Razi University

Technology Assisted Language Education

ISSN: 2981-1465

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

The Effect of Self-Regulated Use of ChatGPT on the Paragraph Writing Skills and Self-Efficacy of Iranian EFL Learners

Malihe Ghorbanzadeh Savar[1], **Hooshang Yazdani**[*2], and **Elham Farahani**

1. Department of English Language and Literature, Faculty of Literature and Languages, University of Arak, Iran. m.ghorbanzadeh.00@phd.araku.ac.ir
2. Corresponding author, Department of English Language and Literature, Faculty of Literature and Languages, University of Arak, Iran. h-yazdani@araku.ac.ir
3. Department of English Language and Literature, Faculty of Literature and Languages, University of Arak, Iran. e-farahani@araku.ac.ir

Article Info

Article type:
Research Article

Article history:
Received May 03, 2026
Received in revised form June 27, 2026
Accepted June 27, 2025
Published online June 30, 2026

Keywords:
ChatGPT, self-regulation, self-efficacy, paragraph writing skills, quantitative approach

Abstract

As technology has advanced rapidly, various AI tools have been integrated into educational contexts. This study investigates the effects of self-regulated use of ChatGPT on paragraph-writing skills and self-efficacy among Iranian EFL learners. To this end, 64 university students were assigned to experimental and control groups in two intact writing classes over a twelve-week intervention. The experimental group received training on self-regulated learning (SRL) strategies based on Zimmerman's (2002) established model. Key strategies covered included: goal setting, self-monitoring, self-evaluation and reflection, and strategic resource management. They also utilized ChatGPT for writing tasks, whereas the control group followed traditional instruction. Pre- and post-tests of compare-contrast and cause-effect paragraph writing were scored for organization, logical development, grammar, mechanics, and style. Self-efficacy was measured using a validated questionnaire. The results indicated significant differences in the post-test mean scores of the experimental and control groups on the compare-and-contrast and cause-and-effect sub-sections, both of which represent large effect sizes. The findings suggest that integrating AI-driven tools facilitates self-regulated learning, enhances writing proficiency, and strengthens learners' writing confidence. These results have implications for educational policy and curriculum development, underscoring that adopting AI tools complements traditional writing instruction by facilitating a more personalized learning experience.

Cite this article: Ghorbanzadeh Savar, M., Yazdani, H., & Farahani, E. (2026). The Effect of Self-Regulated Use of Chat GPT on the Paragraph Writing Skills and Self-Efficacy of Iranian EFL Learners. *Technology Assisted Language Education*, *4*(2), 27-48. doi: 10.22126/tale.2026.13857.1183

Publisher: Razi University

DOI: http//doi.org/10.22126/tale.2026.13857.1183

Introduction

Recent advances in AI-powered language tools have renewed interest in their pedagogical value. Among these tools, ChatGPT has been highlighted for its capacity to generate human-quality text, perform natural language processing, and support dialogue-based interaction. It can adjust to learners' proficiency levels, provide immediate feedback, and promote learner autonomy and self-correction (Chen et al. 2022). Furthermore, technology boosts learners' self-efficacy and autonomy, enabling them to control their learning process by providing a unique sanctuary- a place where learners can experiment, stumble, and grow without the weight of human judgment. (Gikas & Grant, 2013). Teaching writing skills with traditional methods is incredibly challenging because it requires extensive time, effort, and subjective interpretation to monitor and provide insightful feedback on student writing. Additionally, time constraints often diminish learners' motivation to improve their writing abilities. AI-assisted writing tools can enhance writing skills by providing automated feedback on aspects such as organization, logical development of ideas, grammar, mechanics, and the quality and style of expression.

Writing in a foreign language is a complex cognitive task for English as a Foreign Language (EFL) learners, requiring not only linguistic competence but also strategic planning, self-monitoring, and reflective revision. Consequently, self-regulated learning, the capacity to set goals, monitor progress, and adjust strategies, has been widely recognized as essential for effective writing development and for fostering learner autonomy. SRL frameworks emphasize metacognitive engagement, behavioral control, and motivational beliefs (Pintrich, 2004), all of which contribute to improved writing outcomes and heightened self-efficacy. Yet, despite the rapid integration of AI in language classrooms, the intersection of self-regulated use of AI tools with learners' paragraph writing skills and self-efficacy remains underexplored. Existing research on AI-assisted language learning has largely focused on the general impacts of AI tools on writing performance and motivation, with several mixed-method studies showing that learners perceive AI assistance as beneficial for generating ideas, checking grammar, and improving fluency (Gayed et al., 2022; Kohnke et al., 2023; Lin & Mubarok, 2021). However, such studies often do not explicitly examine how learners' self-regulated engagement with these tools shapes writing skill development and writing confidence. This gap is particularly salient in EFL contexts where cultural, educational, and linguistic factors influence how learners interact with technology. This study seeks to fill a gap at the intersection of technology, self-regulation, and paragraph writing, which has been underexplored, especially among Iranian EFL learners. To address these gaps, the following research questions are posed:

- Does the self-regulated use of AI-assisted tools, specifically ChatGPT, have any significant effect on the compare-contrast paragraph-writing skills of EFL learners?
- Does the self-regulated use of AI-assisted tools, specifically ChatGPT, have any significant effect on the cause-effect paragraph writing skills of EFL learners?

- Is there a significant difference in the level of self-efficacy between learners who use AI-assisted tools and those who do not?

The following null hypotheses were constructed to be tested in this study:

H01: Self-regulated use of AI-assisted tools, ChatGPT, has no significant effect on the comparison paragraph writing skills of EFL learners.

H02: Self-regulated use of AI-assisted tools ChatGPT has no significant effect on the cause-and-effect paragraph writing skills of EFL learners.

H03: There is no significant difference in the level of self-efficacy between learners who use AI- assisted tools and those who do not

Literature review

Self-Regulation and Self-Efficacy

Self-regulated learning is a core concept in educational psychology, referring to the ways learners systematically activate and sustain cognition, motivation, behavior, and affect to attain their learning goals (Teng, 2022). Students who effectively self-regulate their learning often achieve higher academic levels. (Broadbent et al., 2019; Richardson et al., 2012). This improved performance can be attributed to their ability to plan, organize, and manage their time and resources efficiently, allowing them to prioritize tasks, avoid procrastination, and engage in deeper, more meaningful learning (Panadero, 2017). Furthermore, SRL promotes intrinsic motivation and self-efficacy because students who believe in their ability to control their learning are more likely to persevere through challenges, see learning as a meaningful and rewarding experience, and develop a growth mind-set. (Dweck, 2006; Ryan & Deci, 2000).

