Razi University

Technology Assisted Language Education

ISSN: 2981-1465

Homepage: https://tale.razi.ac.ir

AI-Powered Writing in the EFL Classroom: The Pedagogical Role of ChatGPT

Behzad Aghajanzadeh Kiasi[1] and Abbas Pourhosein Gilakjani[2]*

1. Department of English Language and Linguistics, College of Humanities, Ra. C., Islamic Azad University, Rasht, Iran. E-mail: aghajanzadeh@iaurasht.ac.ir

2. Corresponding author, Department of English Language Translation, La. C., Islamic Azad University, Lahijan, Iran. E-mail: abbas.pourhosein@iau.ac.ir

Article Info

Article type:
Research Article

Article history:
Received February 20, 2026
Received in revised form June 21, 2026
Accepted June 30, 2026
Published online June 30, 2026

Keywords:
AI integration in EFL classrooms, ChatGPT-assisted writing pedagogy, generative AI in L2 writing, Iranian intermediate EFL learners, writing proficiency learning and retention

Abstract

Generative Artificial Intelligence (AI) systems, such as ChatGPT, hold significant potential to reshape. This study compared the impact of ChatGPT-supported writing activities with intentional instruction on the writing performance of Iranian intermediate EFL learners. Fifty-six learners from Zaban Iran English Language Institute in Rasht, Iran, were selected via the Solutions Placement Test and assigned to one experimental and one control group, each with 28 participants. After establishing homogeneity with a pretest, the experimental group engaged in writing tasks facilitated by ChatGPT, while the control group received traditional writing instruction. An immediate posttest and a delayed posttest two weeks later measured both immediate performance and retention. Analyses using independent and paired-samples t-tests revealed that the experimental group significantly outperformed the control group in writing proficiency and retention, indicating that AI-supported instruction can enhance L2 writing outcomes. The results highlight the pedagogical value of integrating ChatGPT into writing pedagogy and suggest that conventional methods can be strengthened through AI-assisted approaches. Nonetheless, effective implementation requires careful planning, ongoing teacher training, and robust professional support to address challenges such as tool reliability, alignment with learning objectives, and data privacy considerations.

Cite this article: Pourhosein Gilakjani, A., & Aghajanzadeh Kiasi, G. (2026). AI-Powered Writing in the EFL Classroom: The Pedagogical Role of ChatGPT. *Technology Assisted Language Education*, (), 101-122. doi: 10.22126/tale.2026.13655.11

Publisher: Razi University.

DOI: http//doi.org/10.22126/tale.2026.13655.1175

Volume 4. Issue 2. 2026. Pages 101 to 122.

Introduction

Artificial intelligence (AI) has transformed multiple domains, from healthcare and finance to education and entertainment. One of the most notable advances in this domain is ChatGPT, a conversational AI model created by OpenAI (Sallam, 2023). With the growing emphasis on technology-enhanced language education, AI-based tools such as ChatGPT are increasingly being explored for their potential to facilitate learners' skill development. The integration of ChatGPT into educational settings has generated both excitement and concern, as the tool demonstrates the ability to produce academic essays, narratives, and other genres of text (Dergaa et al., 2023). Over the past two decades, AI in education (AIED) research has focused on promoting learner autonomy and partially automating teaching functions (Afzaal et al., 2022). Proponents argue that such tools can help address educational challenges like teacher shortages, increased demand for virtual support, and disparities in learner achievement linked to socioeconomic background (Imran & Almusharraf, 2023).

Chatbots and other AI applications are now recognized as valuable resources for improving efficiency, accessibility, and personalization (Cotton et al., 2023). In particular, ChatGPT offers opportunities for writing practice, interactive communication, instant feedback, and self-directed learning. Firat (2023) points out that AI systems like ChatGPT can foster improved academic performance, the development of soft skills, and individualized learning experiences. Similarly, Brynjolfsson and McAfee (2017) highlight AI's potential to personalize instruction, reduce administrative workload, and provide insights through data-driven analysis.

Among the four core language skills, writing remains the most demanding for learners (Khansir & Gholami Dashti, 2013). Teaching writing is often considered labor-intensive, requiring attention to content, organization, coherence, and rhetorical aspects. AI tools like ChatGPT may ease these difficulties by supporting learners through drafting, revising, and polishing written work. They also help students and researchers produce more structured, fluent, and concise writing, particularly for those who are still developing language proficiency or working under time constraints (Nazari et al., 2021). Additionally, ChatGPT's adaptability allows learners to refine their texts repeatedly, explore stylistic variations, and interact with the system in ways that foster more natural, communicative learning experiences (Kim et al., 2023).

Iranian EFL teachers, much like their peers worldwide, encounter persistent obstacles in teaching writing effectively. Moses and Mohamad (2019) emphasize that instructors often struggle to meet the needs of learners with diverse learning styles, provide individualized feedback in large classes, and motivate students who frequently view writing as intimidating. Overcoming these barriers is vital to supporting writing development among intermediate-level learners in Iran. Traditional teaching practices still dominate many Iranian classrooms. Yet, as Aldabbus et al. (2022) note, these approaches often prioritize rote memorization

rather than fostering essential skills such as paraphrasing, summarizing, and developing ideas. This heavy reliance on conventional methods restricts learners' ability to produce flexible and meaningful writing, leaving them ill-prepared for academic and professional contexts.

In addition to difficulties in acquiring writing skills, Iranian intermediate learners also face challenges in retaining them over time. The lack of technology-enhanced tools in traditional classrooms makes this problem more pronounced. Unlike conventional instruction, ChatGPT provides instant personalized feedback, encourages experimentation, and offers diverse practice opportunities. Without such support, learners may miss out on timely guidance and struggle to refine their writing, which can undermine motivation and slow progress. Despite these affordances, little empirical work has examined ChatGPT's actual impact on L2 writing outcomes, especially in Iran. Most studies so far have focused on the theoretical potential of AI tools rather than on their measurable effects on writing achievement and retention. Febriani (2022) has highlighted that EFL learners often struggle to form logical connections in writing. Similar challenges have also been reported in the Iranian EFL context, where learners experience difficulties in the effective use of cohesive devices, conjunctions, transition markers, and coherence strategies in writing tasks (Samadian & Mohseny, 2019; Shahhoseini et al., 2024). These gaps highlight the need for systematic research to test whether ChatGPT can provide more effective and sustainable writing instruction than traditional approaches.

