

RESEARCH

Open Access



The role of ChatGPT in enhancing EFL students' ESP writing skills: an experimental study of gender and major differences

Hassen A. Jmaiel¹, Reema O. Abukhait¹, Amr M. Mohamed^{2*}, Tahany S. Shaaban¹, Mohammad Hamad Al-khresheh³ and Abdo Hasan AL-Qadri⁴

*Correspondence:

Amr M. Mohamed
amamr83@yahoo.com

¹The Applied College, Northern
Border University, Arar, Saudi Arabia

²North Private College of Nursing,
Arar, Saudi Arabia

³Department of Languages and
Translation, College of Humanities
and Social Sciences, Northern
Border University, Arar, Saudi Arabia

⁴School of Humanities and
Education, Xi'an Eurasia University,
Xi'an, China

Abstract

This experimental study investigates the impact of ChatGPT as a self-directed learning tool on the enhancement of English for Specific Purposes (ESP) writing skills, with a focus on gender and academic major differences involving 117 university students learning ESP writing. Data from pre- and post-tests, analysed using the Wilcoxon signed-rank test and ANOVA, indicate substantial improvements in writing skills across all measured dimensions. Results indicate notable gender differences, with female students demonstrating greater improvements, and underscore the limited impact of academic major on writing skill enhancement. While the study findings suggest a positive impact of ChatGPT on enhancing ESP writing skills, future research should involve more diverse and larger samples and include comparative analyses with other learning tools.

Keywords AI in education, Gender analysis, Major analysis, ESP, Self-Directed Language learning, Educational technology

1 Introduction

The application of AI to language acquisition is a growing research area. ChatGPT has garnered attention among AI technologies for its potential application in EFL writing tasks. The incorporation of ChatGPT in educational settings revolutionises language learning methodologies [27, 28, 44].

Traditional EFL instruction emphasises reading, writing, listening, and speaking competencies. The intricacies of syntax, semantics, and creative expression make writing arduous [30, 42]. Language educators can now augment instruction using AI-powered technologies like ChatGPT. ChatGPT's sophisticated language processing aids students with idea generation, phrase structuring, and grammatical correction [28]. The writing settings for students may be more supportive and empowering. The integration of AI into education poses obstacles.



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Literature evaluated the legitimacy of AI-generated remarks, enquiring whether they could genuinely express human feeling and cultural specificity [40]. Research indicated that schools with inadequate internet infrastructure face challenges in integrating AI, implying that technological disparity may exacerbate educational inequality [10]. Moreover, concerns over excessive technology utilisation [43], the legitimacy of AI-generated content, and the critical thinking and inventiveness of students are prevalent. The efficacy of these strategies in various educational and cultural contexts remains ambiguous [27, 32].

Language education has emphasised self-directed learning as essential to language learning [29]. It involves independent decision-making and language development responsibility. Students are encouraged to actively study languages, challenging the teacher-centered paradigm. This requires goal setting, material selection, and progress monitoring. A learner-centred approach that considers students' needs, interests, and learning styles creates personalised and engaging language experiences. Increased autonomy boosts learner motivation and engagement. Thus, scholars and educators have stressed the need of creating a language-learning environment that fosters student autonomy [6, 18].

The primary objective of this experimental study is to evaluate the efficacy of ChatGPT as a self-directed learning tool in enhancing ESP writing skills among faculty students. The study intends to:

1. Examine potential differences in the augmentation of writing skills by gender.
2. Examine the impact of academic major on the degree of writing skill improvement, investigating whether specific degrees derive greater advantages from the incorporation of ChatGPT.
3. Enhance comprehension of the effective application of ChatGPT in educational settings to foster self-directed learning in language learning.

By addressing these objectives, the study seeks to provide insights into the role of AI in modern language education and its implications for fostering self-directed learning.

2 Literature review

Proficient writing in a foreign language is a crucial skill. It facilitates academic and professional achievement. Nonetheless, non-native speakers may encounter considerable difficulties in writing. This is equally applicable to English and its various contexts, including ESP. In ESP courses, students must compose professional emails and reports, necessitating vocabulary, norms, and writing styles. Inadequate skill acquisition results in substandard writing. Consequently, students ought to cultivate writing skills and methods [2, 30].

Writing is conventionally regarded as an "isolated" endeavor, wherein students engage in thinking, drafting, and revising independently [5]. To attain proficiency in writing, a student must invest thousands of hours in solitary practice. The instruction of writing is labor-intensive and challenging, resulting in minimal practice opportunities for students in educational settings. In 2022, ChatGPT became broadly accessible, prompting some educators to voice apprehensions over pupils potentially utilizing it for academic misconduct, thereby circumventing the arduous task of honing writing skills. Media coverage of these issues likely intensified their apprehensions [5, 21].

2.1 Self-directed learning

A framework known as self-directed learning (SDL) empower the students to independently oversee their own educational journeys. Since the 1970s, adult education research has placed a great deal of emphasis on this [17]. Garrison [11] asserts that “self-directed learning is an essential aspect of adult education that empowers learners to take control of their educational journeys” (p. 18). The three fundamental stages of the SDL paradigm—performance, self-reflection, and foresight—were first described by Zimmerman [46]. The first phase, foresight, is the most fundamental and vital of learning. Students develop goals and make plans after assessing the learning challenge. In order to achieve the learning goal, learners must often adjust their motivation. Performance is the second level, when students actively learn. To meet learning goals, students must actively participate in school and self-regulate. The final level of learning is self-reflection. To improve their educational experience and knowledge, students must evaluate and analyze their own efficacy and learning outcomes [19, 47].

Several studies have investigated SDL to comprehend the learning processes of students. Tough [36] initially claimed that SDL originated from adult education. SDL can be interpreted in two ways: as the personal traits of learners and as a learning process. Researchers advocating for the personal qualities interpretation have listed independence, metacognitive awareness, self-efficacy, deep learning, intrinsic motivation, and prioritisation in learning as prerequisites for self-directed learning [26, 45].

According to Garrison [11], SDL encourages students to take individual and group responsibility for cognitive (self-monitoring) and environmental (self-management) processes in creating and verifying meaningful and useful learning results. Additionally, Garrison stated that task motivation keeps learners engaged and persistent, while entrance motivation draws them to a learning environment or activity. Cognitive and metacognitive thinking is essential for regulating and reflecting on learning and educational processes. Self-management directs learning through the regulation of resources, time, and environment. These three factors are interconnected [45].