Recently, the integration of Artificial Intelligence (AI) into language learning has renewed interest in SRL, as it can enhance personalized, adaptive learning experiences. AI-assisted language learning (AIALL) platforms can provide learners with personalized feedback customized learning paths, and adaptive practice opportunities. (Hwang et al., 2020). Therefore, improving SRL skills like goal setting, planning, and monitoring.

Several studies have reported positive relationships between technology-enhanced SRL and language learning outcomes. For example, Öztürk and Çakiroglu (2021) found that SRL facilitated learning in English speaking, reading, writing, and grammar in a study comparing two groups of university students with and without SRL strategies in a flipped learning context. Similarly, Zhu et al. (2024) found that students with SRL capabilities exhibited enhanced learning outcomes in blended learning settings.

Self-efficacy, a central component of social cognitive theory, refers to an individual's belief in their ability to succeed in specific situations or accomplish a task (Bandura, 1997). In the context of second language learning, self-efficacy refers to learners' belief in their ability to successfully achieve language proficiency. More recently, the rise of digital technologies and online learning platforms has led to studies on the role of self-efficacy in these environments.

Researchers have explored how online learning can be designed to improve self-efficacy, student engagement, and academic outcomes.

Recent studies have increasingly examined the use of AI tools like ChatGPT in language learning, highlighting both the potential benefits and drawbacks of it. Several studies have investigated ChatGPT's effectiveness as a scaffolding tool for idea generation, outlining, and providing feedback on paragraph structure and coherence (Kasneci et al., 2023; Lund & Wang, 2023). While initial results often demonstrate improvements in grammatical accuracy and stylistic fluency when learners use AI-assisted feedback (Farrokhnia et al., 2023), critical analyses reveal several nuanced concerns. The main problem is that students may become over reliant on AI-generated content, which could diminish their critical thinking skills and ability to express original ideas (Cotton et al., 2023). Additionally, because the rationale behind AI suggestions in large language models is often unclear, learners may struggle to develop a thorough understanding of the linguistic principles and writing strategies involved (Holmes et al., 2022). A study by Saenkhum & Kim (2023) showed that while ChatGPT can produce well-written paragraphs, students who relied on it too much often lacked a deep understanding of the topic and had difficulty evaluating the AI-generated content. While some believe AI hinders creative writing (Malik et al., 2023), research by Marzuki et al. (2023) found that EFL teachers reported AI writing tools as beneficial for stimulating creativity and idea expansion, with the tools offering students various angles and perspectives when they struggled with writer's block, thereby helping them overcome creative obstacles. A crucial point is that students require explicit instruction on assessing AI-generated content, using it responsibly and ethically, and finding a balance that leverages AI's benefits without compromising their independence and critical thinking abilities.

While many studies have examined the immediate effects of AI on learners' writing proficiency, there is limited research on its influence on self-efficacy and self-regulated learning strategies. Despite the valuable insights from current studies, a significant gap remains in understanding the impact of these interventions on learners' SRL, particularly within specific cultural and educational contexts. Existing literature frequently overlooks the critical role of learner agency and metacognitive strategies in optimizing the benefits of AI in language learning, leaving a gap in understanding how learners can effectively integrate these tools into their existing SRL processes to improve specific writing skills, such as paragraph construction.

Much of the recent research on AI in education has been conducted in Western, predominantly English-speaking contexts, where learner autonomy, classroom discourse, and access to digital tools may differ substantially from those in Iran (Hwang et al., 2020). Although the pedagogical principles of AI use may be broadly transferable, its affordances and challenges for EFL learners are shaped by local educational and cultural conditions (Godwin-Jones, 2024; Kohnke et al., 2023). In the Iranian EFL context, English is generally learned as a foreign language rather than a medium of everyday communication, which limits learners' exposure to authentic input outside the classroom and increases the influence of teachers, textbooks, and classroom instruction (Rezvani & Ketabi, 2011). Iranian schooling has traditionally been

characterized by teacher-centered pedagogy, grammar-focused instruction, and high-stakes assessment, which can limit learners' opportunities to develop self-regulated learning strategies independently, particularly in productive skills such as writing (Parviz et al., 2025). Hence, this paper aims to investigate how the self-regulated use of ChatGPT affects EFL learners' confidence in their paragraph writing abilities.

Method

Design

This study employs a quasi-experimental method design. It incorporates elements of a pre-test/post-test control-group design, which enables the establishment of a baseline for participants' writing skills and self-efficacy before the intervention. This allows a direct comparison of changes in the experimental and control groups, thereby enabling the determination of the effect of the independent variable.

Participants

The study was conducted in the English Language and Literature Department at Mofid University, Iran, over 12 weeks. The data was collected during class sessions of the course entitled Advanced Writing. Students who attend this course are usually at their sophomore year, as a prerequisite for enrolling in the Essay Writing course. The course was well-suited for collecting data for the present study because it was the first official writing course learners attended at the university level; thus, the likelihood that participants had any prior experience was quite low. Data collection was conducted in two intact classes, with populations of 33 and 31. The use of intact classes means that random assignment of individual students was not possible. However, efforts were made to mitigate potential biases associated with convenience sampling by ensuring homogeneity of the groups through pre-intervention proficiency assessments. The learners' writing pre-test was also considered as a criterion of homogeneity in writing proficiency. The participants were divided into two distinct groups based on their pre-existing class affiliations. The experimental group, comprising 33 students, received instruction supplemented by self-regulated use of AI-assisted tools (ChatGPT), while the control group, consisting of 31 students, received conventional writing instruction without access to the AI-assisted tools.

Instrumentation

English Language Proficiency Test

To ensure homogeneity in participants' general English proficiency at the beginning of the study, the Oxford Placement Test (OPT) was administered.

Volume 4. Issue 2. 2026. Pages 27 to 48.

Writing Pre-test and Post-test

To address the first two research questions of the present study, two writing prompts from the compare-contrast and cause-and-effect genres were chosen for the pre-test, and two writing prompts from the compare-contrast and cause-and-effect genres for the post-test.

Paragraph Writing Skills Assessment Rubric

This study employs the Categorical Instrument for Scoring Second Language Writing, developed by Brown and Bailey (1984). This instrument was chosen for its well-established reputation in the field, particularly for its demonstrated validity and reliability in discerning varying levels of writing competence.