This study aimed to examine whether ChatGPT–mediated instruction yields greater improvements in writing proficiency and retention among Iranian intermediate EFL learners than traditional teaching approaches. Specifically, it evaluated ChatGPT's role as a pedagogical aid in both short-term learning gains and long-term writing retention. Additionally, the study sought to extend research on technology-enhanced language learning by examining the practical integration of advanced AI tools into classroom settings. Findings from this research are expected to guide educators and policymakers on the potential of ChatGPT to complement existing practices, close instructional gaps, and support data-driven decision-making in adopting innovative technologies. Accordingly, the study was directed by the questions proposed below: In this study, the following research questions are addressed:

- Does ChatGPT-assisted instruction lead to greater gains in Iranian intermediate EFL learners' writing skills than the conventional method?
- Does ChatGPT-assisted instruction lead to greater retention of writing skills among Iranian intermediate EFL learners than the conventional method?

Volume 4. Issue 2. 2026. Pages 101 to 122.

Literature Review

Artificial intelligence has its roots in the 1950s, culminating in Alan Turing's proposal of the Turing Test as a method to assess a machine's capacity to replicate human cognitive processes. Early debates distinguished between strong AI, which emphasizes human-like reasoning, and weak AI, which is designed to handle specific, task-oriented processes (Kannan & Munday, 2018). AI has strong connections to computational psychology, as the "mind as computer" perspective conceptualizes cognition as the algorithmic processing of symbols, similar to the workings of a digital computer (Zawacki-Richter et al., 2019).

This study leverages Davis's (1989) Technology Acceptance Model (TAM) to examine learners' acceptance of AI tools. Over the past three decades, across fields such as educational technology and computer-assisted language learning, TAM has been leveraged to explain the drivers of acceptance and resistance to new technologies (Al-Emran et al., 2018; Martín-García et al., 2022; Sun & Mei, 2022). In addition, two theoretical perspectives help explain ChatGPT's role in feedback. Winstone and Carless (2020) view feedback as dialogic, emphasizing interaction and learner engagement. In line with this view, ChatGPT facilitates ongoing dialogue by offering instant guidance. Barrot (2023) further argues that ChatGPT serves as a formative feedback tool, providing personalized, need-based responses throughout the writing process. The current study draws on Barrot's framework to examine ChatGPT's impact on Iranian EFL learners.

Technically, ChatGPT is a large-scale generative model built on transformer architecture. It requires sequential patterns in language and generates fluent, human-like responses (Lund & Wang, 2023; Pavlik, 2023). With its capacity for fine-tuning, it can be adapted for instructional purposes in language classrooms (Cotton et al., 2023; Shen et al., 2023). A major advantage of ChatGPT is its ability to personalize learning by tailoring tasks to learners' needs and proficiency levels. This personalization supports autonomy, encourages reflection, and promotes ownership of learning. For example, Agustini (2023) observed that ChatGPT encouraged self-directed learning and strengthened learners' sense of responsibility.

Recent studies in Iranian and ESP/EFL contexts further support the pedagogical value of technology-mediated instruction and feedback. For example, Zohrabi and Khalili (2024), in a cross-cultural study on medical students' ESP writing courses, found that diverse written corrective feedback strategies significantly improved learners' writing performance and engagement. Their findings highlighted that learners benefited more from feedback approaches that combined explicit correction with opportunities for reflection and revision. The study also emphasized that technology-assisted feedback environments can enhance learners' motivation and awareness of linguistic accuracy. Similarly, Dobakhti et al. (2023) investigated the effects of flipped and online instruction on EFL learners' writing ability and reported significant improvements in learners' writing performance. Their results

demonstrated that technology-enhanced learning environments increased learner participation, promoted autonomous learning, and created more interactive opportunities for writing practice and feedback. These findings are particularly relevant to the present study because ChatGPT similarly provides immediate, individualized, and interactive support during the writing process.

Since its launch in late 2022, ChatGPT has attracted strong interest from educators, and research into ChatGPT's educational applications continues to grow. Studies report that it can generate coherent academic texts, provide feedback on organization and language, and suggest revisions that strengthen arguments and conclusions (Fitria, 2023; Gao et al., 2023). Yan (2023) found that while ChatGPT supported improvements in learners' writing, its use also raised concerns regarding academic honesty and plagiarism.

Kim (2023) emphasized the usefulness of ChatGPT in generating instructional content within the framework of Task-Based Language Teaching (TBLT), and called on educators to adjust both teaching and assessment practices in line with technological advancements. Likewise, Yan (2023) analyzed ChatGPT's use in an L2 writing practicum, concluding that it can enhance writing efficiency and meaningfully contribute to writing pedagogy, while also raising concerns about academic integrity and fairness in education.

Ali et al. (2023) also examined how ChatGPT influenced both learners' and teachers' motivation in English language classrooms. Their findings showed that while the tool significantly supported the development of learners' reading and writing skills, it had only limited effects on speaking and listening. In a related study, Schmidt-Fajlik (2023) compared ChatGPT with Grammarly and ProWritingAid in terms of grammar assistance. The results demonstrated that ChatGPT was more effective than the other tools in identifying and improving grammatical issues among EFL learners.

Extending this line of research, Rudolph et al. (2023) reported that ChatGPT positively shaped teaching and learning practices in higher education. Although they noted some drawbacks, the technology was seen as providing new opportunities for instructors and administrators to enhance teaching, assessment, and student engagement. Similarly, Kovačević (2023) found that ChatGPT could individualize the learning process for students in English for Specific Purposes (ESP), thereby supporting more effective language acquisition and better learning outcomes.

