociocultural theory from 1978 to explain both LOA and AI-mediated learning. Vygotsky said that learning happens first between people and then within the person. The ZPD shows what tasks learners can do with help. In LOA, feedback from peers and teachers acts as more knowledgeable others to help writers do better (Ghaneiarani et al., 2024; Rostampour & Akbarpour, 2025). For instance, when students read each other's drafts, they are doing scaffolded reflection, which is similar to what happens in Vygotskian tutorials. Song and Song (2023) build on Vygotskian ideas and apply them to AI. They say that AI chatbots can act like a partner in a group, giving learners immediate support that helps them reach their ZPD. So, AI-assisted writing practice can be seen as a social interaction in which AI acts as a virtual peer, which fits with the sociocultural focus on distributed cognition.

LOA is also based on formative assessment theory. William and Thompson (2007) mentioned that formative assessment, or assessments for learning, is characterized by feedback that helps people figure out what to do next. LOA does this by making assessment a part of the learning process. Carless (2007) and Er and Farhady (2023) say that LOA constantly combines assessment, learning, and teaching. Each draft and comment is not a final decision, but a guide to how to improve. This point of view helps explain why LOA interventions work: students are not only focused on grades, but also on making progress over time (William & Thompson, 2007). It also stresses the importance of having clear criteria and getting students involved, since formative theory stresses openness and self-evaluation (Stiggings, 2005).

Another useful way to look at this is through the lens of SRL theory. According to SRL, students learn better when they set goals, keep track of their strategies, and think about the results (Zimmerman, 2002). Planning an essay, checking for coherence, and making changes to drafts are all parts of SRL in writing. There have been a number of studies that have shown that teaching SRL strategies can help students write better. For example, Anggraeni et al. (2025) found that adding SRL scaffolds to writing lessons, like teaching metacognitive strategies and self-evaluation, made EFL students' academic writing much better. LOA environments tend to encourage SRL because they make students do assessment tasks and use feedback (Estaji & Safari, 2023; Carless, 2007). Lastly, SRL theory helps us understand AI tools: AI feedback can help students regulate themselves by giving them tasks to do (Sharifzadeh et al., 2025). But SRL research also says that students might not automatically use good strategies if they do not get help. So, using AI or LOA with explicit strategy training may give the most benefits.

However, gaps and challenges were also identified. Empirically, there were also problems and gaps found, though. There haven't been many studies that have looked directly at AI-assisted LOA interventions for writing, especially in Iranian EFL settings. Many of the studies that are already out there use short-term interventions and may not keep track of long-term writing growth. There is not much information about how teachers feel about and use AI and LOA innovations in their classrooms, or how institutional factors like curricula and testing policies affect their use. Also, AI can give feedback, but there are still questions about fairness (like who has access to technology) and how to use AI in a holistic way in the classroom.

3. Method

The present study is mixed-methods research (Creswell & Creswell, 2005). In the quantitative phase, a quasi-experimental research design with a pre-test-post-test-control group design was used. This meant that there had to be three groups: two experimental groups (AI-assisted LOA (AI-LOAG) and teacher-assisted LOA (T-LOAG)) and a control group. The experimental groups were treated with new methods of AI-assisted LOA and teacher-assisted LOA, while the control group was taught using the traditional way. The information gathered from the pretests and posttests was used as the research data for the quantitative part of this study. In this study, academic writing is the dependent variable, and AI-assisted LOA is the independent variable. In the qualitative part, the researchers used an approved open-ended semi-structured interview protocol that was made by Kohnke and Zou (2025) concerning the impression and perceptions of introducing AI tools and applications in the realm of teaching and learning a language.

3.1. Participants

The participants were selected from a population of 250 from two different institutes in Tabriz, namely Pardisan and Goldis. The final sample consisted of 60 lower-intermediate EFL learners, all male, aged 15 to 24, who were native speakers of Azari and were recruited from institutes that taught identical curricula. In an attempt to make the comparison between the two institutions valid, both institutions followed and pursued similar syllabi in instruction, hired similar instructors who were trained to the same academic rigor, and taught in similarly equipped classrooms. The standardization that occurred between the two contexts involved the weekly number of contact hours, assessment procedures, and

10.57647/jee.2026.0901.05

instructional materials, and this made all the contexts methodologically consistent and reduced institutional variability. This technique ensured that study subjects had similar academic backgrounds. Participants were selected through a convenience sampling method because of practical limits like accessibility and the willingness to participate. This approach guaranteed that the participants chosen would be representative of the typical population of lower-intermediate EFL students at institutes, as they were exposed to similar educational backgrounds, standard of language proficiency, and stages of learning. According to the placement criteria of the institutes, they were lower-intermediate students. Nonetheless, to guarantee the integrity and uniformity of the participants, a proficiency test was administered before the commencement of the primary research. Based on the proficiency test, the candidates who scored between 40 and 47 were included in the study and rated as proficient. The participants were then non-randomly assigned to three groups that included two experimental groups (AI-LOAG and T-LOAG) and a control group. To protect anonymity, each participant was assigned a numerical code. The final sample was composed of 20 participants of the Pardisan Institute AI-LOA experimental group and 40 participants of the Goldis Institute: 20 of the LOA experimental group and 20 of the control group. Originally, after conducting the proficiency test, there were 26 homogeneous students in the Pardisan Institute class, but to equalize their numbers with those of the Goldis Institute, six participants were excluded from the main study. These six students were not put into the treatment stage but received routine training by the institute and were not subject to any experimental manipulations or tests. In order to remove the confounding variables of teachers, all groups were instructed by the same teacher. The participants had studied Evolve 1-3. They further taught using the Evolve 4 textbook during the study to maintain consistency in the teaching content and exposure across all of the groups.

3.2. Instruments

To gather the data needed for the study, the researchers applied the following instruments at various stages of the study.