The categorical instrument for scoring second-language writing skills uses five equally weighted criteria, each with a possible score of 20 points, for a total possible score of 100. The criteria are: 1- Organization, 2- Logical Development of Ideas, 3- Grammar, 4- Punctuation, Spelling, and Mechanics, 5- Style and Quality of Expression

To ensure inter-rater reliability, two raters scored the writing samples. Before the main scoring phase, raters participated in a detailed discussion of the rubric's criteria and pilot scoring of sample paragraphs to establish a shared understanding.

English Writing Self-Efficacy Questionnaire

Participants' self-efficacy in English writing was measured using the L2 Learners' Writing Self-efficacy Scale (LWSE). This questionnaire was developed based on the conceptual framework and sub-categories of writing self-efficacy as outlined by Tsao (2021) to assess second language learners' beliefs in their capabilities to perform various writing tasks. The LWSE comprises multiple sub-scales that capture different aspects of writing self-efficacy, including: Self-efficacy for writing ideation, Self-efficacy for writing conventions, Self-efficacy for writing self-regulation. The questionnaire utilizes a Likert-type scale, typically ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

Data Collection Procedure

Sixty-four lower-intermediate learners based on the result of OPT that was administered in the first session of the class in two intact university writing classes engaged in paragraph writing tasks during twelve-weeks in a semester. The learners wrote four paragraphs (except pre and post- test of paragraphs) during the course. The intervention began with explicit instruction on self-regulated learning strategies related to the writing process. This training was based upon established models of SRL in academic contexts (Zimmerman, 2002). The SRL training phase was implemented over 5 sessions. Key strategies covered included:

- Session 1: Goal setting, planning, and task analysis
- Session 2: Strategic prompting and idea generation
- Session 3: Monitoring, self-questioning, and draft control

- Session 4: Revision strategies and rubric-based evaluation
- Session 5: Reflection, self-evaluation, and transfer to independent use

The SRL intervention was delivered in a structured, teacher-monitored format to ensure the experimental group consistently followed the target strategies. During each of the five sessions, students were guided through explicit instruction, modeling, guided practice, and reflection, and their use of SRL strategies was monitored through in-class observation, task completion checks, and instructor feedback. ChatGPT use was also regulated during the intervention to ensure that it served as a support tool rather than a substitute for SRL engagement. In addition, both the experimental and control groups received equal instructional time to control for exposure effects and ensure that any differences in outcomes could be attributed to the SRL-based intervention rather than differences in time on task. After the completion of the twelve-week intervention period, a post-test phase was administered to both the experimental and control groups

Data analysis

The data collected for this study were analyzed using an independent-samples t-test and a Multivariate Analysis of Covariance (MANCOVA). These statistical techniques require that the data be normally distributed. The normality of the data was probed using skewness and kurtosis indices, which examine relative symmetry and peakedness, respectively. The reliability indices for the instruments employed in this study were computed using KR-21, Cronbach's alpha, and Pearson's correlation (inter-rater reliability). four sets of Pearson correlations were run to probe inter-rater reliability indices for pre-tests and post-tests of sub-sections on cause and effect and compare and contrast.

(MANCOVA) was run to compare the experimental and control groups' mean scores on post-tests for the sub-sections of compare and contrast and cause and effect paragraphs. The linear relationships between pre-tests and post-tests for sub-sections of compare-and-contrast and cause-and-effect paragraphs were measured. A one-way ANCOVA was run to compare the two groups' mean post-test self-efficacy scores, controlling for the pre-test.

Findings

The reliability indices for the instruments employed in this study were computed using KR-21, Cronbach's alpha, and Pearson's correlation (inter-rater reliability). The results of KR-21 reliability index for the OPT test. indicated that the OPT test enjoyed a KR-21 reliability index of .74.

Four sets of Pearson correlations were run to probe the inter-rater reliability indices for pre-tests and post-tests of the cause-and-effect and compare-and-contrast sub-sections. Table 4.3 shows the inter-rater reliability indices for pre-tests of compare-and-contrast. Based on these results it can be concluded that there were significant agreements between the two raters on pre-tests of sub-sections of compare and contrast; i.e. organization (r (62) = .736 representing

a large effect size, $p < .05$), development of an idea ($r (62) = .637$ representing a large effect size, $p < .05$), grammar ($r (62) = .747$ representing a large effect size, $p < .05$), mechanics ($r (62) = .561$ representing a large effect size, $p < .05$), and style ($r (62) = .626$ representing a large effect size, $p < .05$).

To probe the first hypothesis, a MANCOVA was run to compare the experimental and control groups' mean scores on post-tests for the sub-sections of compare and contrast: organization, the logical development of ideas, grammar, mechanics, and style and quality of expression.

First, MANCOVA requires linear relationships between pre-tests and post-tests of sub-sections of compare and contrast. The significant results rejected the statistical assumption that the relationships between the pre-tests and post-tests were linear. There were linear relationships between pre-tests and post-tests of; organization ($F (1, 63) = 49.33$, $p < .05$, eta-squared = .555 representing a large effect size), logical development of idea ($F (1, 63) = 53.24$, $p < .05$, eta-squared = .559 representing a large effect size), grammar ($F (1, 63) = 81.77$, $p < .05$, eta-squared = .673 representing a large effect size), mechanics ($F (1, 63) = 31.92$, $p < .05$, eta-squared = .399 representing a large effect size), and style and quality of expression ($F (1, 63) = 27.83$, $p < .05$, eta-squared = .362 representing a large effect size).

Second, MANCOVA requires that linear relationships hold across the two groups; i.e., the assumption of homogeneity of regression slopes. The non-significant interaction ($F (10, 94) = .845$, $p > .05$, Partial eta squared = .082 representing a moderate effect size) between the independent variable, i.e. groups, and covariates, i.e. pre-tests of sub-sections of compare and contrast, indicated that the assumption of homogeneity of regression slopes was retained.

Third, the results of Levene's test of homogeneity of variances indicated that except for post-test of grammar ($F (1, 62) = 1.29$, $p > .05$), the assumption of homogeneity of variances was not retained on post-tests of organization ($F (1, 62) = 4.00$, $p < .05$), logical development of idea ($F (1, 62) = 16.56$, $p < .05$), mechanics ($F (1, 62) = 4.90$, $p < .05$), and style and quality of expression ($F (1, 62) = 4.65$, $p > .05$). In case the assumption of homogeneity of variances is not retained, one can report the main results of MANCOVA at .01 levels instead of .05 (Tabachnick & Fidell, 2019). That was why the MANCOVA results were reported at the .01 level.