Increasing study investigates the impact of technology-mediated environments and digital learning instruments on SDL. Research indicates that modern technologies enhance self-directed learning outside the classroom. Duolingo and Busuu are versatile and widely utilised platforms for self-directed language learning [37].

Empirical research on generative AI in service design and delivery is scarce. Generative AI systems may accommodate various learning styles, demands, methodologies, and objectives [32]; nonetheless, Ali et al. [1] emphasise the importance of cognitive engagement and self-assessment to enable learners to evaluate and reflect on their learning processes and outcomes. The findings suggest that further research is necessary to comprehend how SDL learners employ ChatGPT and the impact of generative AI tools on them.

Lin [22] examines ChatGPT's ability to set learning goals, provide instructional materials, provide personalized feedback, and provide interactive assistance for independent adult learners in asynchronous online settings. Nevertheless, inadequate resources and excessive reliance on AI may hinder critical thinking. The qualitative research conducted by Mogavi et al. [25] examines 1500 significant social media posts pertaining to this subject. Early users of educational ChatGPT value its capacity to deliver individualised feedback and assist in establishing learning objectives and lesson plans; nonetheless, they

express concerns over superficial learning, academic dishonesty, and potential detriment to critical thinking skills. Germinal AI may assist SDL while also providing learners with cognitive shortcuts.

2.2 Gender and majors differences

Research has demonstrated that enhancements in learning methodologies inside generative AI-augmented learning environments foster increases in students' self-efficacy, technological acceptance, and learning motivation (Mohamed et al. [28]; Wu et al. [39]. Recently, Elshaer et al. [9] conducted a survey of university students from various academic disciplines and genders. Gender significantly influenced the association between performance expectancy and ChatGPT utilisation. Male students saw a greater impact from performance anticipation than female students when utilising ChatGPT. Males experienced a greater social influence than females while utilising ChatGPT. The research indicated that study discipline greatly influences social impact and the utilisation of ChatGPT. Social influence affects ChatGPT utilisation more significantly in social sciences than in practical sciences.

Alotabibi and Alshehri (2023) contended that the adoption and use of ChatGPT differ across higher education disciplines. According to them, ChatGPT may be advantageous for data analysis and problem-solving in the fields of computer science and engineering. They noted that these discrepancies underscore the necessity of accounting for disciplinary contexts while employing ChatGPT in educational settings. Similarly, Jabeur et al. [15] suggest that while business students may use it for decision-making, students in the social sciences and arts may utilize it for language and text analysis. Multidisciplinary classes can highlight ChatGPT's many uses, especially its ability to meet integrated domain needs. Additionally, Strzelecki and ElArabawy [34] examined how gender and education affect Polish and Egyptian students' use of generative AI platforms like ChatGPT. Primary findings in both nations showed that gender and field of study affected students' acceptance (effort anticipation, social influence, and favorable conditions) of behavioural intention and AI use in education.

Conversely, Lee et al. [17] discovered that there were no gender or age disparities in computer usage for self-directed learning; nevertheless, older students exhibited higher scores in both learning motivation and anxiety. To enhance university students' utilisation of technology for ESL learning, it is crucial to foster their motivation for learning, particularly among younger students [17].

2.3 Main research question

What is the overall impact of using ChatGPT as a self-directed learning tool on ESP writing skill improvement, and are there any statistically significant differences in this improvement due to gender and academic majors? To answer this research question, the following hypotheses derived from the literature are proposed:

Hypothesis 1 Students demonstrate a statistically significant improvement in ESP writing skills following the utilization of the ChatGPT as a self-directed learning tool.

Hypothesis 2 A statistically significant difference exists in ESP writing skill improvement between male and female students who employ the ChatGPT.

Hypotheses 3 There is a statistically significant difference in ESP writing skill improvement among students from different academic majors using the ChatGPT.

3 Methodology

3.1 Research design

The study employed a quantitative research methodology, specifically a pre-experimental design that hinged upon one-group pretest-posttest approach to evaluate the impact of ChatGPT as a self-directed learning tool on ESP writing skills. This study adopted a one-group pretest-posttest design to assess ChatGPT's impact on EFL students' ESP writing quality. While the absence of a control group limits causal interpretation, this approach aligns with established methodological precedents in longitudinal writing research (e.g., Man and Chau [24], and addresses pragmatic constraints in natural educational settings. The development in ESP writing skills was the dependent variable, whereas gender and academic major were the independent variables. By manipulating independent variables and analyzing their impacts on dependent variables, quantitative research methodology generally entails testing hypotheses and addressing research problems [12, 31]. This approach efficiently ascertains the efficacy of the intervention on participants' ESP writing abilities by methodically gathering quantitative data and doing statistical analysis. Pre-experimental designs are frequently employed in education research to evaluate how new teaching methods affect students' learning results (Bin-Hady et al. [4]; Gurjar [13, 14]). Small sample sizes and non-random participant selection are employed in this design [35]. Therefore, the methodology used in this study is appropriate.

Noteworthy, a within-subjects design or repeated measures was implemented in this study. Unlike the between-subject design, which compares different treatment groups, the within-subjects design compares variations in the outcome variable across different conditions within the same group over time and evaluates the intervention's effectiveness [16]. This design is particularly beneficial when researchers can only access one group and need to observe it over a prolonged period [4]; Shaaban and Mohamed [33]. It is believed that this design enhances reliability and fulfils internal validity criteria [3].

3.2 Participants

This study targeted Saudi undergraduate students studying ESP courses. A cohort of 117 students, aged from 18 to 21, was selected using both convenience and purposive sampling techniques. These non-probability sampling techniques were employed due to the researchers' accessibility to and familiarity with the group, which possesses specific characteristics of critical importance to this study. Non-probability samples can generally be acquired more swiftly and at a reduced financial cost compared to probability sampling methods [23]. The sample group consisted of 60 males and 57 females who were informed of the study's purpose and details and willingly consented to participate. The Common European Framework of Reference placed their English competence at A1–B1. They were majoring in various subjects including E-commerce, Legal Assistant, and Information Technology. All study participants gave informed consent to participate and publish, understanding the study's goal, procedures, and the rights to withdraw without penalty. Before participating, students were notified that their involvement was completely optional and that they could withdraw from the study at any moment without facing any adverse repercussions.