3.2.1. *Oxford Placement Test*

The Oxford Placement Test (OPT), developed by Dave (2004), was systematically applied to assess and verify if the proficiency levels of the English language differed in any significant ways between the experimental and control groups investigated. This is a well-known and standard examination test that properly and effectively ascertains the knowledge of language at various levels of the Common European Framework of Reference for languages (CEFR). As an evaluation instrument, the OPT is considered one of the main characteristics since it serves as the homogenizing tool, which measures a variety of linguistic skills such as grammar, vocabulary, and reading comprehension. The test is a well-structured formal evaluation diluted to six levels of proficiency on the CEFR scale and assigns test scores to well-defined value boundaries for each of the discrete levels: Basic (A1: 0–17), Elementary (A2: 18–29), lower intermediate (B1: 30–39), upper intermediate (B2: 40–47), advanced (C1: 48–54) and very advanced (C2: 54–60). To ensure that all participants fell strictly within the upper-intermediate (B2) band and not the adjacent lower-intermediate (B1), only those whose scores ranged from 40 to 47 were included in the study. This range of scores was a strict cut-off point; any learner who scored 48 or below 40 was left out of the sample. This process ensured that the participants had the same and acceptable standardized level of proficiency that matched the CEFR B2 standard. Those OPT results obtained at the beginning of the research were crucial in the sense that they allowed researchers to place participants whose results would be placed at the upper intermediate level on the scale intentionally to ensure that the language proficiency level was identical in the groups. Moreover, the psychometric characteristics of the OPT have been tested in previous studies that report high reliability and construct validity of the tool in the EFL environments. Normally, the internal consistency of the grammar part of the OPT and the listening part of the test is more than 0.85 Cronbach's alpha (Geranpayeh, 2003). The results of the current study showed that the Cronbach alpha of the OPT on a pilot sample (n = 20) was 0.87, which indicates high internal reliability of this group of Iranian EFL students. In addition, the test was CEFR-based, making the process of decisions made on the placement content valid.

3.2.2. *Academic Writing Tests*

During the data collection procedure, participants were required to write two academic essays (pre-test and post-test). The pre-test was done one session before the treatment began, and the post-test was performed after the eight-session intervention was finished. The main purpose was to assess the academic writing performance of students taking part in the intervention. That was done in line with the suggestion of Jacobs et al. (1981), who recommended that researchers should consider at least two sets of writing performance from every participant before and after the treatment to conclude a reliable stance of the learners' performance. The essays were scored using the rubrics advised by Wang (2024). The

10.57647/jee.2026.0901.05

writing topic was selected from NTC's TOEFL materials and was carefully adapted for upper-intermediate learners to ensure accessibility in terms of vocabulary, structure, and task demands. Although originally designed for higher proficiency levels, TOEFL-style prompts were used to expose learners to authentic academic tasks while maintaining an achievable level of complexity. The adapted version allowed participants to demonstrate key academic writing skills, such as stating a position, providing reasons, and organizing ideas coherently, within a manageable linguistic range. The prompt was reviewed by two EFL instructors to confirm its appropriateness for the learners' level and was then shared with all participants. This prompt formed the basis of an essay task that participants were to carry out before and after the intervention, wherein they wrote an essay of about 150-250 words. Fifty minutes was the time allotted for each session. Inter-rater reliability was used to maintain the validity of the assessment process. To do this, the essays were rated by two independent raters to confirm the consistency of the scoring. Raters were trained to use Wang's (2024) rubric through calibration sessions. High agreement between raters indicated reliable scoring. In cases of substantial disagreement, discussions were conducted between the raters to resolve differences and achieve consensus.

3.2.3. *Face-to-face Interview*

The current study used a semi-structured interview format, which is a good mix of consistency and flexibility. This format permits the interviewer to use a set list of open- and closed-ended questions, and it also permits them to clear up any confusion and ask more questions based on what the participants said. Dörnyei and Taguchi (2009) say that open-ended questions add to the data and let people be spontaneous, especially when the range of possible answers can't be predicted ahead of time. The semi-structured interview protocol for this study was made in line with the research questions and based on what had already been written about the topic. The main goal was to find out what Iranian EFL students thought about using AI-powered assessment tools for writing lessons. The interview questions were obtained and adapted from Kohnke and Zou (2025). Two experts in applied linguistics and educational technology looked over the protocol and suggested rewording some parts to make them clearer and less vague.

A pilot test with two teachers led to smaller changes in the order of the questions. Here is the final version of the interview checklist:

1. How much do you know about ChatGPT? What do you usually use ChatGPT for?

2. What are your teachers' attitudes towards ChatGPT? For example, do you have any experience where teachers encouraged the use of ChatGPT in a specific class? Can you provide me with examples?

3. Could you please briefly describe your experience with ChatGPT in language learning? Can you give me one specific example?

4. Compared with how you studied English before, what do you think are the particular strengths of ChatGPT in language learning? Can ChatGPT help/facilitate you to study independently?

5. What specific language skills do you think ChatGPT may provide more support with?

6. Could you provide examples of effective prompts you have used when interacting with ChatGPT?

7. How can you tell if the information ChatGPT gives you is correct or trustworthy?

8. What problems have you run into while using ChatGPT? What did you do to fix problems when you used ChatGPT in the classroom?

9. What are the main problems with using ChatGPT?

10. What do you think about the possible problems that could come up with academic honesty and moral issues?

11. How can you tell if the information ChatGPT gives you is correct or trustworthy?

12. What problems have you run into while using ChatGPT? How did you fix problems that came up when you used ChatGPT in class?

13. What are the main problems with using ChatGPT?

14. What do you think about the problems that might come up with academic honesty and moral issues?

Interviewers need to know enough about the topic of the interview to have meaningful and informed conversations. One of the researchers in this study, who acted as an interviewer, tried to stay objective while collecting and analyzing the data. However, it is impossible to be completely objective in qualitative research because the results are created through the interactions between the interviewer and the participants. The researchers' backgrounds, beliefs, and assumptions will always affect how they understand the data. Interviewer bias is more likely to happen in face-to-face interviews, so the way the interviewer and interviewee interact with each other can have a big effect on the answers given. The interviewer first tried to record the interviews on audio tape so that they could be transcribed and analyzed more accurately and without bias. However, a few participants did not want to be recorded because they were worried about their privacy. In these situations, the researchers wrote down what the participants said in real time by hand using a pre-made interview checklist. To save time and make sure they were comfortable, some participants chose to write their

answers directly on the interview forms. While these changes were necessary, they were carefully planned so that the data collected would still be complete and accurate. Also, to reduce bias, the researchers thought carefully about their positions. The researchers were TEFL experts who had used AI tools in Iranian universities and language schools before. To reduce confirmation bias, they kept reflexive journals while they were collecting data, writing down their assumptions (for example, their initial belief that AI would make work easier for everyone). Peer debriefing with two outside qualitative researchers made sure that the analysis was even more rigorous (Guba & Lincoln, 1994).