In addition to these findings, Koraishi (2023) highlighted ChatGPT's usefulness for material design and assessment in EFL contexts. The study indicated that the tool can simplify the creation of engaging, context-sensitive resources aligned with learner needs. Moreover, it can deliver immediate, personalized feedback that helps learners monitor and improve their performance, enriching the overall learning process. On the other hand, despite these advantages, several risks remain associated with ChatGPT's use. Kohnke et al. (2023) cautioned that the tool's paraphrasing ability can lead to plagiarism if outputs are used

without proper attribution. They also observed that identical prompts may yield inconsistent responses, raising questions about reliability and precision. These challenges suggest the importance of responsible implementation and critical awareness when incorporating ChatGPT into language education.

Method

Design

Given the study's quantitative nature and focus on addressing the research questions, a quasi-experimental design was employed. This design was characterized by the non-random selection of participants from intact classes, random assignment of learners into study groups, pretest administration, implementation of the intervention program, and posttest administration. The purpose was to examine whether providing writing instruction through ChatGPT, compared with the conventional method, significantly affects Iranian EFL learners' writing performance and retention. The dependent variables were learners' writing performance and retention, while the independent variables were the use of ChatGPT as an instructional intervention and conventional methods of writing instruction.

Participants

The sample comprised 56 intermediate-level EFL learners, drawn from a population of 74 students, after homogenization via the Solutions Placement Test (SPT). The participants, aged 17 to 26, included both male and female learners, all of whom were native Persian speakers. They were taking an English course at Zaban Iran English Language Institute in Rasht, Iran (see Table 1). The learners were assigned to one experimental and one control group, each with 28 learners. It is worth mentioning that participation was voluntary, and students were assured that withdrawal would have no academic consequences. Confidentiality was maintained, and all learners provided informed consent following a detailed explanation of the study (including its purpose, procedures, potential risks, and benefits).

Table 1

Demographic Characteristics of the Participants

Variable	Details
Initial Participants	74 intermediate EFL learners
Sample size	56 learners
Age Range	17–26 years
Gender	Male and female learners
Native Language (L1)	Persian

English Proficiency Level	Intermediate
Educational Context	English language learners at Zaban Iran English Language Institute
Location	Rasht, Iran

Instruments

To ensure more comprehensive findings, several instruments were employed: the Oxford Solutions Placement Test (SPT), a pretest, ChatGPT, an immediate posttest, a delayed posttest of L2 writing, and an analytic rubric for scoring students' writing papers. The writing tests were selected from TOEFL Writing Questions (2022, 2nd ed.) and the sample essays and prompts available at TOEFL Resources (https://www.toeflresources.com/sample-toefl-essays/). To ensure comparability, the writing topics were chosen at the same proficiency level for the groups.

ChatGPT (Chat Generative Pre-Trained Transformer) was used in the experimental group. It is an artificial intelligence language model developed by OpenAI that utilizes deep learning techniques to generate human-like text responses. It was used because it is designed to understand and generate text based on its input (Brown et al., 2020). Moreover, to enhance scoring reliability, learners' writing papers were scored independently by two qualified raters, and rater consistency was established to minimize subjectivity in the evaluation process.

Procedure

A total of 56 EFL learners were selected from a pool of 74 intermediate-level students based on the Solutions Placement Test (SPT). Only those who scored between 30 and 40 were included, ensuring homogeneity at the intermediate level. Participants were then divided into an experimental group (n = 28) and a control group (n = 28) by non-random assignment. To establish baseline performance, all participants took a pretest writing task entitled "Movies and television strongly influence the way people behave. Do you agree or disagree? Use reasons and specific examples to support your answer."

The experimental group received instruction through ChatGPT across 16 sessions (two sessions per topic, eight topics total). At the beginning of the treatment, learners were introduced to the objectives of the study and received training on how to interact with ChatGPT and interpret its feedback. They were also informed about ethical use, including the importance of using AI-generated feedback for learning and revision rather than copying complete responses. Instruction followed a three-phase cycle:

1. Learners produced a short email (e.g., describing their favorite sport).

2. ChatGPT corrected grammar, vocabulary, and organization, providing an improved version.

3. Learners studied revisions, discussed suggested changes, and produced a new piece on a related topic (e.g., writing about sports and healthy living).

The teacher guided learners in analyzing ChatGPT's feedback by discussing common errors, suggested revisions, and alternative expressions in class. Learners compared their original drafts with the revised versions and identified areas requiring improvement. This iterative process encouraged learners to internalize ChatGPT's feedback and apply it to new contexts. The control group, however, followed a conventional, product-oriented approach that emphasized writing, teacher-led correction, revision, and rewriting. Both groups completed eight topics over the 8-week treatment. To measure progress, an immediate posttest and a delayed posttest (two weeks later) were administered. The posttest prompt was "Overall, the widespread use of the internet has a mostly positive effect on life in today's world." The delayed posttest used the prompt "Are you a pessimist or an optimist? Elaborate." These alternative topics controlled for testing effects while maintaining equivalent difficulty.

For scoring, Brown's (2007) analytic rubric was used to evaluate content, organization, vocabulary, grammar, and mechanics. Two independent raters scored the writings, verifying inter-rater reliability with Pearson's r. Finally, the research questions were addressed by calculating descriptive statistics (mean scores) and employing inferential statistics, specifically paired-samples and independent-samples t-tests.

Results

To assess scoring consistency, inter-rater reliability (IRR) was calculated for the three writing tasks, and to further establish reliability, "Single Measures" for Intraclass Correlation Coefficients (ICCs) were calculated separately for the writing tests. Table 2 displays the ICC results with 95% confidence intervals.

Table 2

Intraclass Correlation Coefficients for Pretest, Posttest, and Delayed Posttest

Test	ICC (Average Measures)	95% CI (Lower–Upper)	F	df1	df2	Sig.
Pretest	.889	.790 – .941	9.018	54	54	.000
Posttest	.825	.670 – .908	5.726	54	54	.000
Delayed Posttest	.832	.749 – .921	6.987	54	54	.000

The inter-rater reliability (r) for the pretest scores was estimated at 0.889, with a relatively wide 95% confidence interval (0.790,0.941). Reliability for the posttest scores was slightly lower at

0.825, with a 95% CI of 0.670 to 0.908. The estimated reliability between the two raters for the delayed posttest (r2= .832) was also acceptable. These findings confirm that the scoring procedure was consistent and dependable across raters and testing occasions.