Finally, the results of the Box's test of homogeneity of covariance matrices (Box's $M = 15.38$, $p > .001$) indicated that the assumption of homogeneity of covariance matrices was retained. It should be noted that the results are reported at the .001 level (Pallant, 2016; Tabachnick & Fidell, 2019; and Field, 2024).

Table 1

Descriptive Statistics for Post-tests of Sub-Sections of Compare and Contrast by Group with Pre-tests

Dependent Variable	Groups	Mean	SD	95% Confidence Interval	
				Lower Bound	Upper Bound
Post-CCOrg	Experimental	15.652[a]	2.210	15.079	16.226
	Control	12.596[a]	2.863	12.003	13.189
Post-CCDev	Experimental	14.057[a]	2.586	13.447	14.668
	Control	11.423[a]	2.719	10.792	12.053
Post-CCGram	Experimental	16.515[a]	2.077	16.021	17.009
	Control	14.419[a]	2.541	13.909	14.929
Post-CCMech	Experimental	14.422[a]	2.948	13.721	15.123
	Control	13.002[a]	2.496	12.278	13.727
Post-CCStyle	Experimental	14.604[a]	2.208	14.083	15.125
	Control	11.841[a]	2.243	11.303	12.379

a. Covariates appearing in the model are evaluated at the following values: Pre-CCOrg = 10.94, Pre-CCDev = 10.20, Pre-CCGram = 12.70, Pre-CCMech = 12.42, Pre-CCStyle = 10.39.

Table 1 shows the mean scores for the experimental and control groups on post-tests of the sub-sections of compare and contrast, after controlling for the effect of the pre-tests. The results indicated that the experimental group had higher mean scores than the control group.

Table 2

Multivariate Analysis of Covariance for Post-tests of Sub-Sections of Compare and Contrast by Group with Pre-tests

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.472	9.473	5	53	.000	.472
	Wilks' Lambda	.528	9.473	5	53	.000	.472
	Hotelling's Trace	.894	9.473	5	53	.000	.472
	Roy's Largest Root	.894	9.473	5	53	.000	.472
Group	Pillai's Trace	.556	13.287	5	53	.000	.556
	Wilks' Lambda	.444	13.287	5	53	.000	.556
	Hotelling's Trace	1.254	13.287	5	53	.000	.556
	Roy's Largest Root	1.254	13.287	5	53	.000	.556

The results of MANCOVA *(F (5, 53) = 13.28, p < .01*, partial eta squared = .556, representing a large effect size) indicated that there were significant differences between the experimental

and control groups' mean scores on post-tests of sub-sections of compare and contrast after controlling for the effect of pre-tests. Thus, the first hypothesis was rejected.

Table 3

Tests of Between-Subjects Effects for Post-tests of Sub-Sections of Compare and Contrast by Group with Pre-tests

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Groups	Post-CCOrg	138.714	1	138.714	53.085	.000	.482
	Post-CCDev	103.076	1	103.076	34.882	.000	.380
	Post-CCGram	65.205	1	65.205	33.707	.000	.372
	Post-CCMech	29.935	1	29.935	7.673	.008	.119
	Post-CCStyle	113.357	1	113.357	52.601	.000	.480
Error	Post-CCOrg	148.944	57	2.613			
	Post-CCDev	168.436	57	2.955			
	Post-CCGram	110.264	57	1.934			
	Post-CCMech	222.382	57	3.901			
	Post-CCStyle	122.837	57	2.155			
Total	Post-CCOrg	13379.000	64				
	Post-CCDev	10992.000	64				
	Post-CCGram	15766.000	64				
	Post-CCMech	12555.000	64				
	Post-CCStyle	11661.000	64				

Table 3 shows the results of the Between-Subjects Effects, which compare the two groups' mean scores for each subsection of compare and contrast. It can be concluded that:

- The experimental group (M = 15.62) had a significantly higher mean score than the control group (M = 12.59) on the post-test of the organization sub-section of compare and contrast ($F(1, 57) = 53.08$, $p < .05$, partial eta squared = .482, representing a large effect size).
- The experimental group (M = 14.05) had a significantly higher mean score than the control group (M = 11.42) on post-test of logical development of idea sub-section of compare and contrast ($F(1, 57) = 34.86$, $p < .05$, partial eta squared = .380 representing a large effect size).

- The experimental group (M = 16.51) had a significantly higher mean score than the control group (M = 14.41) on the post-test of the grammar sub-section of compare and contrast ($F(1, 57) = 33.70$, $p < .05$, partial eta squared = .372, representing a large effect size).
- The experimental group (M = 14.42) had a significantly higher mean score than the control group (M = 13.00) on the post-test of the mechanic sub-section of compare and contrast ($F(1, 57) = 7.67$, $p < .05$, partial eta squared = .119, representing a moderate effect size).

Finally, the experimental group (M = 14.60) had a significantly higher mean score than the control group (M = 11.84) on post-test of style and quality of expression sub-section of compare and contrast ($F (1, 57) = 52.60$, $p < .05$, partial eta squared = .480 representing a large effect size).

In order to probe the second hypothesis, MANCOVA was run to compare the experimental and control groups' mean scores on post-tests of sub-sections of the cause and effect paragraph;

First, MANCOVA requires linear relationships between pre-tests and post-tests of sub-sections of cause and effect. The significant results rejected the statistical assumption that the relationships between the pre-tests and post-tests were linear. There were linear relationships between pre-tests and post-tests of; organization ($F (1, 63) = 39.38$, $p < .05$, eta-squared = .524 representing a large effect size), logical development of idea ($F (1, 63) = 42.16$, $p < .05$, eta-squared = .500 representing a large effect size), grammar ($F (1, 63) = 13.28$, $p < .05$, eta-squared = .388 representing a large effect size), mechanics ($F (1, 63) = 22.80$, $p < .05$, eta-squared = .398 representing a large effect size), and style and quality of expression ($F (1, 63) = 35.76$, $p < .05$, eta-squared = .451 representing a large effect size).

Second, MANCOVA requires that linear relationships hold across the two groups; i.e., the assumption of homogeneity of regression slopes. The non-significant interaction ($F (10, 94) = 1.34$, $p > .05$, Partial eta squared = .125 representing a moderate effect size) between the independent variable, i.e., groups, and covariates, i.e., pre-tests of sub-sections of cause and effect, indicated that the assumption of homogeneity of regression slopes was retained.