3.3. Data Collection Procedure

This study used a quasi-experimental design with a pre-test and a post-test to look at how AI-LOA affected the academic writing of Iranian EFL students. The process was based on strict, trustworthy, and moral principles. It took place in a set order from November 2024 to January 2025 at the Pardisan and Goldis Language Institutes in Tabriz, Iran. The Ethical Committee of the Pardisan and Goldis Language Institutes approved the study. They were in charge of making sure that all ethical research issues were taken care of before the study began. Recruitment took place during regular class times. The researchers explained the study's purpose, how it would be done, and that it was voluntary. They also promised that participants would remain anonymous and that they could leave the study at any time without any problems. All subjects gave their informed written consent by signing printed forms that were returned and kept in a locked filing cabinet that the researchers could access.
Before the main data collection, a pilot study was done to make the research tools used clearer and more reliable. Twenty students who were at the same level of proficiency as the main sample took the OPT and academic writing test. The goal of this pilot study was to find the unclear parts of both instruments, figure out how long it would take to fill out both scales, and check the accuracy of the Persian version of the instructions for both tools. The pilot study showed that both tools were very reliable, which meant that they could both be used in the main study.

Two tests were used to determine baseline equivalence. First, the OPT was given, and it was in a 60-minute session with all learners present to make sure everyone was on the same page. Researchers collected answer sheets by hand and scored them using the official scoring key. To keep things consistent, only participants with scores of 40 to 47 (upper-intermediate, B2) were included. There were originally 250 male students, but only 60 were kept. Some were not kept because their scores were outside of this range. A colleague double-checked the scores for accuracy after they were entered by hand into a spreadsheet. There were three groups of participants: AI-LOAG (n = 20, Pardisan), T-LOAG (n = 20, Goldis), and control (n = 20, Goldis). This non-random assignment made sure that all the institutions were the same. To account for differences between teachers, all groups were taught by the same teacher, who spoke both Azari and English fluently.

Before the treatment, we gathered baseline information about how well the students wrote in school. First, there was an 80-minute writing test that served as a pre-test. The question for the essay was, "*Do you agree or disagree with the statement that only students should be able to use ChatGPT?*" People wrote between 150 and 250 words. The topic was chosen on purpose because it was something that everyone could relate to right away. They were all EFL students who had learned about or seen AI in school. The topic was interesting and relevant, which made students think critically and use their own experiences and opinions on AI tools in school. Two independent raters used Wang's (2024) analytic rubric to score the essays, which were written by hand on standardized answer sheets and collected by hand. The rubric looked at grammar, vocabulary, organization, and content. We wrote down the scores by hand on a scoring sheet and then figured out how reliable they were between raters. We talked about any differences to make sure everyone understood. It was printed out and put into the spreadsheet.

The AI-LOAG group learned how to use artificial intelligence tools, mostly ChatGPT, in a structured LOA framework over the course of eight 90-minute sessions. The same steps were followed in each session: input, practice, feedback, and revision. The students started by writing about academic topics that they had found and changed to fit the TOEFL prompts. After brainstorming and outlining, they wrote first drafts and sent them to ChatGPT using pre-made prompts from the teacher that were meant to get academic-level responses (for example, Evaluate the argument with examples and appropriate cohesion). After turning in their work, students got AI-generated feedback on how well they organized their ideas, used grammar, varied their vocabulary, and made their points clear. The teacher taught the students how to critically look at AI responses, tell the difference between surface-level and deeper writing problems, and make changes as needed. There were times for students to talk to each other about their AI feedback and think about how to make changes during each session. Mini-lessons in class dealt with common mistakes that students made in their writing, based on patterns found in AI feedback. The teacher helped the students use the AI tool in a way that encouraged metacognitive strategies like setting goals,

keeping track of progress, and thinking about what they had done, all within the LOA framework. We kept track of our progress every week by using self-assessment checklists and drafts that the teacher marked up to see how our writing skills were getting better.

The T-LOAG group had the same eight sessions and curriculum themes as the AI-LOAG group, but instead of an AI tool, they got feedback directly from the teacher. At the start of each session, there was a short lecture or demonstration of writing strategies that were related to TOEFL themes, such as how to write a thesis, develop a paragraph, and use coherence devices. Then, the students wrote answers to the same questions that were used in the AI-LOAG. After writing, students turned in their drafts to their teachers for feedback. The teachers then gave the students feedback in the next session based on the same criteria used in the AI-LOAG: content, organization, cohesion, vocabulary, and grammar. Feedback focused on what students did well, what they could do better, and how to rewrite their work. This helped students understand the standards for good writing and good performance. Every other week, the teacher also held short one-on-one meetings with students to explain feedback and help them come up with ways to revise their work. In later sessions, peer assessment activities were added so that students could use rubrics to evaluate each other's writing. Weekly reflective logs helped students keep track of their progress and figure out their writing habits and problems that kept coming up. The whole way of teaching was based on the ideas of LOA, which stresses assessment for learning rather than assessment of learning and gives students more control over their learning through structured feedback loops, guided practice, and continuous improvement.

The control group, on the other hand, learned how to write in a traditional way without using AI tools or learning-focused assessment methods. Using the TOEFL textbook, lessons were centered on the teacher and followed a presentation-practice-production (PPP) model. Students worked on writing assignments on their own, and the whole class sometimes got general feedback without any individual formative assessment. There was no focus on revision, self-assessment, or peer feedback. Instead, the lessons focused mostly on grammar, vocabulary, and how well the structure worked.

The instructor, who was also the first researcher, had formal teaching credentials as an appointed teacher under the Ministry of Education in Iran. All groups followed the same Evolve 4 curriculum and session length. He was a good teacher for online EFL courses that included tests, so he was able to make sure that all groups got the same lessons. His background allowed him to use the same teaching methods over and over again and run the AI-LOA and LOA procedures fairly, making sure that the effects of AI-LOA and LOA interventions were kept separate from each other.

After the treatment, data were collected in the same way as the pre-tests, and inter-rater reliability was calculated. Any differences were talked about to make sure the post-test was consistent. We kept all the data and made it anonymous by giving each participant a code (like AI-LOAG-01, T-LOAG-01, or CG-01). To keep outside factors to a minimum, all of the institutes used the same classroom conditions (lighting, seating, etc.) and times (morning sessions). We used SPSS 30 to analyze the data. We compared the scores from the pre- and post-tests using Analysis of Covariance (ANCOVA) and Analysis of Variance (ANOVA) to make sure the results were statistically sound.

The qualitative part of the study was meant to profoundly improve understanding of AI-assisted instruction and its effects on Iranian EFL learners' academic success. The researchers used the accredited open-ended interview protocol created by Kohnke and Zou (2025) for the introduction of AI chatbots in learning a language. The sample size was based on data saturation, which is when no new themes came up in interviews that followed each other (Saunders et al., 2018). After 15 people took part, saturation was reached, which is in line with what is recommended for phenomenological studies. We transcribed and analyzed 15 semi-structured interviews using Braun and Clarke's (2006) thematic analysis to find patterns in how people thought about AI scaffolding. Figure 1 shows how the researchers in this study collected and analyzed qualitative data. The second phase of the research interviews took place at the same time as the first to make data collection as efficient as possible. The researchers used semi-structured interviews to gather qualitative data in an inductive way. After that, the data were systematically coded and grouped into new themes. This made it possible to find larger patterns that could help with the development of theory (Creswell et al., 2003). Also, parts of grounded theory methodology were used to try to come up with a theory based on the data given by participants (Long et al., 1993).