Table 1 Pre-assessment test specification

Question	Skill Assessed	Proficiency Level	Maximum Score
1	Requesting a meeting	A1	20
2	Writing a thank-you email	A2	20
3	Complaining and requesting	B1	20

Table 2 Test rubrics

Criteria	Points	Description
Content	8	Provides a clear reason for the meeting and suggests suitable meeting times.
Format and Structure	6	Follows proper email format, including appropriate salutations, subject lines, and closing remarks.
Grammar and Punctuation	4	Demonstrates basic grammatical accuracy and punctuation usage.
Clarity and Cohesion	2	Expresses ideas clearly and coherently.
Total	20	

3.3 Study instrument

The study utilized formal email writing assignments as pre- and post-tests to gauge students' formal email writing abilities. Experts in ESP teaching reviewed the pre- and post-tests items to ensure their face and content validity. Consequently, students were given specific prompts and tasked with composing professional emails. Tasks were assigned according to students' competency level (see Table 1). Detailed instructions came with each task, including goal, target audience, and format. By completing these assignments, students were able to assess their email writing skills.

To assess students' writing performance, a rubric was created. Table 2 outlines the rubric's evaluation criteria, which include overall effectiveness, organization, clarity, language, and vocabulary. To ensure a consistent and objective assessment of students' writing skills, a specific rating scale was assigned to each criterion.

3.4 Procedures

For a two-month period, the sample group had access to ChatGPT and was introduced to the self-directed learning process. Students were taught how to use ChatGPT efficiently during this time, especially how to create prompts to review their drafts and get comments. Along with receiving a range of tools, participants also received thorough instruction in professional email writing practices. They were also urged to use peer review as a tool for tracking and assessing their own development. Teachers helped students along the way, gave them extra materials, and assessed their completed email writing assignments. Notably, ChatGPT was utilized to promote independent language learning and improve formal email writing abilities through self-directed learning.

3.5 Data collection

Data was initially gathered through a pre-test designed to gauge students' proficiency in writing formal emails prior to the intervention. Students composed formal emails based on provided scenarios. Evaluation criteria encompassed four categories including, Content, Format and Structure, Grammar and Punctuation, and Clarity and Cohesion. Eventually, a post-test, also in the form of email writing tasks, was administered to evaluate students' performance and measure progress following the intervention. It is essential

to emphasize that ethical standards were strictly maintained, primarily by securing informed consent from participants and ensuring the confidentiality of their data.

3.6 Data analysis

In this study, data analysis was conducted using various analytical tools to investigate the role of ChatGPT in enhancing self-directed ESP writing skills. As with all quantitative research, the initial step in statistical analysis after data collection involves describing the characteristics of the sample responses, including distribution, central tendency, and variability. The ensuing step in the data analysis process is inferential statistics, which aims to confirm or refute the study hypotheses. Levene's Test for Equality of Variances, the Wilcoxon signed-rank test, and Analysis of Variance (ANOVA) were all included in this phase. Where Levene's test indicated heterogeneity of variance ($p < .05$; see Tables 8 and 12), Welch's ANOVA was employed for the Vocabulary and Cohesion subskill to ensure statistical rigor, as it does not assume equal variances across groups. For all other variables, standard ANOVA procedures were retained based on adequate sample sizes and a balanced group design, which support the robustness of ANOVA to moderate violations of the homogeneity of variance assumption [7]. No within-subjects analyses were conducted; therefore, sphericity corrections were not applicable. These tests facilitated the comparison of pre-and post-test scores and identified any statistical differences in writing proficiency between male and female participants from different academic majors after using ChatGPT as a self-directed learning tool.

3.7 Ethical considerations

This study was approved by Northern Border University's Permanent Committee for Research Ethics in Saudi Arabia. The study adhered to the guidelines set forth in the Declaration of Helsinki. All relevant study information, such as the title, research goal, participant rights, withdrawal terms, and informed consent form, are included in the first section of the survey link. Informed consent was indicated by the participants' selection of the "I consent to participate" option. Throughout the implementation of this study, researchers strictly adhered to ethical criteria to ensure and maintain the credibility of the research. One of the stringent measures implemented to guarantee the participants' privacy and confidentiality was data anonymization. Additionally, the study's design placed a high priority on voluntary participation and provided the freedom to withdraw from the study at any time without facing repercussions, thereby prioritizing the rights and welfare of participants.

4 Findings

4.1 Pre-post changes in ESP writing skills with ChatGPT

The descriptive statistics presented in Table 3 provide substantial evidence supporting Hypothesis 1, which posits that students exhibit a statistically significant improvement in ESP writing skills after utilizing the ChatGPT as a self-directed learning tool. Across all measured variables, the post-test means consistently exceeded the pre-test means, indicating notable progress in students' writing abilities.

- Content: The mean score increased from 3.949 in the pre-test to 5.145 in the post-test, reflecting a significant enhancement in students' ability to produce relevant, well-developed content.

Table 3 Descriptive statistics of Pre- and Post-Test writing quality scores for the experimental group

Writing skills	N	Post-Test			Pre-Test		
		Mean	SD	SE	Mean	SD	SE
Content	117	5.145	1.733	0.160	3.949	1.613	0.149
Format and Structure	117	4.051	1.151	0.106	3.256	1.190	0.110
Grammar and Punctuation	117	2.581	0.843	0.078	2.145	0.912	0.084
Vocabulary and Cohesion	117	1.581	0.495	0.046	1.274	0.502	0.046
Total	117	13.359	3.818	0.353	10.667	3.853	0.356

Table 4 Comparison of Pre- and posttest writing skills in the experimental group using Wilcoxon Signed-Rank tests

Writing skills	W	z	P	Rank-Biserial Correlation	SE Rank-Biserial Correlation	95% CI for Rank-Biserial Correlation
Content	3704.5	7.578	<0.001	0.935	0.123	0.897–0.960
Format and Structure	3041.5	6.818	<0.001	0.877	0.128	0.805–0.924
Grammar and Punctuation	1153.0	5.795	<0.001	0.961	0.164	0.926–0.979
Vocabulary and Cohesion	561.0	5.012	<0.001	1.000	0.197	1.000–1.000
Total	5972.0	8.707	<0.001	0.956	0.109	0.934–0.971