Third, the results of Levene's test of homogeneity of variances. indicated that except for post-tests of grammar ($F (1, 62) = .230$, $p > .05$), and style and quality of expression ($F (1, 62) = 2.94$, $p > .05$), the assumption of homogeneity of variances was not retained on post-tests of organization ($F (1, 62) = 4.37$, $p < .05$), logical development of idea ($F (1, 62) = 5.74$, $p < .05$), and mechanics ($F (1, 62) = 11.56$, $p < .05$). In case the assumption of homogeneity of variances is not retained, one can report the main results of MANCOVA at .01 levels instead of .05 (Tabachnick & Fidell, 2019). That was why the MANCOVA results were reported at the .01 level.

Finally, the results of Box's test of homogeneity of covariance matrices. (Box's M = 22.71, $p > .001$) indicated that the assumption of homogeneity of covariance matrices was retained. It

should be noted that the results reported at the .001 level (Pallant, 2016; Tabachnick & Fidell, 2019; and Field, 2024).

Table 4

Descriptive Statistics for Post-tests of Sub-Sections of Cause and Effect by Group with Pre-tests

Dependent Variable	Groups	Mean	SD	95% Confidence Interval	
				Lower Bound	Upper Bound
Post-CEOrg	Experimental	15.577[a]	2.612	14.889	16.264
	Control	13.160[a]	2.658	12.450	13.871
Post-CEDev	Experimental	13.796[a]	2.322	13.232	14.360
	Control	11.766[a]	2.252	11.183	12.349
Post-CEGram	Experimental	16.677[a]	1.838	16.175	17.179
	Control	14.247[a]	2.106	13.728	14.766
Post-CEMech	Experimental	13.808[a]	2.575	13.121	14.496
	Control	12.720[a]	1.973	12.009	13.431
Post-CEStyle	Experimental	14.854[a]	2.583	14.259	15.448
	Control	11.543[a]	2.860	10.928	12.158

a. Covariates appearing in the model are evaluated at the following values: PreCEOrg = 10.33, PreCEDev = 9.61, PreCEGram = 11.84, PreCEMech = 11.52, PreCEStyle = 9.06.

Table 4 shows the mean scores for the experimental and control groups on post-tests of cause-and-effect subsections, after controlling for pre-test effects. The results indicated that the experimental group had higher mean scores than the control group on post-tests of the cause-and-effect sub-sections, after controlling for pre-test scores.

Table 5

Multivariate Analysis of Covariance for Post-tests of Sub-Sections of Cause and Effect by Group with Pre-tests

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.647	19.469	5	53	.000	.647
	Wilks' Lambda	.353	19.469	5	53	.000	.647
	Hotelling's Trace	1.837	19.469	5	53	.000	.647
	Roy's Largest Root	1.837	19.469	5	53	.000	.647
Group	Pillai's Trace	.513	11.179	5	53	.000	.513
	Wilks' Lambda	.487	11.179	5	53	.000	.513
	Hotelling's Trace	1.055	11.179	5	53	.000	.513

	Roy's Largest Root	1.055	11.179	5	53	.000	.513

Table 5 shows the main results of MANCOVA. The results ($F(5, 53) = 11.17$, $p < .01$, partial eta squared = .513, representing a large effect size) indicated that there were significant differences between the experimental and control groups' mean scores on post-tests of sub-sections of cause and effect. Thus, the second hypothesis was rejected.

Table 6

Tests of Between-Subjects Effects for Post-tests of Sub-Sections of Cause and Effect by Group with Pre-tests

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Groups	Post-CEOrg	80.944	1	80.944	22.361	.000	.282
	Post-CEDev	57.113	1	57.113	23.448	.000	.291
	Post-CEGram	81.857	1	81.857	42.359	.000	.426
	Post-CEMech	16.408	1	16.408	4.528	.038	.074
	Post-CEStyle	151.951	1	151.951	56.090	.000	.496
Error	Post-CEOrg	206.327	57	3.620			
	Post-CEDev	138.836	57	2.436			
	Post-CEGram	110.150	57	1.932			
	Post-CEMech	206.532	57	3.623			
	Post-CEStyle	154.417	57	2.709			
Total	Post-CEOrg	13792.000	64				
	Post-CEDev	10888.000	64				
	Post-CEGram	15696.000	64				
	Post-CEMech	11640.000	64				
	Post-CEStyle	11856.000	64				

Table 6 shows the results of the Between-Subjects Effects, which compare the two groups' mean scores for each subsection of cause and effect. It can be concluded that:

- The experimental group (M = 15.57) had a significantly higher mean score than the control group (M = 13.16) on the post-test of the organization subsection of cause and

effect (F(1, 57) = 22.36, p < .05, partial eta squared = .282, representing a large effect size).

- The experimental group (M = 13.97) had a significantly higher mean score than the control group (M = 11.76) on post-test of logical development of idea sub-section of cause and effect (F (1, 57) = 23.42, p < .05, partial eta squared = .291 representing a large effect size).
- The experimental group (M = 16.67) had a significantly higher mean score than the control group (M = 14.24) on the post-test of the grammar sub-section of cause and effect (F(1, 57) = 42.35, p < .05, partial eta squared = .426, representing a large effect size).
- The experimental group (M = 13.80) had a significantly higher mean score than the control group (M = 12.72) on the post-test of the mechanic sub-section of cause and effect (F(1, 57) = 4.52, p < .05, partial eta squared = .074, representing a moderate effect size).
- Finally, the experimental group (M = 14.85) had a significantly higher mean score than the control group (M = 11.54) on post-test of style and quality of expression sub-section of cause and effect (F (1, 57) = 56.09, p < .05, partial eta squared = .496 representing a large effect size).

To test the third hypothesis, the researcher attempted to run a One-Way ANCOVA to compare the two groups' post-test self-efficacy scores, controlling for the pre-test; however, the assumption of homogeneity of regression slopes was not met. That was why two separate Independent-Samples t-tests were run to compare the two groups' mean scores on the pre-test and post-test of self-efficacy.