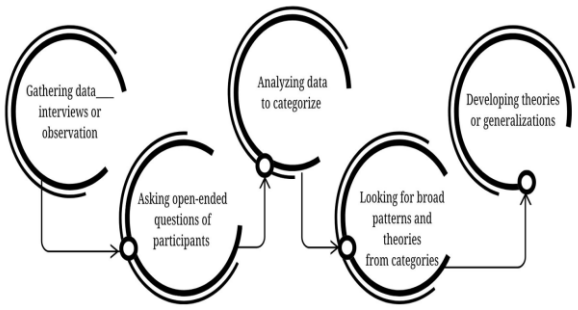

Figure 1. Inductive Logic Used in Qualitative Studies (adopted from Creswell et al., 2003)

10.57647/jee.2026.0901.05

At the start of each interview, there was a short summary of the study's goals and the reasons for doing the research to make sure the participants understood and were interested.

The researchers told the participants that they were interested in the topic to build trust and encourage open communication. The conversation was guided by a set of pre-planned interview questions, but participants were free to go into more detail, which led to richer data and a better understanding of their experiences with AI-LOA. By strictly following Guba and Lincoln's (1994) qualitative rigor criteria of credibility, dependability, confirmability, and transferability, the study's data collection tools were shown to be trustworthy. This made sure that the semi-structured interviews were reliable and valid. Triangulation with 15 interviews, member checking with six participants, and peer debriefing by two experts all helped to make the study more credible. It was possible to rely on the results by adapting validated checklists for the Iranian EFL context, testing the tools with students, and keeping a record of methodological decisions that were reviewed by experts in the field. Reflexive memos, intercoder reliability (88% agreement, Kappa = 0.85), and the use of in vivo coding and verbatim quotes to ground interpretations in participants' voices were all ways to make sure that confirmability was taken into account. Finally, transferability was improved by giving a lot of information about the demographics of the participants, the types of institutions they were in, and the regions they were in. This helped readers see how the findings could be applied to other EFL settings.

3.4. Data Analysis

In the quantitative phase, concerning the first research question, the collected data were entered into SPSS 30 for further statistical analysis. At first, the OPT scores were checked to see if the two groups were similar. The researchers showed descriptive statistics for both the pre-test and post-test scores, such as the mean, standard deviation (SD), and standard errors (SEs). They used the Pearson correlation coefficient to check how reliable the two raters were when they rated the same thing. The researchers used Levene's Test of Equality of Error Variances to see if the data were normally distributed. When the data were normal, ANCOVA and ANOVA were used to look at how the independent variables affected the dependent variables.

The main way to interpret the data was through thematic analysis in order to answer the research questions. We used member checking and low-inference interpretation to make sure that the thematic analysis was reliable and thorough, following the rules for qualitative research (Guba & Lincoln, 1994; Braun & Clarke, 2006). Member checking meant sharing synthesized thematic summaries with six randomly chosen participants to make sure that the interpretations were correct and matched what the participants meant. This step took place about two weeks after the first data collection, and summaries were sent in Persian via email to make sure everyone could read them. Participants were asked to confirm or clarify the themes, which led to changes. Creswell and Creswell (2005) say that this iterative feedback loop made the results more real. Researchers wanted to avoid bias and base their conclusions on what the participants said, so they used low-inference interpretation. We used in vivo labels for 70% of the first codes to keep the participants' language and lower the chance of misinterpretation (Braun & Clarke, 2006). There were a lot of verbatim quotes in the results to back up the themes with raw data. A bilingual applied linguistics expert checked the English translations for accuracy. As Guba and Lincoln (1994) suggested, reflexive journaling was used throughout the analysis to keep track of assumptions (like the initial belief that all participants would resist AI) and how they changed over time. Weekly peer debriefing sessions with two outside qualitative researchers (one an expert in EFL and the other in educational technology) helped reduce bias even more. They coded 30% of the transcripts on their own, and there was 88% agreement between coders (Kappa = 0.85). Consensus discussions were used to settle differences, which ensured that the analysis was thorough. These strategies worked together to make sure that the thematic findings were trustworthy, clear, and closely connected to the participants' real-life experiences.

Table 1. Descriptive Statistics of Academic Writing Pre-and Post-test Scores

					95% Confidence Interval for Mean			
	N	Mean	SD	SE	Lower Bound	Upper Bound	Min	Max
Pre-AI-LOAG	20	58.45	2.502	0.559	57.27	59.62	54.00	63.00
Pre-T-LOAG	20	56.65	2.796	0.625	55.34	57.95	51.00	61.00
Pre-CG	20	58.15	2.183	0.488	57.12	59.17	54.00	62.00
Post-AI-LOAG	20	95.40	3.067	0.685	93.96	96.83	91.00	101.00
Post-T-LOAG	20	95.50	2.910	0.650	94.13	96.86	90.00	100.00
Post-CG	20	58.45	2.981	0.666	57.05	59.84	53.00	63.00

10.57647/jee.2026.0901.05

4. Results

4.1. Results for the First Research Question

To answer the first research question, we first used an ANOVA to find out if there were any differences in the pretest scores for academic writing between the three groups. This first look at the groups made sure they were similar before the intervention. Table 1 shows the descriptive statistics for the groups before and after the test. Table 1 demonstrates that there were small differences in the pre-test scores between the groups. The AI-LOAG had a mean score of 58.45 (SD = 2.502, SE = 0.559), the T-LOAG had a mean score of 56.65 (SD = 2.796, SE = 0.625), and the control group had a mean score of 58.15 (SD = 2.183, SE = 0.488). The T-LOAG had the lowest mean and the most variability. The post-test scores, on the other hand, showed that the experimental groups had higher means: the AI-LOAG had a mean of 95.40 (SD = 3.067, SE = 0.685), the T-LOAG had a mean of 95.50 (SD = 2.910, SE = 0.650), and the control group had a mean of 58.45 (SD = 2.981, SE = 0.666).