Table 3 presents a descriptive analysis for the groups' writing pretest. The purpose of this test was to determine any initial differences in writing ability prior to the intervention.

Table 3

The Results of The Descriptive Statistics Reported for the Writing Pretest

Test	N	Range	Minimum	Maximum	Sum	Mean	Std. Deviation	Variance
Pre. Cont.	28	4.5	12.75	16.25	217.5	14.6	1.339	1.795
Pre. Exp.	28	5.5	12	16.5	217.75	14.32	1.678	2.817

As shown, the experimental group achieved a mean score of 14.32, compared with 14.60 for the control group. The scores across both groups were relatively close, suggesting that the groups were comparable in their initial writing ability. Figure 1 illustrates the distribution of pretest scores across groups, further highlighting the comparability of their performance before the treatment. Overall, the descriptive analysis indicates that participants demonstrated moderate, rather than advanced, writing proficiency at baseline.

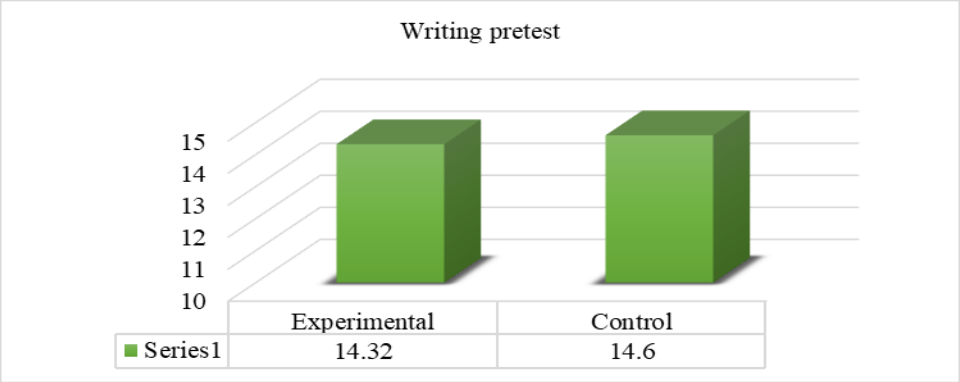

Figure 1 *The Means Plot for the Pretest of The Study Groups*

Upon finalizing the treatment, an immediate writing posttest was administered to determine the groups' performance. The descriptive statistics for the participants' posttest scores are presented in Table 4.

Table 4

The Results of The Descriptive Statistics for The Writing Posttest

Volume 4. Issue 2. 2026. Pages 101 to 122.

Test	N	Range	Minimum	Maximum	Sum	Mean	Std. Deviation	Variance
Post. Cont.	28	4.5	13	17.5	228.5	15.23	1.174	1.379
Post. Exp.	28	3.75	14.5	19.25	252.5.	17.70	1.214	1.475

As evidenced in Table 4, the experimental group's mean score (M = 17.70) was 2.47 points higher than the control group's mean (M = 15.23). This substantial difference suggests that learners who received ChatGPT-assisted instruction showed greater improvement in writing proficiency than those taught through conventional methods. Figure 2 illustrates the progression of groups across the tests. The upward trend confirms that both groups improved, but the control group was clearly outperformed by the experimental group, reflecting the significant impact of ChatGPT on L2 writing development.

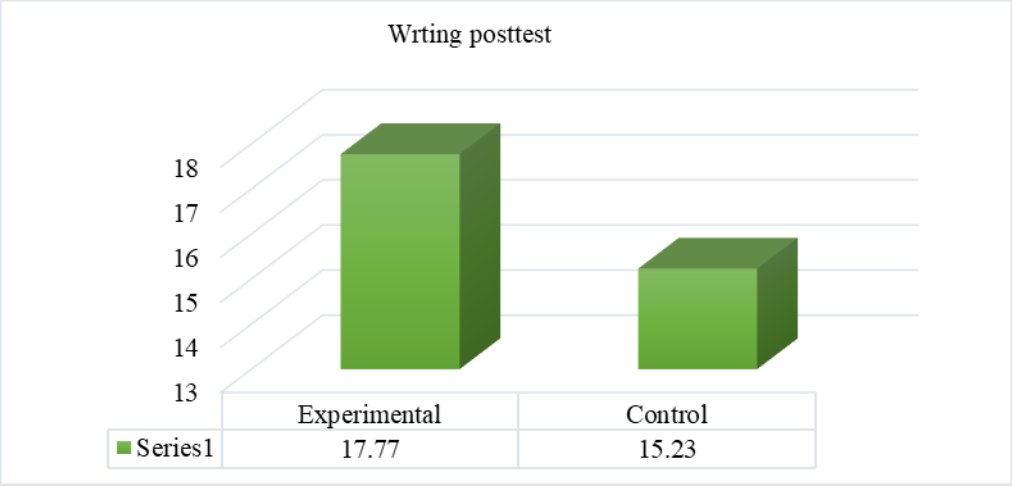

Figure 2 *The Means Plot for the Posttest of the Study Groups*

To evaluate the retention of L2 writing skills following the intervention, the researcher administered a delayed posttest to both groups. The reports for the posttest and delayed posttest scores are presented in Table 5.

Table 5

The Results of Descriptive Statistics for the Writing Posttest and Delayed Posttest

	Groups	N	Mean	Std. Deviation	Std. Error Mean

Volume 4. Issue 2. 2026. Pages 101 to 122.

Posttest scores	Cont.	28	15.23	1.803	.403
	Exp.	28	17.70	1.627	.363
Delayed posttest Scores	Cont.	28	14.12	1.080	.241
	Exp.	28	17.20	1.141	.255

As shown in Table 5, the experimental group consistently outperformed the control group across both assessments. In the immediate posttest, the experimental group recorded a mean score of 17.70, exceeding the control group's mean of 15.23 by 2.47 points. Notably, with a lower standard deviation (SD=1.62) than the control group (SD=1.80), the experimental group demonstrated greater score consistency. The experimental group sustained its high performance with a mean score of 17.20 on the delayed posttest. In contrast, the control group′s mean score declined to 14.12. Although the standard deviations were nearly identical, the experimental group showed remarkable retention of writing performance in comparison to the decline observed in the control group.