Table 7

Testing Homogeneity of Regression Slopes for Pre-test and Post-test of Self-Efficacy

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	1081.442	1	1081.442	37.538	.000	.385
PreSelfEfficacy	1488.914	1	1488.914	51.682	.000	.463
Group * PreSelfEfficacy	382.937	1	382.937	13.292	.001	.181
Error	1728.552	60	28.809			
Total	257107.000	64				

An Independent-Samples t-test was run to compare the experimental and control groups' mean pre-test self-efficacy scores to demonstrate that the two groups were homogeneous in terms of self-efficacy prior to treatment. The results indicated that the two groups were virtually the same at the outset: the experimental group scored an average of 50.18 (SD = 9.80), while the control group scored 48.19 (SD = 9.41).

The results of the inferential statistical test are displayed in Table 8. The analysis is presented in two rows, and for this study, the results from the first row—"Equal variances assumed"—were selected for interpretation. This decision was justified by the non-significant outcome of Levene's Test for Equality of Variances ($F = .133$, $p > .05$), which confirmed that the assumption of homogeneity of variances between the two groups was met.

The Independent-Samples t-test yielded a non-significant result ($t(62) = .903$, $p > .05$), with a calculated Cohen's d effect size ($d = .226$) indicating a weak magnitude. This confirms that there was no statistically significant difference between the mean scores of the experimental and control groups on pre-test of self-efficacy.

Table 8

Independent Samples Test for Pre-test of Self-Efficacy by Group

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.404	.528	.903	62	.370	2.370	2.625	-2.879	7.618
Equal variances not assumed			.900	60.431	.372	2.370	2.634	-2.898	7.637

Cohen's d = .226

An Independent-Samples t-test was run to compare the experimental and control groups' mean post-test scores on self-efficacy to probe the third hypothesis. The results indicated that the experimental group ($M = 70.70$, $SD = 7.34$) had a higher mean than the control group ($M = 53.42$, $SD = 8.12$) on post-test of self-efficacy.

Table 9

Descriptive Statistics for Post-test of Self-Efficacy by Group

	Group	N	Mean	Std. Deviation	Std. Error Mean
Post-Self-Efficacy	Experimental	33	70.70	7.346	1.279
	Control	31	53.42	8.127	1.460

The results of the inferential statistical test are displayed in Table 10. the results from the first row—"Equal variances assumed"—were selected for interpretation. This decision was justified by the non-significant outcome of Levene's Test for Equality of Variances ($F = .563$, $p > .05$),

which confirmed that the assumption of homogeneity of variances between the two groups was met.

The Independent-Samples t-test yielded a significant result ($t(62) = 8.93$, $p < .05$), with a calculated Cohen's d effect size ($d = 2.23$) indicating a large effect. This confirms that the experimental group had a significantly higher mean score than the control group on post-test of self-efficacy. Therefore, it can be confidently concluded that the third hypothesis was rejected.

Table 10

Independent Samples Test for Post-test of Self-Efficacy by Group

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.563	.456	8.932	62	.000	17.278	1.934	13.411	21.145
Equal variances not assumed			8.903	60.379	.000	17.278	1.941	13.396	21.159

Cohen's d = 2.23

Discussion

Regarding the first research question, the MANCOVA results indicated robust, statistically significant advantages for the experimental group across all subskills after controlling for pre-test scores; the first hypothesis was therefore rejected.

Self-regulated use of ChatGPT improved rhetorical organization and idea development. The large effects on organization (partial $\eta 2 = .482$) and logical development (partial $\eta 2 = .380$) suggest that when learners engage with ChatGPT in a self-regulated manner, they are better able to construct coherent compare-and-contrast paragraphs. This aligns with SRL frameworks (Zimmerman, 2002). ChatGPT likely provided immediate model texts, structural templates, and opportunities for iterative revision, enabling learners to internalize discourse patterns and apply them to organize comparisons and elaborations. The observed gains can also be situated within sociocultural theory (Vygotsky, 1978). ChatGPT functioned as a mediating tool that augmented learners' zone of proximal development by providing proximal support for structuring arguments and signaling cohesive moves (e.g., topic sentence, comparison markers, concluding synthesis).

Grammar and mechanics improved through iterative, scaffolded feedback. Significant gains in grammar (partial η2 = .372) and mechanics (partial η2 = .119) indicate that self-regulated AI use improved fundamental linguistic accuracy and surface conventions. ChatGPT's ability to generate grammatically polished examples and to offer corrective suggestions likely supported noticing and form-focused practice (Schmidt, 1990). The larger effect for grammar relative to mechanics suggests ChatGPT is especially helpful for syntactic and morphosyntactic restructuring (rewording sentences, correcting tense/aspect) while mechanical conventions (punctuation, spacing) improved to a lesser degree—consistent with the moderate effect seen for mechanics.

The very large effect for style and quality (partial η2 = .480) points to substantial improvements in learners' expressive resources—lexical sophistication, idiomatic phrasing, and cohesive linking. ChatGPT's exposure to vast written corpora likely provided varied stylistic exemplars that learners could emulate. From a cognitive apprenticeship perspective (Collins et al., 1989), ChatGPT provided models of expert-like expression; self-regulated learners could observe, imitate, and adapt these models in iterative drafting, improving rhetorical fluency and expression.

The combination of SRL and AI affordances appears crucial. Self-regulation likely determined the quality of interaction. learners who planned prompts, evaluated AI outputs, and revised their drafts actively could harness ChatGPT's potential for feedback, exemplification, and idea generation. Prior studies (Borenstein & Howard, 2021; Mollick & Mollick, 2023; Selwyn, 2022) indicate that uncontrolled reliance on AI can produce surface-level gains or overdependence; however, the present results show substantial learning when AI use is self-regulated—suggesting that learner agency mediates effective uptake of AI-provided support.

The present findings corroborate and extend recent research showing positive effects of AI-assisted writing tools on L2 composition (Gao & Zhang, 2023; Li & Kormos, 2024; Wang & Lee, 2022). Unlike studies reporting mixed or limited gains, which often occur when AI is used passively or without metacognitive engagement, this study's explicit, self-regulated use of AI design aligns with the literature that emphasizes a guided, reflective use of digital tools. The significant effects for organization and style here are notably large compared with many intervention studies, suggesting that ChatGPT's generative capacity, when coupled with learners' self-regulatory practices, can lead to pronounced writing development.

Regarding the second research question, the MANCOVA results indicated robust, statistically significant advantages for the experimental group across all subskills after controlling for pre-test scores; the second hypothesis was therefore rejected.