This means that there was not much difference between the AI-LOAG and T-LOAG, but there was a big difference between the control group and the other two groups. These descriptive statistics show that both AI-LOA and T-LOA instructions were very important for helping students improve their academic writing. Furthermore, the Pearson correlation coefficient was utilized to check the reliability of the ratings and see how consistent the two raters were with each other. Table 2 shows the steps for these analyses. Table 2 shows that the inter-rater correlation for the control group was almost perfect for the pretest scores, with r = 0.945 ($p < 0.001$), which means that the scores were very consistent. The pretest scores for the experimental groups were also very reliable (r =0.945, $p < 0.001$), which shows that the raters used the same scoring criteria for both groups at baseline before the intervention. The scores after the test showed even more agreement, with correlations rising to r = 0.993 ($p <0.001$) for the control and experimental groups, respectively. Given this high level of agreement between the two raters, we can say that they both gave the same level of accuracy and objectivity when judging how well the participants did on the pre- and post-test. Also, Levene's Test of Equality of Error Variances was used to see if the error variance of the dependent variables was the same for all groups. These results are shown in Table 3.

The test for homogeneity of variance for the academic writing pre-test gave a $F(2, 57) = 0.768$, $p = 0.469$, which is higher than the .05 threshold, as shown in Table 3. This means that the assumption of equal error variance across groups is met, which means that ANOVA is a good way to look at the dependent variables. The academic writing post-test did not show a significant difference (p =0.462), which means that the differences in post-test scores were about the same across groups. This means that ANCOVA was appropriate. Table 4 shows the ANOVA results for the pre-test scores for the three groups: AI-LOAG, T-LOAG, and the control group.

Table 2. Inter-Rater Correlation for the Academic Writing Test Scores

		Rater 1	Rater 2
Pretest of Control Group (Rater 1)	Pearson Correlation	1	0.945**
	Sig.(2-tailed)		0.000
	N	20	20
Pretest of Experimental Groups (Rater 2)	Pearson Correlation	0.945**	1
	Sig.(2-tailed)	0.000	
	N	40	40
Posttest of Control Group (Rater 1)	Pearson/Correlation	1	0.993**
	Sig.(2-tailed)		0.000
	N	20	20
Posttest of Experimental Groups (Rater 2)	Pearson/Correlation	0.993**	1
	Sig.(2-tailed)	0.000	
	N	40	40

**. Correlation/is/significant at the 0.01 level (2-tailed)

Table 3. Levene's Test of Equality of Error Variances in Writing Scores

Tests	Levene Statistic	df1	df2	Sig.
Academic Writing Pre-test	0.768	2	57	0.469
Academic Writing Post-Test	0.784	3	57	0.462

Tests the null hypothesis that the error variance of the dependent variable is equal across groups

10.57647/jee.2026.0901.05

Table 4. Results of ANOVA Analysis for Pre-Test Scores

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	37.200	2	18.600	2.961	0.060
Within Groups	358.050	57	6.282	-	-
Total	395.250	59	-	-	-

Table 5. Tukey HSD Pairwise Comparisons for Pre-Test Scores

					95% Confidence Interval	
(I) group	(J) group	MD (I-J)	SE	Sig.	Lower Bound	Upper Bound
AI-LOAG	T-LOAG	1.800	0.792	0.068	-0.1072	3.7072
	CG	0.300	0.792	0.924	-1.6072	2.2072
T-LOAG	AI-LOAG	-1.800	0.792	0.068	-3.7072	0.1072
	CG	-1.500	0.792	0.150	-3.4072	0.4072
CG	AI-LOAG	-0.300	0.792	0.924	-2.2072	1.6072
	T-LOAG	1.500	0.792	0.150	-0.4072	3.4072

Table 6. ANCOVA Results: Effects of AI-LOA and T-LOA on Post-Test Scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	18261.146[a]	3	6087.049	680.338	0.000	0.973
Intercept	597.058	1	597.058	66.732	0.000	0.544
Pre-Test	7.713	1	7.713	0.862	0.357	0.015
group	18108.970	2	9054.485	1012.003	0.000	0.973
Error	501.037	56	8.947	-	-	-
Total	433265.000	60	-	-	-	-
Corrected Total	18762.183	59	-	-	-	-

a. R Squared = 0.973 (Adjusted R Squared = 0.972)

As shown in Table 4, the ANOVA yielded an F-value of 2.961 with 2 degrees of freedom between groups and 57 degrees of freedom within groups, with a significance level ($p = 0.060$). The result is not statistically significant because the p-value is higher than the usual alpha level of 0.05. This means that there were no significant differences in the academic buoyancy levels of the three groups before the intervention. This suggests that all three groups started from a statistically similar level of academic writing. These results back up the idea that the groups were equal before the test, which makes the comparisons on post-test measures more valid. Further pairwise comparisons using the Tukey HSD post hoc test, as shown in Table 5, support this conclusion. This is close to the 0.05 level, but it is still not statistically significant. The other comparisons, AI-LOAG vs. Control (MD = 0.300, p = 0.924) and T-LOAG vs. Control Group (MD = -1.500, p = 0.150), also did not show any significant differences, since all of them had confidence intervals that included zero. The ANOVA and Tukey HSD results both show that there were no big differences in pre-test scores between the AI-LOAG, T-LOAG, and control groups. These results show that the groups were statistically the same at the beginning based on their pre-test scores. An ANCOVA was also done on AI-LOA and T-LOA's effectiveness in academic writing skills to answer the second research question. This was done because it controlled for any differences that may have existed before the test in academic writing scores and gave a more accurate comparison of the three groups' post-test results. The results of the ANCOVA are shown in Table 6. The ANCOVA results in Table 6 show that the way the test was given had a big effect on academic writing ($F = 1012.003$, $p < 0.001$). The partial eta squared value of 0.973 shows that the effect size is strong, which shows that the type of LOA had a big impact on how well students could handle academic challenges. Table 7 shows the adjusted post-test scores, which take into account the differences in pretest scores that were already there as a covariate. The mean pretest score was 57.75. The AI-LOAG and T-LOAG comparison had a mean difference of 1.800 and a p-value of 0.068. Table 5 shows that none of the group comparisons showed differences that were statistically significant. Table 7 shows that the adjusted post-test means for the AI-LOAG, T-LOAG, and control groups are 95.297, 95.661, and 58.391, respectively. The standard errors for these groups are 0.678, 0.691, and 0.672, respectively. The adjusted means show the differences in the pretest scores, where the T-LOAG had the lowest mean (56.65), followed by the control group (58.15) and the AI-LOAG (58.45).