Wilcoxon signed-rank test

- **Format and Structure:** The increase from a pre-test mean of 3.256 to a post-test mean of 4.051 indicates that the students improved in structuring and organizing their writing more effectively.
- **Grammar and Punctuation:** A smaller but still significant increase from 2.145 to 2.581 highlights students' progress in grammatical accuracy and punctuation use.
- **Vocabulary and Cohesion:** Although the difference was less pronounced (1.274 to 1.581), it still indicated an improvement in students' use of vocabulary and their ability to maintain cohesion in their writing.
- **Total Score:** The overall improvement was evident, with the total score increasing from 10.667 to 13.359, indicating significant advancement across all measured aspects of writing.

The data support this hypothesis, demonstrating that students improved their ESP writing skills after using the ChatGPT for self-directed learning. The consistent improvement across all writing components further substantiates the effectiveness of ChatGPT as a learning tool.

Table 4 presents the results of the Wilcoxon signed-rank test comparing pre-and post-test scores in the experimental group, revealing statistically significant improvements across all variables after utilizing ChatGPT as a self-directed learning tool. The p-values for all variables were <0.001, indicating highly significant changes in students' writing skills.

- **Content:** With a W value of 3704.5 and a z-score of 7.578, the results demonstrate a statistically significant improvement in content development. The rank-biserial correlation of 0.935 (95% CI: 0.897–0.960) reflects a strong effect size, suggesting that the students made substantial progress in producing well-developed and relevant content.
- **Format and Structure:** The W value of 3041.5 and a z-score of 6.818 indicate a significant improvement in the organization and structure of students' writing. A

rank-biserial correlation of 0.877 (95% CI: 0.805–0.924) underscores the magnitude of this enhancement.

- Grammar and Punctuation: A *W* value of 1153.0 and a *z*-score of 5.795 highlight significant gains in grammatical accuracy and punctuation. The rank-biserial correlation of 0.961 (95% CI: 0.926–0.979) reflects a very strong effect, further confirming students' improvement in this area.
- Vocabulary and Cohesion: Although the numerical increase is smaller, the *W* value of 561.0 and *z*-score of 5.012 reflect a significant improvement in vocabulary and cohesion. The rank-biserial correlation of 1.000 (95% CI: 1.000–1.000) indicated a perfect correlation, demonstrating that all students improved in this category. The rank-biserial correlation of 1.000 (95% CI: 1.000–1.000) is unusual and may reflect stringent scoring criteria or minimal baseline variation.
- Total Score: The total score reflects an overall significant improvement, with a *W* value of 5972.0 and *z*-score of 8.707. The rank-biserial correlation of 0.956 (95% CI: 0.934–0.971) further confirmed the substantial overall enhancement of students' ESP writing skills.

These results support the hypothesis that students exhibit significant gains in writing proficiency after using ChatGPT for self-directed learning. The consistently high rank-biserial correlations across all the variables further highlight the magnitude and consistency of these improvements. The consistently high rank-biserial correlations across all variables highlight the magnitude of improvement, though the perfect correlation ($r = 1.000$) in Vocabulary/Cohesion warrants further investigation to rule out rubric artifacts or ceiling effects.

4.2 Gender variations in ChatGPT-Assisted writing improvement

The results presented in Table 5 provide substantial evidence for Hypothesis 2, which postulates a statistically significant difference in ESP writing skill improvement between male and female students utilizing ChatGPT as a self-directed learning tool.

- Pre-post-test: The analysis revealed a highly significant *F*-value of 1244.235 ($p < .001$) with a large effect size ($\eta^2 = 0.789$), indicating a substantial overall improvement in ESP writing skills across both genders from pre-test to post-test. This pronounced effect underscores the significant impact of the ChatGPT on writing skill development for all students, irrespective of gender.
- Pre-Post Gender Interaction: The interaction between pre-post-test results and gender is also statistically significant, with an *F*-value of 47.924 ($p < .001$) and an associated eta-squared ($\eta^2 = 0.030$). While the interaction effect confirms that gender influences the degree of improvement, the relatively small effect size suggests that the difference between male and female students, although significant, is modest compared to the overall gains observed from the intervention.

Table 5 Within-Subjects effects for Pre-Post test and gender interaction

Source	SS	Df	MS	F	<i>p</i>	η^2
Pre-Post Test	17,266.905	9	1,918.545	1,244.235	< 0.001	0.789
Pre-Post * Gender Interaction	665.066	9	73.896	47.924	< 0.001	0.030
Residuals	1,595.915	1,035	1.542			

Type III sum of squares used. η^2 = eta-squared (effect size). SS = sum of squares; MS = mean square

Table 6 Between-Subjects effects of gender on ESP writing skills improvement

Source	SS	Df	MS	F	p	η^2
Gender	970.108	1	970.108	80.446	< 0.001	0.044
Residuals	1386.792	115	12.059			

Type III sum of squares used. η^2 = eta-squared (effect size). SS Sum of squares, MS mean square. Welch's ANOVA results reported where variance homogeneity assumption was tested and found violated

Table 7 Descriptive statistics of writing skills scores by gender and timepoint

Writing Skills	Gender	N	Pre-Test		Post-Test	
			Mean (SD)	SE	Mean (SD)	SE
Content	Female	57	4.667 (1.725)	0.229	6.404 (1.334)	0.177
	Male	60	3.267 (1.148)	0.148	3.950 (1.111)	0.143
Format and Structure	Female	57	3.632 (1.248)	0.165	4.579 (1.068)	0.141
	Male	60	2.900 (1.020)	0.132	3.550 (0.999)	0.129
Grammar and Punctuation	Female	57	2.684 (0.848)	0.112	3.193 (0.611)	0.081
	Male	60	1.633 (0.637)	0.082	2.000 (0.582)	0.075
Vocabulary and Cohesion	Female	57	1.544 (0.600)	0.079	1.930 (0.258)	0.034
	Male	60	1.017 (0.129)	0.017	1.250 (0.437)	0.056
Total	Female	57	12.614 (4.078)	0.540	16.105 (2.807)	0.372
	Male	60	8.817 (2.501)	0.323	10.750 (2.640)	0.341

- Residuals: The residuals represent the portion of variance not explained by the model that is relatively low, emphasizing the robustness of the findings.