These findings align with the current literature, which suggests that AI technologies, such as ChatGPT, can provide targeted feedback, thereby enhancing learners' writing abilities (Huang & Mizumoto, 2024). By facilitating self-regulated learning, AI enables learners to actively engage with their writing processes, reflect on their performance, and make informed revisions based on instant feedback. This process fosters a deeper understanding of writing

conventions and encourages learners to develop critical thinking skills—with specific attention to constructing coherent topic sentences and developing arguments cohesively.

The significant improvements observed in the organization and logical development of ideas may suggest that ChatGPT effectively helps students structure their thoughts in a logical sequence, clarifying the relationships between components of their writing. Moreover, the enhanced performance in grammar and style indicates that the AI tool can assist learners in refining their language use, promoting a more effective communication style, which is vital for EFL learners aiming to succeed in diverse linguistic contexts (Liu et al. 2023).

The findings from the Independent-Samples t-test analysis clearly refuted the third hypothesis. From a theoretical perspective, this pattern is consistent with sociocognitive accounts of language learning in which performance is influenced by the dynamic interaction among cognitive processes, behavioral regulation, and learners' beliefs about their competence. When learners are given repeated opportunities to plan, draft, monitor, and revise with ChatGPT in a self-directed manner, they are not simply receiving feedback; they are actively participating in a cycle of observation, comparison, decision-making, and revision that can strengthen both writing performance and perceived self-efficacy. Unlike teacher-fronted feedback that is often delayed, limited, or externally controlled, ChatGPT provides immediate, individualized, and low-stakes responses that learners can consult repeatedly. This immediacy may reduce uncertainty during paragraph composition and help learners notice problems in organization, coherence, vocabulary choice, and grammatical accuracy as they occur.

In self-regulated learning terms, such support may enhance metacognitive monitoring and strategic control, because learners are prompted to evaluate their own drafts, compare alternative phrasings, and make conscious revisions. Over time, this recursive process can cultivate a sense of ownership over writing, especially relevant in contexts where learners may have limited confidence in their English writing and few opportunities for extended feedback.The improvement in paragraph writing skills may also be linked to reduced cognitive load. For Iranian EFL learners, writing in English often requires simultaneous attention to content generation, discourse organization, lexical retrieval, grammatical encoding, and mechanical accuracy. ChatGPT can partially externalize some of these burdens by offering language models, reformulations, and organizational suggestions, thereby freeing cognitive resources for higher-order composition. Importantly, when this support is self-regulated rather than fully directive, learners remain engaged in decision-making rather than passively copying generated output. This distinction matters because learning is more likely when learners must evaluate feedback, select revisions, and justify their choices. In this sense, the tool may function as a scaffold that supports performance while still preserving productive struggle.

The observed results align with prior research suggesting that technology-enhanced feedback can contribute positively to learners' agency and self-perception (Huang & Mizumoto, 2024). The personalized nature of AI feedback allows learners to engage in self-regulation, fostering reflection on their writing strategies and strengthening their confidence in their

competencies (Van Horn, 2024). In contrast, traditional teacher feedback, while invaluable, may not offer the same level of individualized engagement or immediacy as AI tools, potentially contributing to lower self-efficacy scores among students who rely solely on human feedback.

Moreover, the findings illuminate the growing relevance of AI in educational contexts, particularly in language learning, where developing self-efficacy is crucial for learners navigating complex language tasks (Athanassopoulos et al., 2023). The immediate feedback provided by AI can help learners identify and rectify errors more efficiently, fostering a sense of accomplishment and reinforcing their motivation to improve.

Conclusion

Sixty-four lower-intermediate EFL learners, determined by the Oxford Placement Test (OPT), participated in this study in two intact university writing classes. During a 12-week semester, learners completed paragraph-writing tasks and produced four instructional paragraphs, in addition to pre- and post-test paragraphs. The intervention began with explicit instruction in SRL strategies for the writing process, grounded in established SRL models (Zimmerman, 2002). The training focused on strategies such as goal setting and planning, self-monitoring, self-evaluation, reflection, and strategic resource management.

To address the first two research questions, participants completed pretests involving compare-and-contrast and cause-and-effect paragraph writing to measure their baseline writing ability. After the intervention, both experimental and control groups completed a comparable post-test writing task. Learners' writing was assessed using Brown and Bailey's (1984) Categorical Instrument for Scoring Second Language Writing, which evaluates organization, logical development of ideas, grammar, mechanics, and style. Multivariate Analysis of Covariance (MANCOVA) was conducted to compare the groups' post-test scores. The results showed that the experimental group achieved higher mean scores than the control group across the assessed components in both paragraph types.

To examine the third research question, the LWSE was administered before and after the intervention. Independent-samples t-tests revealed no significant difference between groups at the pre-test stage, but indicated higher post-test self-efficacy scores for the experimental group.

The findings suggest that integrating AI-based tools such as ChatGPT into writing instruction can provide timely feedback and support more autonomous, self-regulated learning. Interaction with such tools may help learners improve paragraph organization and development through revision and practice while also strengthening their confidence in writing.

One of the primary methodological limitations of the present study is the use of intact university classes rather than random assignment to experimental and control groups. Since learners were already enrolled in existing university classes, true random assignment was not feasible. However, the use of intact classes is both a practical and an ethically sound necessity in educational research conducted within institutional settings. Disrupting existing class structures to achieve random assignment would have been logistically impractical and ethically

problematic, as it would interfere with the regular academic schedule and students' right to continuity in their learning environment. This approach is widely accepted and commonly used in quasi-experimental EFL research (Dörnyei, 2007; Creswell, 2012), in which the natural classroom serves as the primary unit of instruction. Moreover, pre-tests were administered to both groups prior to the intervention to verify initial equivalence, thereby partially mitigating the threat to internal validity posed by non-random assignment. Although the study aimed to examine self-regulated use of ChatGPT, the actual degree to which participants regulated their interaction with the tool — including goal-setting, monitoring, and self-evaluation — could not be fully observed or verified. Individual differences in metacognitive skills and self-regulation strategies may have varied considerably among participants, introducing variability that the study design could not entirely account for.

Future research could examine the effectiveness of ChatGPT across different educational levels and proficiency groups, as well as investigate its long-term impact on writing development and self-efficacy employing true experimental designs with random assignment of participants to experimental and control groups, where institutional and ethical conditions permit. Additional studies may also explore contextual factors such as access to technology, familiarity with digital tools, and learner autonomy, and compare ChatGPT with other AI-assisted writing technologies.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W. H. Freeman.