Table 7. Estimates of Adjusted Post-Test Scores, Accounting for Pretest Scores

group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
AI-LOAG	95.297[a]	0.678	93.939	96.655
T-LOAG	95.661[a]	0.691	94.277	97.046
CG	58.391[a]	0.672	57.045	59.737

a. Covariates appearing in the model are evaluated at the following values: Academic Writing Pre-test = 57.7500

Table 8. Comparisons of AI-LOA and T-LOA Effects on Post-Test Scores

					95% Confidence Interval for Difference[b]	
(I) group	(J) group	MD (I-J)	SE	Sig.[b]	Lower Bound	Upper Bound
AI-LOAG	T-LOAG	-0.364	0.988	1.000	-2.802	2.074
	CG	36.906*	0.947	0.000	34.569	39.243
T-LOAG	AI-LOAG	0.364	0.988	1.000	-2.074	2.802
	CG	37.270*	0.975	0.000	34.863	39.677
CG	AI-LOAG	-36.906*	0.947	0.000	-39.243	-34.569
	T-LOAG	-37.270*	0.975	0.000	-39.677	-34.863

*. The mean difference is significant at the 0.05 level

b. Adjustment for multiple comparisons: Bonferroni

Moreover, Table 7 shows that the small differences in academic writing that were already there were controlled for by using the pretest scores as a covariate. This means that the post-test results mostly show the effects of the AI-LOAG and T-LOA interventions, not the differences that were already there. Table 8 shows how the post-test scores for academic writing compare between groups. Table 8 shows that there was no statistically significant difference between the AI-LOAG and T-LOAG (MD = -0.364, p = 1.000). This means that both methods were equally good at helping students write academically. The AI-LOAG and T-LOAG groups did, however, much better than the control group (p <0.001), with mean differences of 36.906 and 37.270, respectively. These results show that both assessment methods greatly improved students' ability to handle academic challenges compared to regular teaching.

4.2. Results for the Second Research Question

The second half of this section shows what we found from a thematic analysis of semi-structured interviews. It is organized around the study's second research question and shows how Iranian EFL learners feel about using AI-LOA to teach academic writing. The data were written down in Persian, reviewed multiple times, and coded semantically to find themes. These themes were then refined by constantly comparing them to find the most important patterns. To make sure the research was rigorous, six participants checked the themes for accuracy, and low-inference interpretation focused on the participants' exact words, with long verbatim quotes to reduce bias and add depth to the story. We translated important themes and quotes into English and kept the names of the people who took part secret (for example, P3 and Q12). More quotes have been added to make the participants' points of view clearer, making sure that the findings are based on their voices. Four main themes came up in the thematic analysis of the semi-structured interviews with 15 Iranian EFL learners about how they feel about AI (specifically ChatGPT) in academic writing instruction:

Theme 1: Perceived Usefulness of AI Feedback

Most students said that ChatGPT's feedback, which included corrections and explanations, helped them improve their grammar, vocabulary, and sentence structure. They liked how quickly and clearly the answers came, which often helped them understand the rules of language better. As P2 explained, "*I always had problems with grammar. When I use ChatGPT, it corrects my mistakes and tells me why. I think I am learning faster this way.*" P7 described ChatGPT as a supportive presence: "*When I write something and put it in ChatGPT, it shows me a better way to say it. It's like having a teacher all the time*." These comments show that learners see AI as more than just a way to fix mistakes; they see it as a way to get personalized help that helps them improve.

Theme 2: Increased Autonomy and Motivation

AI-assisted assessment seemed to help students become more independent and interested in writing. Some participants mentioned they felt empowered and in charge of their own learning. P4 shared, "*Before, I didn't like writing. Now I feel more confident because I can check my writing first with AI.*" P13 echoed this by stating, "*I don't need to wait for my teacher anymore. I*

10.57647/jee.2026.0901.05

can write and improve by myself." These findings show that ChatGPT helped students go from passively receiving feedback to actively engaging with writing as a process that repeats itself. Some even noted writing more frequently at home, as P1 said, "*Now I write more at home because ChatGPT helps me. I didn't do that before.*"

Theme 3: Trust and Skepticism Toward AI Responses

Most learners trusted AI-generated suggestions, but some were worried about how reliable they were, especially when it came to tone or content development. P8 mentioned, "*Sometimes the sentences it gives me are too formal or not natural. I have to change them a lot*." P5 recounted a specific experience, "*I don't always trust the information. One time it gave me a wrong example, so now I check with my teacher if I'm not sure*." These comments show that the person has a deep understanding of what AI can and can't do. Students seemed to be getting better at digital literacy by learning when to trust or question AI output. This shows that their language learning skills are getting better.

Theme 4: Ethical Concerns and Misuse Potential

A few participants who took part brought up worries about how the technology could be misused, like copying AI-generated text without making changes or thinking about it. P12 observed, "*Some of my friends just copy the answer from ChatGPT and submit it. They don't even read it.*" P6 added, "*It's useful, but only if you use it the right way. If you just paste the answer, it's cheating*." These statements show that people think there is a conflict between using AI as a learning tool and using it as a shortcut. Students clearly wanted teachers to help them make sure they used it correctly. According to P11, "*Teachers should teach us how to use it. We don't always know if it's okay to use ChatGPT for everything*." These results show that Iranian EFL students generally see AI tools like ChatGPT as helpful tools for writing essays, especially for self-directed learning and revision. However, to use them well, you need to be digitally literate, aware of ethics, and have a teacher to help you find the right balance between freedom and academic honesty.

5. Discussion

The current study investigated how AI-assisted LOA affected the academic writing skills of Iranian EFL learners using a mixed-methods approach that combined quantitative results with qualitative insights into how learners felt about the experience. The numbers showed that both AI-LOAG and T-LOAG made participants' academic writing skills much better than the control group that got regular instruction. There was no statistically significant difference between the AI-LOAG and T-LOAG, though. This means that AI-assisted methods are just as good as teacher-led methods at helping people write better. This answers the first research question by showing that AI-assisted LOA is a good alternative to traditional writing instruction. It is more effective than traditional writing instruction and has the same benefits as teacher-guided LOA.

The qualitative results, which were obtained by analyzing interviews by theme, gave us a better understanding of what learners thought about AI in writing instruction, which answered the second research question.

Four main themes came up: how useful people thought AI feedback was, how it made them feel more independent and motivated, how much they trusted and doubted AI responses, and how worried they were about AI's ethical issues and potential for abuse. Learners liked that ChatGPT gave them quick, detailed feedback on their grammar, vocabulary, and sentence structure. One participant said, *"It corrects my mistakes and tells me why. I think I am learning faster this way"* (P2). This fits with earlier research that showed AI's ability to give timely, personalized feedback (Hwang et al., 2023; Song & Song, 2023). In addition, students said they felt more independent and motivated, as shown by comments like *"I can write and improve by myself"* (P13).

This supports the idea that AI encourages SRL (Anggraeni et al., 2025). But people are worried about AI's reliability, like when they say, *"Sometimes the sentences it gives me are too formal or not natural"* (P8), and they are also worried about how it might be misused, like by copying text that AI made (Al-Abbas et al., 2023; Farrokhnia et al., 2024).