Figure 3 illustrates these findings, highlighting the sustained advantage of ChatGPT-assisted instruction in promoting not only immediate writing improvement but also long-term retention.

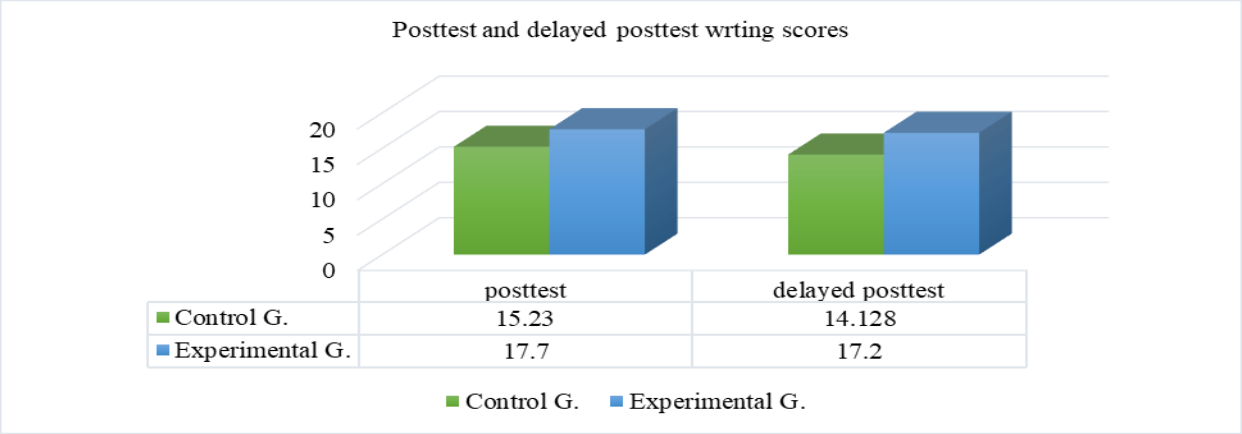

Figure 3 *The Comparison of the Means for the Groups on the Posttest and Delayed Posttest of Writing*

To verify that the data were suitable for parametric statistical procedures, the normality of the writing scores was checked for both the control and experimental groups across the pretest, posttest, and delayed posttest. This assumption was evaluated using the Shapiro–Wilk test, along with examination of skewness and kurtosis indices. As summarized in Table 6, all

distributions satisfied the normality requirement, indicating that the data met the necessary conditions for subsequent parametric analyses.

Table 6

Tests of Normality for Writing Scores (Pretest, Posttest, and Delayed Posttest)

Test	Groups	N	Skewness	Kurtosis	Shapiro–Wilk Statistic	Sig. (p)	Normality Decision
Pretest	Control	28	0.12	-0.44	0.967	0.431	Normal
Pretest	Experimental	28	-0.18	-0.57	0.961	0.312	Normal
Posttest	Control	28	-0.09	-0.62	0.958	0.281	Normal
Posttest	Experimental	28	0.21	-0.49	0.973	0.522	Normal
Delayed Posttest	Control	28	0.15	-0.41	0.969	0.398	Normal
Delayed Posttest	Experimental	28	-0.24	-0.36	0.975	0.584	Normal

All calculated skewness and kurtosis values fell within the acceptable threshold of ±1, and the Shapiro-Wilk results ($p > .05$) showed no significant deviation from normality in either group across the three testing phases. These outcomes confirmed that the data were approximately normally distributed, validating the use of parametric techniques in the following analyses.

Since the normality assumption was confirmed across all sets of writing scores, parametric statistical tests were used to examine potential differences between the control and experimental groups. Consequently, independent-samples t-tests were conducted to compare the mean scores on the pretest, posttest, and delayed posttest, aiming to examine the influence of ChatGPT-supported instruction on the development and retention of learners' writing proficiency.

To assess whether the groups differed in performance on the pretest and posttests in a statistically meaningful way, inferential statistics were employed to analyze the participants' writing test scores. Table 7 presents the outcomes of the pretest independent-samples t-test. Levene's test was used to evaluate whether the error variances are equal across groups (homogeneity of variance).

Table 7

Results of Independent-Samples T-Test on the Pretest Scores of Writing for the Groups

	Levene's Test	t-test for Equality of Means

Volume 4. Issue 2. 2026. Pages 101 to 122.

		F	Sig.	t	df	Sig. (2-tailed)	MD	SED	95% Confidence Interval of the Difference	
									Lower	Upper
Writing pretest	Equal variances assumed	.389	.538	-1.594	54	.122	-.733	.460	-1.676	.209
	Equal variances not assumed			-1.594	53.52	.122	-.733	.460	-1.676	.210

According to Table 7, Levene's test produced a p-value of 0.538, indicating homogeneity of variances ($p > 0.05$). The pretest t-test yielded a p-value of 0.122, which is not statistically significant at the 0.05 level, suggesting no difference between groups. The pre-intervention writing scores showed no significant difference between groups: control M = 14.60 (SD = 1.33) vs. experimental M = 14.32 (SD = 1.67); t (54) = -1.59, p = .122 (two-tailed). This implies comparable starting proficiency. To determine whether the treatment influenced writing ability, an additional independent-samples t-test was performed on the posttest. See Table 8 for the posttest results.

Table 8

Results of Independent-Samples T-Test on The Posttest Scores of Writing for the Groups

		Levene's Test		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	MD	SED	95% Confidence Interval of the Difference	
									Lower	Upper
Writing posttest	Equal variances assumed	.434	.515	-3.36	54	.002	-1.466	.4362	-2.360	-.573
	Equal variances not assumed			-3.36	53.96	.002	-1.466	.4362	-2.360	-.573

Levene's test for equality of variances indicated $p > 0.05$ ($p = 0.515$), supporting equal variances. The posttest comparison produced a significant result ($p = .002$), with the control group at M = 15.23 (SD = 1.17) and the experimental group at M = 17.70 (SD = 1.21); t(54) =

-3.36, p = .002 (two-tailed). The data imply the experimental group improved more than the control group on the posttest.

A paired-samples comparison was used to quantify within-group improvements from pretest to posttest, as detailed in Table 9.