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics, 1*(1), 61–65. https://doi.org/10.1007/s43681-020-00002-7

Broadbent, J., Panadero, E., Lodge, J. M., & Elliott, S. (2019). The role of academic self-efficacy in the relationship between self-regulated learning and academic achievement. *Educational Psychology Review, 31*(3), 975–996. https://doi.org/10.1007/s10648-019-09469-9

Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. *Language Testing, 1*(1), 21–38.

Chen, Z., Chen, W., Jia, J., & Le, H. (2022). Exploring AWE-supported writing processes: An activity theory perspective. *Language Learning & Technology, 26*(2), 129–148. https://hdl.handle.net/10125/73495

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction* (pp. 453–494). Lawrence Erlbaum.

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 60*(5), 1–12. https://doi.org/10.1080/14703297.2023.2190148

Creswell, J. (2007). *Qualitative inquiry and research design: Choosing among fiveapproaches. Thousand Oaks*, CA: Sage.

Dörnyei, Z. (2007). *Research methods in applied linguistics. Quantitative, qualitativeand mixed methodologies*. Oxford: Oxford University Press.

Dweck, C. S. (2006). *Mindset: The new psychology of success.* Random House.

Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, advance online publication. https://doi.org/10.1080/14703297.2023.2195846

Field, A. (2024). *Discovering statistics using IBM SPSS statistics* (7th ed.). Sage.

Gao, L., & Zhang, L. J. (2023). Understanding university students' experiences and perceptions of ChatGPT: A case study in an EFL writing context. *Frontiers in Psychology, 14*, Article 1222829. https://doi.org/10.3389/fpsyg.2023.1222829

Gayed, J. M., Jonson Carlon, M. K., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence, 3*, Article 100055. https://doi.org/10.1016/j.caeai.2022.100055

Gikas, J., & Grant, M. M. (2013). Mobile computing devices in higher education: Student perspectives on learning with cellphones, smartphones, and social media. *The Internet and Higher Education, 19*, 18–26. https://doi.org/10.1016/j.iheduc.2013.06.002

Holmes, W., Bialik, M., & Fadel, C. (2022). *Artificial intelligence in education: Promises and implications for teaching and learning.* Center for Curriculum Redesign.

Huang, J., & Mizumoto, A. (2024). Examining the effect of generative AI on students' motivation and writing self-efficacy. *Digital Applied Linguistics*, *1*, 102324. https://doi.org/10.29140/dal.v1.102324

Hwang, G. J., Xie, H., & Chen, G. D. (2020). Definition, roles, and potential research issues of artificial intelligence in education: An overview. *Computers & Education: Artificial Intelligence, 1*, Article 100001. https://doi.org/10.1016/j.caeai.2020.100001

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Nerdel, C. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Li, Z., & Kormos, J. (2024). The effects of using ChatGPT on self-efficacy and writing performance in English as a foreign language. *Journal of Second Language Writing, 65*, Article 101996. https://doi.org/10.1016/j.jslw.2024.101996

Lin, C.-J., & Mubarok, H. (2021). Learning Analytics for Investigating the Mind Map-Guided AI Chatbot Approach in an EFL Flipped Speaking Classroom. *Educational Technology & Society*, *24*(4), 16–35. https://www.jstor.org/stable/48629242

Liu, Q., He, X., Li, M., Huang, J., Gan, W., & Wu, Z. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized

controlled trial. *Frontiers in Psychology*, *14*, Article 1249991. https://doi.org/10.3389/fpsyg.2023.1249991

Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News, 40*(3), 26–29. https://doi.org/10.1108/LHTN-01-2023-0009

Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., & Darwis, A. (2023). Exploring artificial intelligence in academic essays: Higher education students' perspectives. *International Journal of Educational Research Open, 5*, Article 100296. https://doi.org/10.1016/j.ijedro.2023.100296

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

Mollick, E. R., & Mollick, L. (2023). *Using AI to implement effective teaching strategies in classrooms: Five strategies, including prompts.* SSRN. https://doi.org/10.2139/ssrn.4391243

Öztürk, M., & Çakıroğlu, Ü. (2021). Flipped learning design in EFL classrooms: Implementing self-regulated learning strategies to develop language skills. *Smart Learning Environments, 8*, Article 2. https://doi.org/10.1186/s40561-020-00147-8

Pallant, J. (2016). *SPSS survival manual* (6th ed.). Allen & Unwin.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, Article 422. https://doi.org/10.3389/fpsyg.2017.00422

Parviz, M., Hosseini, S. A., & Rezaei, S. (2025). AI-driven vs. traditional language assessment: Effects on Iranian EFL learners' motivation, anxiety, and proficiency in a high-stakes exam context. *Language Testing in Asia*, *15*, Article 12. https://doi.org/10.1186/s40468-025-00395-4

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*, 385–407. https://doi.org/10.1007/s10648-004-0006-x

Rezvani, E., & Ketabi, S. (2011). On the effectiveness of using web-and print-based materials in teaching grammar to Iranian EFL learners. *Procedia-Social and Behavioral Sciences*, *15*, 376-381.https://doi.org/10.1016/j.sbspro.2011.03.105

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. https://doi.org/10.1037/a0026838

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Saenkhum, T., & Kim, S. H. (2023). Generative artificial intelligence and second language writing. *Journal of Second Language Writing, 62*, Article 101066. https://doi.org/10.1016/j.jslw.2023.101066

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education, 57*(4), 620–631. https://doi.org/10.1111/ejed.12532

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Teng, L. S. (2022). Developmental trajectories of SRL: Evidence from a case study. In L. S. Teng (Ed.), *Self-regulated learning and second language writing: Fostering strategic language learners* (pp. 183–207). Springer.

Tsao, J. (2021). Effects of EFL learners' L2 writing self-efficacy on engagement with written corrective feedback. *Asia-Pacific Education Researcher, 30*, 575–584. https://doi.org/10.1007/s40299-021-00591-9

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

Wang, P., & Lee, C. (2022). Examining the effectiveness of artificial intelligence-powered writing assistants in enhancing EFL students' writing performance. *Computer Assisted Language Learning, 35*(8), 1634–1658. https://doi.org/10.1080/09588221.2021.1877525

Zhu, J., Yang, Y., & Yan, Z. (2024). Relationships between teacher feedback and English writing proficiency in Chinese students: The mediating effect of writing self-regulated learning strategies. *System, 123*, Article 103338. https://doi.org/10.1016/j.system.2023.103338

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*(2), 64–70.