These results are in line with what other studies have found, but they also add new information that is specific to this situation. The fact that AI and teacher-assisted LOA work equally well is similar to studies that show AI can provide the same benefits as human feedback (Biju et al., 2024; Sharifzadeh et al., 2025; Zhao, 2022). However, the fact that there is no significant difference between AI-LOAG and T-LOAG suggests that AI can be used as a scalable alternative in places with few resources, like Iran. Song and Song (2023) found that AI tools make people more interested in what they're doing, which is similar to the qualitative themes of autonomy and motivation. Uyar and Büyükahıska (2025) also talked about trust issues and ethical concerns, which are similar to debates about AI's limitations and the risks of its misuse. Some studies have shown that AI scores are different from those of human raters (Al-Abbas et al., 2023), but this study's focus on formative feedback within an LOA framework reduced those problems by putting more emphasis on the process than the final

10.57647/jee.2026.0901.05

evaluation. The effects on EFL education are very important. In practice, AI-assisted LOA is a resource-efficient way to give personalized feedback in big classes, which is important in Iran because they need writing support that can grow (Asaad & Shabdin, 2021). In theory, the study supports using AI in LOA frameworks to improve formative assessment principles like giving feedback on time and giving students more control (Carless, 2007; William & Thompson, 2007). The results also show that AI should not replace teacher guidance, but rather work with it to address concerns about reliability and encourage ethical use, making sure that students think critically instead of passively.

Nevertheless, this study has limitations. The sample was made up of only male lower-intermediate students from two schools in Tabriz, which means that the results can't be applied to other genders, skill levels, or regions. The eight-session intervention was long enough to see short-term improvements, but it can't prove that writing skills will improve over time or that they can be used in high-stakes situations. The non-random convenience sampling and single-teacher design may introduce selection biases, even though they control for differences between teachers. Finally, using only one AI platform may not show how many different tools are out there.

Future research should use randomized, mixed-gender samples in different EFL settings to confirm and build on these results. Longitudinal studies could look at how AI-LOA affects writing over a semester and whether the initial gains in autonomy last.

Comparing different AI tools and prompt architectures would help us figure out the best ways to improve feedforward quality. Also, looking into what teachers think and what the school policies declare can help us understand what makes it easier or harder for AI-LOA to be used in educational institutions.

6. Conclusion

This study demonstrates that using AI to help with LOA is a good way to enhance the academic writing skills of Iranian EFL students. It works as well as teacher-assisted methods and better than traditional methods. The current research addresses AI's strengths, such as providing immediate feedback and encouraging independence, as well as its weaknesses, like reliability and ethical use. These results show that AI could be a useful tool for teaching writing in EFL, especially in places where there aren't many resources. They also stress the importance of teacher supervision to get the most out of AI. This study improves our understanding of AI's role in language education by connecting technology and pedagogy. It also calls for careful use of AI to support learner-centered assessment around the world.

10.57647/jee.2026.0901.05

Authors Contribution
All the authors have participated sufficiently in the intellectual content, conception, and design of this work or the analysis and interpretation of the data (when applicable), as well as the writing of the manuscript.

Availability of data and materials
The data that support the findings of this study are available from the corresponding author upon reasonable. request.

Conflict of interest
The author states that there is no conflict of interest.

References

Al-Abbas, H. U. A. I., Halim, H. H., & Nurjati, N. N. (2023). Harnessing the use of artificial intelligence in language assessment: A systematic comprehensive review. *Tell-Us Journal*, *9*(3), 723–745. http://dx.doi.org/10.22202/tus.2023.v9i3.7366

Aliakbari, M., Barzan, P., & Maadikhah, M. M. (2025). Privacy issues in online learning: A review of literature and suggestions for further research. *Language, Identity and the Digital Realm*, *1*(1), 1–17. https://doi.org/10.22034/lidr.2025.722877

Anggraeni, C. W., Mujiyanto, J., Rustipa, K., & Widhiyanto. (2025). Effects of utilizing self-regulated learning-based instruction on EFL students' academic writing skills: A mixed-method investigation. *Asian-Pacific Journal of Second and Foreign Language Education*, *10*(1), 1–15. https://doi.org/10.1186/s40862-024-00317-6

Asaad, H. Q. M., & Shabdin, A. A. (2021). The predictive role of morphological awareness and productive vocabulary knowledge in L2 postgraduate students' academic writing. *Eurasian Journal of Applied Linguistics*, *7*(1), 24–44. https://doi.org/10.32601/ejal.911149

Biju, N., Abdelrasheed, N. S. G., Bakiyeva, K., Prasad, K. D. V., & Jember, B. (2024). Which one? AI-assisted language assessment or paper format: An exploration of the impacts on foreign language anxiety, learning attitudes, motivation, and writing performance. *Language Testing in Asia*, *14*(1), 1–45. https://doi.org/10.1186/s40468-024-00322-z

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education,22*(2), 151–167. https://doi.org/10.1080/713695728

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

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57–66. https://doi.org/10.1 080/14703290601081332

Chen, T. J. (2023). ChatGPT and other artificial intelligence applications speed up scientific writing. *J. Chin. Med. Assoc, 86*(1), 351–353. https://doi.org/10.1097/JCMA.0000000000000900

Creswell, J. W., & Creswell, J. D. (2005). Mixed methods research: Developments, debates, and dilemmas. *Research in Organizations: Foundations and Methods of Inquiry*, *2*(1), 315–326.

Creswell, J. W., Clark, V. L. P., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed. *Handbook of Mixed Methods in Social & Behavioral Research*, *209*(1), 209–240.

Dave, A. (2004). *Oxford placement test 2: Test pack*. Oxford University Press.

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research*. Routledge.

Er, H. K., & Farhady, H. (2023). The effects of learning-oriented assessment on academic writing. *International Journal of Assessment Tools in Education*, *10*(2), 197–217. https://doi.org/10.21449/ijate.1222410

Esfandiari, R., Meihami, H., & Jahani, F. (2022). Exploring Iranian postgraduate EFL students' academic writing experiences and expectations: A dynamic narrative approach. *Tesl-Ej*, *25*(4), 14–29.

Estaji, M., & Safari, F. (2023). Learning-oriented assessment and its effects on the perceptions and argumentative writing performance of impulsive vs. reflective learners. *Language Testing in Asia*, *13*(1), 1–31. https://doi.org/10.1186/s40468-023-00248-y

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, *61*(3), 460–474. https://doi.org/10.1080/14703297.2023.2195846

Geranpayeh, A. (2003). A quick review of the English Quick Placement Test. *Research Notes, 12*(3), 8–10.