The data support Hypothesis 2 by confirming a statistically significant gender difference in ESP writing skill improvement following the use of ChatGPT. However, while this difference was significant, the overall effect of ChatGPT on writing skills was considerably more pronounced than gender-based variation. Both male and female students demonstrated substantial gains, with gender playing a secondary, albeit statistically significant, role in the extent of their improvement.

Table 6 presents the Between-Subjects Effects, examining the specific role of gender in ESP writing skill improvement following ChatGPT utilization. The results indicate that gender has a statistically significant effect, as evidenced by an F-value of 80.446 ($p < .001$). However, the effect size is relatively small, with eta-squared ($\eta^2 = 0.044$) indicating that gender explains only 4.4% of the variance in writing improvement, highlighting that most of the gains were consistent across genders. This finding aligns with the Within-Subjects Effects presented in Table 5, where the interaction between the pre- and post-test results and gender also reached statistical significance ($\eta^2 = 0.030$), albeit with a modest impact. Collectively, these results indicate that, while gender differences are statistically significant, they contribute minimally to the overall improvement in ESP writing skills. The primary factor contributing to improvement, as highlighted by both tables, is the utilization of the ChatGPT, which results in substantial gains in ESP writing skills for all students, irrespective of gender. Gender differences, although present, represent only a minor portion of the overall effect, reinforcing the conclusion that the ChatGPT significantly enhances writing performance across the sample population, with gender playing a secondary role.

The descriptive statistics presented in Table 7 show gender differences in ESP writing skill improvement, building upon the findings of Within-Subjects Effects (Table 5) and Between-Subjects Effects (Table 6). The data revealed consistent patterns across various

writing skills, reinforcing the hypothesis that gender influences the observed improvements following the utilization of ChatGPT.

- **Content:** Female students demonstrated higher pre-test ($M = 4.667$) and post-test ($M = 6.404$) mean scores than their male counterparts (pre-test $M = 3.267$; post-test $M = 3.950$), indicating both superior initial performance and greater overall improvement. This aligns with the significant gender effect noted in Table 6, in which gender accounted for 4.4% of the variance ($\eta^2 = 0.044$).
- **Format and Structure:** Female students improved from a pre-test mean of 3.632 to 4.579, while male students increased from 2.900 to 3.550. This difference corresponds with the interaction effect identified in Table 5 ($\eta^2 = 0.030$), highlighting that, although both genders improved, females exhibited greater gains in organizing and structuring their writing.
- **Grammar and Punctuation:** The pre-test to post-test improvement for females (pre-test $M = 2.684$; post-test $M = 3.193$) was notably more substantial than for males (pre-test $M = 1.633$; post-test $M = 2.000$). These findings reinforce the patterns established in the interaction analysis in Table 5, emphasizing the modest yet significant role of gender in this skill area.
- **Vocabulary and Cohesion:** Female students again demonstrated larger improvements (pre-test $M = 1.544$; post-test $M = 1.930$) than males (pre-test $M = 1.017$; post-test $M = 1.250$). Although the overall difference is smaller relative to the other dimensions, it supports the trend of enhanced performance among females across all skills.
- **Total Score:** Female students exhibited a more pronounced increase in total scores (pre-test $M = 12.614$; post-test $M = 16.105$) than males (pre-test $M = 8.817$; post-test $M = 10.750$). This finding is consistent with the overall results in Tables 5 and 6, which revealed significant but modest gender differences in total improvement.

Table 7 confirms that, while both male and female students demonstrated significant improvements in their ESP writing skills after utilizing the ChatGPT, female students experienced larger gains across all assessed dimensions. These results align with the previously identified gender effect, suggesting that gender is a statistically significant factor in writing skill development, albeit a secondary factor. The descriptive statistics provided compelling evidence supporting the hypothesis of gender differences in ESP writing skill improvement.

The results presented in Table 8 demonstrate statistically significant variance differences ($p < .05$) for content (pre-test), Vocabulary and Cohesion (Pre- and Post-tests), and total (pre-test), indicating a violation of the homogeneity of the variance assumption for

Table 8 Levene's test for equality of variances across Pre- and Post-test writing skills by gender

Variable	df1	df2	Pre-Test		Post-Test	
			F	p-value	F	p-value
Content	1	115	9.188	0.003	1.433	0.234
Format and Structure	1	115	2.438	0.121	0.383	0.537
Grammar and Punctuation	1	115	0.837	0.362	2.141	0.146
Vocabulary and Cohesion	1	115	210.288	< 0.001	36.135	< 0.001
Total	1	115	10.370	0.002	0.073	0.788

Levene's test indicated significant violations of variance homogeneity for Vocabulary and Cohesion in both pre- and post-test scores ($p < .001$). Therefore, Welch's ANOVA was used and confirmed statistically significant gender differences (Pre-test: $t(91) = 6.49$, $p < .001$; Post-test: $t(93) = 10.31$, $p < .001$). These results support using Welch's adjustment under variance inequality

Table 9 Within-Subjects effects of time and major on writing skills

Source	SS	Df	MS	F	p-value	η^2
Pre-Post Test	12,707.095	9	1,411.899	917.920	< 0.001	0.733
Pre-Post * Major	682.838	18	37.935	24.663	< 0.001	0.039
Residuals	1,578.143	1,026	1.538			

Type III sum of squares used. η^2 = eta-squared (effect size). SS Sum of squares, MS mean square. Time pre-post test comparison

Table 10 Between-Subjects effects of academic major on writing skills improvement

Source	SS	df	MS	F	p-value	η^2
Major	978.477	2	489.239	40.462	< 0.001	0.056
Residuals	1,378.422	114	12.091			

Type III sum of squares used. η^2 = eta-squared (effect size). SS = sum of squares; MS = mean square

these variables. Conversely, for other writing skills, namely Content (post-test), Format and Structure, and Grammar and Punctuation, the p-values exceeded 0.05, confirming that the assumption of equal variances was met, thus necessitating no adjustments for these variables in subsequent analyses.