Table 9

Results of Paired-Samples Statistics for the Writing Mean Scores

Variable	Groups			Mean	N	Std. Deviation	Std. Error Mean
Writing	Cont.	Pair 1	Pretest	14.6	28	1.33	1.79
			Posttest	15.23	28	1.17	1.37
	Exp.	Pair 1	Pretest	14.32	28	2.58	.714
			Posttest	17.70	28	1.67	2.81

The mean writing test score increased from pretest to posttest for both groups, with the experimental group moving from M = 14.60 to M = 15.23 and the control group from M = 14.32 to M = 17.70, indicating greater improvement for the experimental group on the posttest. To confirm within-group changes, paired-samples t-tests were run comparing pretest and posttest scores within each group; results are summarized in Table 10.

Table 10

Results of Paired-Samples T-Test For Groups' Pretest and Posttest Of Writing

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	SD	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pretest (exp.) - Posttest (exp.)	-.338	.9376	.242	-1.253	-.214	-3.029	54	.001
Pair 2	Pretest (Cont.) -Posttest (cont.)	-0.63	1.676	.433	-2.395	-.538	-3.388	54	.004

*. The mean difference is significant at the.05 level.

Table 10 indicates a significant pretest–posttest difference for the control group at α = 0.05, but the practical magnitude of improvement was limited. The experimental group demonstrated a significant difference between pretest and posttest scores, t(54) = -3.03, p = .009 (two-tailed).

The mean difference was -0.338, and the 95% confidence interval ranged from 1.25 to 2.14. Although both groups showed progress, the experimental group's gain was greater in magnitude, indicating that the treatment improved the participants' writing. To further examine the durability of the treatment effect, an additional independent-samples t-test was conducted (see Table 11) to compare the posttest and delayed posttest scores between the two groups.

Table 11

Independent Samples T-Test for the Delayed Posttest of Writing Scores

		Levene's Test for Equality of Variances		t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Dif.	Std. Error Dif.	95% Confidence Interval of the Diff.
									Lower / Upper
posttest scores	Equal variances assumed	.25	.61	-2.70	54	.000	.95	.35	-1.66 / -.23
	Equal Variances not assumed		.36	-2.70	53.88	.010	.95	.35	-1.66 / -.23

The assumption of homogeneity of variances was met (Levene's test p = .61). Table 11 reports a significant group difference in the post-test scores: control M = 14.12 (SD = 1.08) vs. experimental M = 17.15 (SD = 1.14), t(54) = -2.70, p = .010 (two-tailed). At the delayed posttest, the between-groups difference was -0.95 (95% CI: -1.66 to -0.23). The corresponding eta-squared was $\eta^2 = .161$, reflecting a moderate effect size (Cohen, 1988). These results imply a meaningful and enduring impact of the intervention on writing performance. A paired-samples t-test was used to evaluate posttest versus delayed posttest changes in writing tests within each group, with the results reported in Table 12.

Table 12

Paired Samples T-Test for the Posttest and Delayed Posttest of Writing for the Two Groups

Groups						t	df	Sig. (2-tailed)
	Mean	SD	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			

Con.	posttest scores – delayed posttest scores	–	-1.11	13.7	.306	-1.84	-.55	3.91	54	.001
Exp.	posttest scores – delayed posttest scores	–	-0.5	17.7	.395	-3.17	-1.52	5.27	54	.06

Table 12 indicates that both groups showed declines from posttest to delayed posttest. The control group experienced a significant drop ($t(54) = -3.91$, $p = .001$), with mean difference -1.11 (95% CI: -1.84 to -0.55). The experimental group showed a non-significant decline ($t(54) = -1.27$, $p = .216$), with a mean difference of -0.50 (95% CI: [-1.52, 0.52]). These findings suggest that, by the delayed posttest, both groups experienced some decline in performance, but the decrease was significant only in the control group. In contrast, the experimental group largely maintained the gains they had achieved in the posttest, pointing to a more enduring effect of the treatment on writing ability.

Discussion

This study aimed to examine the integration of technology in EFL classrooms and to explore effective strategies for improving Iranian learners' writing proficiency. More specifically, it investigated the role of ChatGPT in promoting both immediate writing development and longer-term retention. The statistical analyses provided strong support for the effectiveness of ChatGPT-assisted writing instruction. The pretest results confirmed that the two groups were comparable in terms of their initial writing ability. However, the post-intervention test showed a statistically significant difference in performance between the experimental and control groups, with a mean posttest score of 17.70 in the experimental group compared to 15.23 in the control group. This improvement highlights ChatGPT's added value in facilitating learners' writing progress beyond what traditional instruction can achieve.

The delayed posttest further demonstrated the retention benefits of ChatGPT-supported learning. Although both groups showed some decline over time, the reduction in the control group's scores ($M = -1.11$, $p = .001$) was statistically significant, while the decrease in the experimental group ($M = -0.50$, $p = .06$) was not. This indicates that learners who practiced with ChatGPT retained their writing skills more effectively. Together, these findings suggest that ChatGPT not only supports short-term gains but also contributes to more enduring learning outcomes.

These results are consistent with earlier studies on AI in education. Barrot (2023), for example, reported greater improvements in writing among learners using ChatGPT compared to those receiving conventional instruction. Likewise, Agustini (2023) emphasized the system's ability to deliver individualized feedback, which enhances self-reflection, autonomy, and learner responsibility. The adaptive and personalized nature of ChatGPT's feedback may account for the effectiveness observed in the present study.

Other scholars have also highlighted ChatGPT's role in providing immediate and targeted feedback, which reduces delays in the learning process and supports learners in organizing their texts more effectively (Gao et al., 2023). Similarly, Fitria (2023) observed that students perceive ChatGPT's corrective feedback as reliable and supportive, leading to improved writing efficiency. Yan (2023) likewise described ChatGPT as a flexible tool capable of creating interactive and adaptive learning environments.

Barrot (2023) further argued that ChatGPT often produces more detailed and balanced feedback than teachers, especially in grammar and sentence-level issues. This resonates with the present study, in which learners in the experimental group not only performed better but also showed more consistent results, as reflected in lower score variability.