Ghaneiarani, S., Alavi, S. M., & Kaivanpanah, S. (2024). Enhancing the writing ability of Iranian EFL learners through learning-oriented assessment: Peer and teacher feedback implementation. *Language Testing in Asia*, *14*(1), 1–39. https://doi.org/10.1186/s40468-024-00298-w

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of Qualitative Research*, *2*(105), 163–194.

Hsiao, J. C., & Chang, J. S. (2023). Enhancing EFL reading and writing through AI-powered tools: design, implementation, and evaluation of an online course. *Interact. Learn. Environ, 16*(1), 1–15. https://doi.org/10.1080/10494820.2023.2212716

Hwang, W. Y., Nurtantyana, R., Purba, S. W. D., Hariyanti, U., Indrihapsari, Y., and Surjono, H. D. (2023). AI and recognition technologies to facilitate English as a foreign language writing for supporting personalization and contextualization in authentic contexts. *J. Educ. Comput. Res., 61*(1), 1008–1035. https://doi.org/10.1177/07356331221137253

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Jacobs, H.L., Zingraf, S.A., Wormuth, D.R., Hartfiel, V.F., and Hughey, J.B. (1981). *Testing ESL composition*. Newbury House Publishers.

Jafarnia, A., Hariri, H., & Parvizi, G. R. (2023). Unlocking the potential: Exploring the multifaceted impact of artificial intelligence integration in language learning. *Language Education and Technology, 3*(2), 23–47.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*(1), 102–124. https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., & Zou, D. (2025). The role of ChatGPT in enhancing English teaching: A paradigm shift in lesson planning and instructional practices. *Educational Technology & Society*, *28*(3), 4–20. https://doi.org/10.30191/ETS.202507_28(3).SP02

Kovalenko, I., & Baranivska, N. (2024). Integrating artificial intelligence in English language teaching: Exploring the potential and challenges of AI tools in enhancing language learning outcomes and personalized education. *Європейські соціо-правові та гуманітарні студії*, *4*(1), 86–95. https://doi.org/10.61345/2734-8873.2024.1.9

Lee, I. (2019). *Classroom writing assessment and feedback in L2 school contexts*. Springer Singapore.

Liu, C., Hou, J., Tu, Y. F., Wang, Y., and Hwang, G. J. (2021). Incorporating a reflective thinking-promoting mechanism into artificial intelligence-supported English writing environments. *Interact. Learn. Environ., 31*(2), 3340–3359. https://doi.org/10.1080/10494820.2021.1927115

Loncar, M., Schams, W., and Liang, J. S. (2023). Multiple technologies, multiple sources: Trends and analyses of the literature on technology-mediated feedback for L2 English writing published from 2015-2019. *Comput. Assist. Lang. Learn., 36*(1), 722–784. https://doi.org/10.1080/09588221.2021.1943452

Long, D. R., Strauss, A., & Corbin, J. (1993). Basics of qualitative research: Grounded theory procedures and techniques. *The Modern Language Journal, 77*(2), 221–235. https://doi.org/10.2307/328955

Mallahi, O. (2024). Exploring the status of argumentative essay writing strategies and problems of Iranian EFL learners. *Asian-Pacific Journal of Second and Foreign Language Education*, *9*(1), 1–19. https://doi.org/10.1186/s40862-023-00241-1

Mudawy, A. M. A. & Mousa, A. A. E. (2017). Academic writing among Saudi university students: problems and solutions. *International Journal of Science and Research (IJSR), 6*(5), 1–4.

Pineteh, E. A. (2014). The academic writing challenges of undergraduate students: A South African case study. *International Journal of Higher Education, 3*(1), 12–22. http://dx.doi.org/10.5430/ijhe.v3n1p12 Rostampour, M., & Akbarpour, L. (2025). Investigating the impact of mobile-assisted metalinguistic corrective feedback on Iranian EFL learners' grammar proficiency: A mixed-methods approach. *Journal of Education Experiences*, *8*(2), 215-239. https://doi.org/10.71922/jee.2025.1219901

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., ... & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality & Quantity*, *52*(1), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

Sharifzadeh, E., Tayyebi, G., & Akbarpour, L. (2025). AI vs. human feedback in EFL writing: Iranian learners' perceptions. *Journal of Education Experiences*, *8*(2), 75-109. https://doi.org/10.71922/jee.2025.1209120

Singh, M. K. M. (2019). Academic reading and writing challenges among international EFL master's students in a Malaysian University: The voice of lecturers. *Journal of International Students, 9*(4), 972–992. https://doi.org/10.32674/jis.v9i3.934

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*(1), 126–143. https://doi.org/10.3389/fpsyg.2023.1260843

Stiggings, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*(4), 324–328. https://doi.org/10.1177/003172170508700414

Topping, K. J. (2009). Peer assessment. *Theory Into Practice*, *48*(1), 20–27. https://doi.org/10.1080/00405840802577569

Uyar, A. C., & Büyükahıska, D. (2025). Artificial intelligence as an automated essay scoring tool: A focus on ChatGPT. *International Journal of Assessment Tools in Education*, *12*(1), 20–32. https://doi.org/10.21449/ijate.1517994

Vygotsky, L. S. (1978*). Mind in society: The development of higher psychological processes*. Harvard University Press.

William, D., & Thompson, M. (2007). *Integrating assessment with learning: what will it take to make it work?* In: Dwyer, CA, (ed.) The Future of Assessment: Shaping Teaching and Learning. (pp. 53–82). Lawrence Erlbaum Associates.

Wang, J. (2024). An empirical study on continuation writing in senior high school under the assessment and feedback of ChatGPT. *Journal of Theory and Practice of Contemporary Education*, *4*(05), 7–22. https://doi.org/10.53469/jtpce.2024.04(05).02

Zeng, W., Huang, F., Yu, L., & Chen, S. (2018). Towards a learning-oriented assessment to improve students' learning: A critical review of literature. *Educational Assessment Evaluation and Accountability, 30*(1), 211–250. https://doi.org/10.1007/s11092-018-9281-9

Zhang, R., and Zou, D. (2023). A review of research on technology-enhanced peer feedback for second language writing based on the activity theory framework. *Educ. Inf. Technol., 28*(1), 6727–6753. https://doi.org/10.1007/s10639-022-11469-8

Zhao, X. (2022). Leveraging artificial intelligence (AI) technology for English writing: Introducing Wordtune as a digital writing assistant for EFL writers. *RELC J, 89*(1), 10–40. https://doi.org/10.1177/00336882221094089

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Practice, 41*(1), 64–70. https://doi.org/10.1207/s15430421tip4102_2