4.3 Differences in ChatGPT-enhanced writing by academic major

The results in Table 9 provide critical insights into Hypothesis 3, which posits a statistically significant difference in ESP writing skill improvement among students from different academic majors who utilize the ChatGPT.

- Pre-post-test: The analysis revealed a highly significant F-value of 917.920 ($p < .001$) with a large effect size ($\eta^2 = 0.733$). This indicates a substantial overall improvement in ESP writing skills from pre-to post-test across all students, irrespective of their academic major. The significant pre-post effect underscores the considerable impact of ChatGPT as a learning tool for enhancing writing skills.
- Pre-Post Major Interaction: The interaction effect between pre-post-test results and students' academic majors is also statistically significant, with an F-value of 24.663 ($p < .001$) and a smaller effect size ($\eta^2 = 0.039$). While this confirms that academic major influences writing skill improvement, the small effect size suggests that these differences are relatively minor compared to the overall improvement. Although students from different majors exhibit some variation in their improvement, the primary factor driving the observed gains.
- Residuals: Low residuals further underscore the robustness of the model in explaining the variance in writing skill improvement, with minimal unexplained variation.

The findings support Hypothesis 3, confirming that academic major has a statistically significant effect on ESP writing skill improvement. However, the modest effect size ($\eta^2 = 0.039$) indicates that, while there are differences among majors, they are relatively small. The primary driver of improvement is the overall impact of the ChatGPT, which significantly enhances writing skills across all academic disciplines.

Table 10 indicates that academic major exerts a statistically significant effect on ESP writing skill improvement, as evidenced by an F-value of 40.462 ($p < .001$) and an effect size of $\eta^2 = 0.056$. However, despite statistical significance, this effect size indicates that academic major explains only a small portion (5.6%) of the variance in writing improvement, suggesting that the major's impact is modest compared to the substantial overall

gains observed. The low residual variance underscores the model's efficacy in capturing the key factors driving improvement. These findings demonstrate that while academic majors play a role, their impact is limited, with the primary driver of ESP writing improvement being the intervention (ChatGPT), which confers benefits to students across all majors.

The descriptive statistics in Table 11 provide detailed insights into ESP writing skill improvement across three academic majors: e-commerce, information technology, and legal assistance. These findings align with the results presented in Tables 9 and 10, supporting Hypothesis 3, which posits significant differences in writing skill improvement among students from different majors using the ChatGPT.

- **Content:** Information Technology students exhibited the greatest improvement, with their mean score increasing from 4.667 (pre-test) to 6.404 (post-test), while students from E-commerce and Legal Assistant majors demonstrated more moderate gains. This finding aligns with the Between-Subjects Effects in Table 10, where academic major was found to have a statistically significant effect on writing improvement ($F = 40.462$, $p < .001$), although the effect size was modest ($\eta^2 = 0.056$). This suggests that, while major plays a role, it does not account for a large proportion of the variance.

Table 11 Descriptive statistics for Pre- and Post-test writing skills by major

Writing Skills	Time	Major	N	Mean	SD	SE
Content	Pre-test	E-commerce	21	2.810	0.873	0.190
		Information Technology	57	4.667	1.725	0.229
		Legal Assistant	39	3.513	1.211	0.194
	Post-test	E-commerce	21	3.762	0.944	0.206
		Information Technology	57	6.404	1.334	0.177
		Legal Assistant	39	4.051	1.191	0.191
Format and Structure	Pre-test	E-commerce	21	2.762	0.944	0.206
		Information Technology	57	3.632	1.248	0.165
		Legal Assistant	39	2.974	1.063	0.170
	Post-test	E-commerce	21	3.619	0.805	0.176
		Information Technology	57	4.579	1.068	0.141
		Legal Assistant	39	3.513	1.097	0.176
Grammar and Punctuation	Pre-test	E-commerce	21	1.524	0.680	0.148
		Information Technology	57	2.684	0.848	0.112
		Legal Assistant	39	1.692	0.614	0.098
	Post-test	E-commerce	21	2.000	0.548	0.120
		Information Technology	57	3.193	0.611	0.081
		Legal Assistant	39	2.000	0.607	0.097
Vocabulary and Cohesion	Pre-test	E-commerce	21	1.000	0.000	0.000
		Information Technology	57	1.544	0.600	0.079
		Legal Assistant	39	1.026	0.160	0.026
	Post-test	E-commerce	21	1.286	0.463	0.101
		Information Technology	57	1.930	0.258	0.034
		Legal Assistant	39	1.231	0.427	0.068
Total	Pre-test	E-commerce	21	8.095	2.119	0.462
		Information Technology	57	12.614	4.078	0.540
		Legal Assistant	39	9.205	2.628	0.421
	Post-test	E-commerce	21	10.667	2.176	0.475
		Information Technology	57	16.105	2.807	0.372
		Legal Assistant	39	10.795	2.885	0.462

Table 12 Levene's test for equality of variances across Pre- and Post-test writing skills by major

Writing Skills	df1	df2	Pre-Test		Post-Test	
			F	p-value	F	p-value
Content	2	114	6.638	0.002	1.216	0.300
Format and Structure	2	114	1.304	0.275	1.913	0.152
Grammar and Punctuation	2	114	0.736	0.481	1.251	0.290
Vocabulary and Cohesion	2	114	105.877	<0.001	18.164	<0.001
Total	2	114	6.416	0.002	1.142	0.323

Levene's test showed variance violations for Vocabulary and Cohesion ($p < .001$); thus, Welch's ANOVA was used. For other skills, standard ANOVA was retained due to equal variances and balanced group sizes

- **Format and Structure:** Information Technology students also outperformed their peers in this dimension, with their mean improving from 3.632 (pre-test) to 4.579 (post-test). E-commerce and legal assistance students showed smaller improvements. This is consistent with the Within-Subjects Effects in Table 9, where the interaction between the pre-and post-test results and academic major was significant ($F = 24.663$, $p < .001$), but the effect size was relatively small ($\eta^2 = 0.039$). This suggests that although differences exist across majors, the overall effect of the intervention (ChatGPT) is more dominant.
- **Grammar and Punctuation:** Information Technology students again performed better, with a post-test mean of 3.193, compared to 2.000 for both E-commerce and Legal Assistant students. This reinforces the findings in Table 9, where academic majors had a statistically significant, though minor, impact on grammar and punctuation improvement.
- **Vocabulary and Cohesion:** In this dimension, improvements were modest across all majors, with Information Technology students having a slightly higher post-test mean (1.930) than their peers in E-commerce and Legal Assistant. This result aligns with the smaller effect sizes observed in Table 9 for the interaction of academic major and pre-post improvement.
- **Total Score:** Information Technology students achieved the most substantial overall improvement, with a post-test mean of 16.105, compared to 10.667 for E-commerce and 10.795 for Legal Assistant students. This confirms the findings in Table 10, where major accounted for a small proportion of the total variance ($\eta^2 = 0.056$), although the differences were still statistically significant, indicating that major influences the total score improvements to a modest extent.