Beyond writing, ChatGPT has potential applications in vocabulary development, academic writing, and problem-solving across disciplines (Firat, 2023). Its versatility underscores its value as a pedagogical tool, particularly in contexts where access to individual feedback is limited. Nonetheless, the results must be interpreted alongside possible drawbacks. Kohnke et al. (2023) warned that ChatGPT-generated texts may resemble existing materials too closely, raising plagiarism concerns. In addition, output consistency remains an issue, as responses can vary significantly depending on prompt design. These limitations highlight the importance of carefully, ethically integrating ChatGPT into EFL teaching, with ongoing teacher involvement to monitor and supplement its use.

Overall, the findings provide strong evidence that ChatGPT-supported instruction enhances learners' writing proficiency and retention more effectively than traditional methods. By delivering personalized, timely, and consistent feedback, it serves as a valuable complement to conventional pedagogy. At the same time, responsible use requires careful attention to academic integrity and the system's limitations.

contexts.

Conclusion

The present study investigated the impact of ChatGPT-assisted instruction on the writing ability of Iranian intermediate EFL learners. The findings demonstrated that learners who received ChatGPT-supported instruction significantly outperformed those in the conventional instruction group in terms of writing development and retention. These

findings suggest that ChatGPT can function as an effective supplementary tool for enhancing writing instruction in EFL contexts.

The results also indicate that AI-assisted feedback may promote learner engagement, autonomy, and motivation by providing immediate, personalized, and interactive support during the writing process. Nevertheless, the study emphasizes that ChatGPT should not replace teacher feedback entirely. Human instruction remains essential for addressing higher-order dimensions of writing, such as argumentation, coherence, audience awareness, and critical thinking. Therefore, the most effective instructional approach may involve combining AI-supported feedback with teacher-guided instruction. Overall, the findings contribute to the growing body of research on AI-assisted language learning and provide empirical evidence supporting the pedagogical potential of ChatGPT in the Iranian EFL context.

The findings of this study offer several implications for EFL teachers, learners, teacher trainers, and educational policymakers. First, integrating ChatGPT into writing instruction may provide learners with greater opportunities for autonomous practice, individualized feedback, and continuous revision. Such features can make writing instruction more interactive and learner-centered. Second, teachers should receive appropriate training on the pedagogical and ethical use of AI tools in language classrooms. Learners also need guidance in using ChatGPT responsibly to avoid over-reliance, plagiarism, and uncritical acceptance of AI-generated responses. Developing digital literacy skills is therefore essential for the effective integration of AI technologies into education. Third, institutions and policymakers may need to reconsider assessment practices and instructional policies in response to the increasing use of generative AI in education. Establishing clear ethical guidelines and practical frameworks for AI-assisted learning can help maximize educational benefits while minimizing potential risks.

Despite its contributions, the present study has several limitations. First, the study was conducted at a single language institute in Rasht with a relatively small sample of intermediate EFL learners, which may limit the generalizability of the findings to other educational settings and proficiency levels. Second, the duration of the treatment was limited to 16 instructional sessions over approximately eight weeks. Although the intervention was sufficient to identify short-term improvements in writing performance, longer studies may provide clearer evidence regarding the sustained effects of ChatGPT-assisted instruction on long-term writing development and retention. Third, the study focused exclusively on writing skill and did not investigate the effects of ChatGPT on other language skills such as speaking, listening, or reading. In addition, only intermediate-level learners participated in the study, excluding beginner and advanced learners from the investigation.

Finally, the study examined only the use of ChatGPT in a formal classroom setting and did not explore informal, self-directed, or blended learning environments. Furthermore, issues such as digital literacy, plagiarism, ethical concerns, and the reliability of AI-

generated responses were not examined in depth and therefore remain important areas for further investigation.

Future studies should investigate the effectiveness of ChatGPT-assisted instruction across different proficiency levels, age groups, and educational contexts. Researchers may also examine the role of ChatGPT in developing other language skills, including speaking, reading, and listening. In addition, comparative studies investigating ChatGPT alongside other AI-based educational tools or instructional approaches may provide deeper insights into the relative effectiveness of AI-assisted language learning. Further research should also explore ethical concerns related to plagiarism, academic integrity, learner dependency, and the reliability of AI-generated feedback. Longitudinal studies with larger samples and more diverse contexts are recommended to better understand the long-term pedagogical impact of generative AI tools in EFL education.

Bio-data

First Author: Collected the data, wrote the first draft, designed, and conducted the procedures.

Second Author: Reviewed the final draft and made necessary revisions.

Funding: This research received/ did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest: The authors declare that they have no competing interests.

Reference

Afzaal, M., Imran, M., Du, X., & Almusharraf, N. (2022). Automated and human interaction in written discourse: A contrastive parallel corpus-based investigation of meta-discourse features in machine-human translations. *SAGE Open, 12*(4), 1-18. https://doi.org/21582440221142210

Agustini, N. P. O. (2023). Examining the role of ChatGPT as a learning tool in promoting students' English language learning autonomy relevant to Kurikulum Merdeka Belajar. *Edukasia: Jurnal Pendidikan Dan Pembelajaran, 4*(2), 921-934. https://doi.org/10.62775/edukasia.v4i2.373

Aldabbus, S., & Almansouri, E. (2022). Academic writing difficulties encountered by University EFL learners. *British Journal of English Language Linguistics, 10*(3), 1-11. https://doi.org/10.37745/bjel.2013/vol10n3111

Al-Emran, M., Mezhuyev, V., & Kamaludin, A. (2018). Technology acceptance model in M-learning context: A systematic review. *Computers & Education, 125*, 389-412. https://doi.org/10.1016/j.compedu.2018.06.008

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on learning motivation: Teachers and students' voices. *Journal of English Studies in Arabia Felix, 2*(1), 41-49. https://doi.org/10.56540/jesaf.v2i1.51

Barrot, J. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57,* 10-21. https://doi.org/10.1016/j.asw.2023.100745

Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). Pearson Education.

Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence. *Harvard Business Review, 95*(4), 3-11. Retrieved from https://hbr.org/2017/07/the-business-of-artificial-intelligence

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating. Ensuring academic integrity in the era of ChatGPT. Retrieved from https://doi.org/10.35542/osf.io/mrz8h

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information System Quarterly, 13*(3), 983-1003. https://doi.org/10.2307/249008

Dergaa, I., Chamari, K., Zmijewski, P., & Saad, H. B. (2023). From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport, 40*(2), 615-622. https://doi.org/10.5114/biolsport.2023.125623

Dobakhti, L., Zohrabi, M. & Masoudi, S. (2023). Scrutinizing the utility of flipped and online instructions for ameliorating EFL learners' writing ability. *Journal of Modern Research in English Language Studies, 10*(3), 71-94. https://doi.org/10.30479/jmrels.2023.18587.2192

Febriani, T. N. (2022). Writing is challenging: Factors contributing to undergraduate students' difficulties in writing English essays. *Erudita: Journal of English Language Teaching, 2*(1), 83-93. https://doi.org/10.28918/erudita.v2i1.693

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching, 6*(1), 57-63. https://doi.org/10.37074/jalt.2023.6.1.22

Fitria, T. N. (2023). Artificial intelligence (AI) technology in Open AI ChatGPT application: A review of ChatGPT in writing English essay. *ELT J. English Lang. Teach. 12*, 44-58. https://doi.org/10.15294/elt.v12i1.64069

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with

detectors and blinded human reviewers. *NPJ Digit Med., 6*(1), 75-82. https://doi.org/10.1038/s41746-023-00819-6

Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 56*(3), 23-31. https:// doi.org/10.1007/s10639-023-12146-0

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher educa¬tion level: A systematic review of the literature. *Contemporary Educational Technology, 15*(4), 46-58. https://doi.org/10.30935/cedtech/13605

Kannan J., & Munday P. (2018). New trends in second language learning and teaching through the lens of ICT: Networked learning, and artificial intelligence. *Círculo de Lingüística Aplicada a la Comunicación, 76,* 13-30. https://doi.org/10.5209/CLAC.62495

Khansir, A. A., & Gholami Dashti, J. (2013). The effect of question-generation strategy on Iranian EFL learners' reading comprehension development. *English Language Teaching, 7*(4), 38-45. https://doi.org/10.5539/elt.v7n4p38

Kim, S. G. (2023). Using ChatGPT for language editing in scientific articles. *Maxillofacial Plastic Reconstr. Surgery, 45*(2), 13-20. https://doi.org/10.1186/s40902-023-00381-x

Kim, S., Shim, J., & Shim, J. (2023). A study on the utilization of Open AI ChatGPT as a second language learning tool. *Journal of Multimedia Information System, 10*(1), 79-88. https://doi.org/10.33851/JMIS.2023.10.1.79

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC J. 54,* 537-550. https://doi.org/10.1177/00336882231162868

Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education & Technology (LET Journal), 3*(1), 55-72. https://langedutech.com/letjournal/index.php/let/article/view/48

Kovačević, D. (2023). *Use of ChatGPT in ESP teaching process*. In *Proceedings of the 22nd International Symposium INFOTEH-JAHORINA* (pp. 1–5). East Sarajevo, Bosnia and Herzegovina. https://doi.org/10.1109/INFOTEH57020.2023.10066303

Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News, 12*(2), 14-22. https://doi.org/10.1108/LHTN-01-2023-0009

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology, 74*(5), 570-581. https://doi.org/10.1002/asi.24750

Martín-García, A. V., Redolat, R., & Pinazo-Hernandis, S. (2022). Factors influencing intention to technological use in older adults: The TAM model application. *Research on Aging, 44*(7-8), 573-588. https://doi.org/10.1177/01640275211063797

Moses, R. N., & Mohamad, M. (2019). Challenges faced by students and teachers on writing skills in ESL contexts: A literature review. *Creative Education, 10*(13), 3385-3391. https://doi.org/10.4236/ce.2019.1013260

Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon, 7*(5), 19-29. https://doi.org/10.1016/j.heliyon.2021.e07014

Pavlik, J. V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator, 12*(3), 23-29. https://doi.org/10.1177/10776958221149577

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Sallam, M. (2023). The utility of ChatGPT as an example of large language models in healthcare education, research and practice: Systematic review on the future perspectives and potential limitations. *Healthcare (Basel), 11*(6), 887-896. https://doi.org/10.3390/healthcare11060887

Samadian, S., & Mohseny, A. (2019). Analysis of cohesion and coherence in writing performance of Iranian intermediate EFL learners. *Issues in Language Teaching, 8*(2), 213-242. https://doi.org/10.22054/ilt.2019.21262.193

Schmidt-Fajlik, R. (2023). ChatGPT as a grammar checker for Japanese English language learners: A comparison with Grammarly and pro writing aid. *Asia CALL Online J. 14,* 105-119. https://doi.org/10.54855/acoj.231417

Shahhoseini, H., Rezvani, R., Yazdani, S., Behrouzi, M., & Molaei, A. (2024). *Transitional problems in Iranian EFL learners' writing process: A think-aloud protocol study*. *Journal of Language and Translation, 14*(3), 191–204. https://doi.org/10.30495/TTLT.2024.1047695

Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). *ChatGPT and other large language models are double-edged swords.* Radiological Society of North America.

Sun, P. P., & Mei, B. (2022). Modeling preservice Chinese-as-a-second/foreign-language teachers' adoption of educational technology: A technology acceptance perspective. *Computer Assisted Language Learning, 35*(4), 816-839. https://doi.org/10.1080/09588221.2020.1750430

Winstone, N., & Carless, D. (2020). *Designing effective feedback processes in higher education.* Routledge.

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Educ. Inf. Technology, 25*(4), 1-25. https://doi.org/10.1007/s10639-023-11742-4

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1-27. https://doi.org/10.1186/s41239-019-0171-0

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., & He, L. (2023). *A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT.* https://doi.org/10.48550/arXiv.2302.09419

Zohrabi, M., & Khalili, A., (2024). A cross-cultural study into the utility of diverse written corrective feedback strategies in medicine students' ESP writing courses. *International Journal of Society, Culture & Language, 12*(2), 150-169. https://doi.org/10.22034/ijscl.2024.2025600.3436