The data in Table 11 corroborate the findings from Tables 9 and 10, demonstrating that academic major has a statistically significant, albeit modest, effect on ESP writing skill improvement. Information Technology students consistently showed the highest gains, while students from E-commerce and Legal Assistant majors experienced smaller improvements. However, as indicated by both Tables 9 and 10, the primary factor driving writing improvement was the intervention (ChatGPT), with academic majors playing a secondary role in shaping the extent of improvement.

Table 12 shows that the assumption of equal variances was violated ($p < .05$) for several key variables, specifically content (pre-test), Vocabulary and Cohesion (pre-test and post-test), and total (pre-test). This finding suggests a significant heterogeneity in the variability between the majors in these domains. Conversely, for other variables, such as content (post-test), Format and Structure, and Grammar and Punctuation, the p-values exceed 0.05, indicating that the assumption of equal variances is met for these measures.

5 Discussion

The recent study investigated the impact of using ChatGPT as a self-directed learning (SDL) tool on ESP writing skill improvement, and whether there are any statistically significant differences in this improvement due to gender and academic majors. The analysis indicated that ChatGPT has the potential to support self-directed learning and may contribute to improvements in ESP writing skills, with variation observed across gender and academic major. Female students, particularly those majoring in language-related fields, seemed to benefit more from using ChatGPT. These findings suggest that AI-powered tools can be tailored to specific learner needs to maximize their impact. The following subsections elaborate on these key findings while maintaining distinct focus areas.

5.1 ChatGPT's impact on ESP writing quality

The findings presented in Table 1 provide compelling evidence for Hypothesis 1, demonstrating a significant improvement in students' English for Specific Purposes writing skills after using ChatGPT as a self-directed learning tool. The substantial increase in mean scores across all writing dimensions content, format and structure, grammar and punctuation, and vocabulary and cohesion indicate that the intervention effectively enhanced students' writing abilities. These results strongly support the hypothesis that ChatGPT significantly improves ESP writing skills, as the tool enhanced all aspects of writing, from content to grammar. The descriptive statistics and results from the Wilcoxon signed-rank test consistently showed notable gains across all writing components. These comprehensive findings collectively demonstrate ChatGPT's multifaceted benefits for ESP writing development. These findings are in line with previous research on AI-powered language learning, underscoring its potential to facilitate language acquisition [41].

5.2 Gender variations in writing improvement

Regarding the second Hypothesis: Gender Differences in ESP Writing Skill Improvement, the ANOVA results showed a significant main effect of gender, indicating that female students improved more than male students after using ChatGPT as a self-directed learning tool. However, the effect size was small ($\eta^2 = 0.044$), suggesting that gender accounts for only about 4% of the variance in improvement, indicating that the majority of gains were similar across genders. This result confirms other studies showing that gender can influence language learning results [8]. Although the intervention was beneficial to both male and female students, it seemed that female students were more open to the tool or may have utilized it more successfully. Although there were gender variances, both sexes showed notable gains in their ESP writing abilities. Notably, both genders demonstrated substantial gains, with the gender difference representing a secondary consideration. This demonstrates how AI-powered resources, such as ChatGPT, can enhance language learning for a range of students. The gender differences observed warrant particular attention in both implementation strategies and future research directions. Recent research on generative AI chatbots, like Teacher GAIA [1], emphasizes the need of self-evaluation and cognitive engagement. These components help students assess and consider their learning objectives and procedures. To investigate the factors behind this gender disparity, more research is needed. Variations in language

learning methods, motivation levels, and prior language experience are a few examples of potential contributing factors. By recognizing these elements, researchers and educators may create plans that improve the efficiency of AI-powered language learning resources for both male and female students.

5.3 Major-based divergences in writing gains

According to the current study, students from different academic majors who used ChatGPT showed statistically significant differences in their improvement of ESP writing skills. However, the effect size was small ($\eta^2 = 0.056$), indicating that academic major explained only 5.6% of the variance in improvement, highlighting that the substantial overall gains were largely consistent across majors despite these minor differences. After utilizing ChatGPT as a self-directed learning SDL tool, students from various majors saw differing degrees of development in their English for Specific Purposes (ESP) writing abilities, according to the ANOVA results, which revealed a significant main impact of academic major. Information Technology (IT) majors showed the biggest improvement, whilst E-commerce (EC) and Legal Assistant (LA) students had less increase. Crucially, all major groups showed meaningful progress, with the between-group differences being relatively minor in practical terms. It's crucial to keep in mind, though, that the ChatGPT intervention had a big overall impact on all majors. This is in line with other studies that demonstrated how ChatGPT and other AI-powered language learning programs may raise language competency regardless of the learners' backgrounds or academic specializations [20]. The major-specific patterns revealed here highlight the importance of disciplinary considerations when implementing AI writing tools. There are a number of reasons why different majors differ from one another, such as disparities in prior language proficiency, different degrees of motivation, or the particular requirements of each major. Additionally, the variations in variance among writing components suggest that ChatGPT's impact may vary depending on the specific skill being worked on. The current study unequivocally shows that the use of ChatGPT as a self-directed learning tool was the main driver of ESP writing skill growth, even though an academic major affects the degree of progress. According to recent research on generative AI chatbots, such as Teacher GAIA, this discovery highlights the potential of AI-powered tools to provide students across a range of academic fields with individualized and successful language learning experiences [1]. Furthermore, Warschauer and Xu's [38] suggestion to incorporate technology into language instruction offers a substantial chance to improve language instruction. This strategy has the ability to revolutionize conventional teaching techniques and enhance students' overall learning results by offering creative and captivating learning opportunities.

5.4 Pedagogical implications

The findings of this study carry significant pedagogical implications for educators, curriculum designers, and policymakers seeking to integrate AI tools like ChatGPT into ESP instruction. First, higher education institutions should proactively revise ESP course syllabi to include structured training on AI-assisted writing, ensuring students learn to use these tools critically rather than dependently. For example, writing assignments could be scaffolded to include both AI-generated drafts and human revisions, fostering metacognitive awareness of strengths and limitations in AI outputs. Second, ESP

instructors should receive targeted professional development on prompt engineering and AI literacy, enabling them to model effective ChatGPT use for discipline-specific writing tasks such as technical reports or case study analyses. Third, assessment rubrics must be adapted to evaluate not only final writing quality but also the process of human-AI collaboration, emphasizing originality and critical engagement with AI suggestions. Finally, given the observed gender and major-based differences, instructors should adopt differentiated support strategies—such as tailored feedback sessions for non-humanities majors or peer mentoring programs for female students—to ensure equitable learning gains. These evidence-based adjustments would maximize ChatGPT's pedagogical value while maintaining academic rigor in EFL contexts.

6 Conclusion

This experimental study sought to examine the effects of ChatGPT as a self-directed learning tool on ESP writing skills among university students. The research findings demonstrate strong evidence of an association between ChatGPT use and improvements across multiple dimensions of students' writing proficiency, including content, organization, grammar, and vocabulary. Notably, the results revealed a gender-based disparity in writing skill acquisition, with female students exhibiting greater gains in writing proficiency compared to their male counterparts. Additionally, the study identified a modest correlation between academic major and the degree of writing skill improvement. These findings collectively highlight the potential influence of both gender and academic major on writing competency enhancement facilitated by AI-assisted learning tools. A critical factor contributing to these enhancements was the tool's capacity to provide personalized feedback, adaptive practice, and immediate support.

While confirming ChatGPT's pedagogical value, the study also stresses the necessity for future studies involving larger and varied samples to further substantiate these results. In essence, this research accentuates the potential of ChatGPT as an invaluable tool in language learning, thereby contributing to the body of evidence upholding the incorporation of AI tools in educational contexts to promote autonomous learning and enhance language skills. The following limitations and future directions refine these conclusions:

While this study offers valuable insights into the potential of AI technologies in ESP instruction, several methodological limitations must be acknowledged. Most significantly, the use of a one-group pretest-posttest design without a control or comparison group is a key limitation, as it restricts our ability to make definitive causal claims about ChatGPT's impact. Consequently, potential confounding variables and maturation effects cannot be ruled out. Additionally, the relatively small sample size, limited to ESP students, may affect the generalizability of the findings, highlighting the need for future research across different academic disciplines, proficiency levels, and cultural backgrounds. Although some violations of the homogeneity of variance assumption were detected, no specific statistical corrections were applied; nevertheless, the strong pre-post effects observed (all $p < .001$) suggest that the overall conclusions remain robust. Despite these constraints, the consistency of significant pre-post effects across writing skills supports the reliability of our core findings. Future studies employing randomized controlled designs with larger, more diverse participant pools are recommended to validate and extend these preliminary conclusions.

Author contributions

All authors made significant contributions to this study. They were involved in conceptualizing the research, developing the methodology, collecting and analyzing the data, as well as writing and revising the manuscript. All authors have reviewed and approved the final published version.

Funding

Not applicable.

Data availability

The data are available from the authors upon reasonable request.

Declarations**Ethics approval**

This study was approved by Northern Border University's Research Ethics Committee (# 24/14/33) and conducted in accordance with the Declaration of Helsinki.

Consent to participate

Informed written consent was obtained from all participants prior to the study. For participants under 18, consent was provided by a parent or legal guardian.

Consent for publication

All participants (or their parent/legal guardian for minors) provided written consent for publication of the study's findings.

Competing interests

The authors declare no competing interests.

Received: 7 April 2025 / Accepted: 10 July 2025

Published online: 12 July 2025

References

1. Ali F, Choy D, Divaharan S, Tay H, Chen H. Supporting self-directed learning and self-assessment using teacher GAIA, a generative AI chatbot application: learning approaches and prompt engineering. *Learning: Res Pract.* 2023;9(2):135–47. <https://doi.org/10.1080/23735082.2023.2258886>.
2. Al-khresh M, Mohamed A, Bin Ali R. A study on writing apprehension among Saudi undergraduate EFL students. *Lang Teach Res Q.* 2023;34:1–18. <https://doi.org/10.32038/ltrq.2023.34.01>.
3. Ary D, Jacobs LC, Sorensen C. Introduction to research in education (8th ed.). Wadsworth; 2010.
4. Cohen L, Manion L, Morrison K. Research methods in education. 6th ed. Routledge; 2007.
5. Cotton DRE, Cotton PA, Shipway JR. Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innovations Educ Teach Int.* 2023;61(2):228–39. <https://doi.org/10.1080/14703297.2023.2190148>.
6. David L, Weinstein N. Using technology to make learning fun: technology use is best made fun and challenging to optimize intrinsic motivation and engagement. *Eur J Psychol Educ.* 2024;39(2):1441–63. <https://doi.org/10.1007/s10212-023-00734-0>.
7. Delacre M, Leys C, Mora YL, Lakens D. Taking parametric assumptions seriously: arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA. *Int Rev Social Psychol.* 2019;32(1):13. <https://doi.org/10.5334/irsp.198>.
8. Ellis R. Second Language acquisition. The United States: Oxford; 1997. p. 98.
9. Elshaer IA, Hasanein AM, Sobaih AEE. The moderating effects of gender and study discipline in the relationship between university students' acceptance and use of ChatGPT. *Eur Journal Invest Health Psychol Education.* 2024;14(7):1981–95. <https://doi.org/10.3390/ejihpe14070132>.
10. Gabriel S, Kühn A, Dumitrescu R. Strategic planning of the collaboration between humans and artificial intelligence in production. *Procedia CIRP.* 2023;120:1309–14. <https://doi.org/10.1016/j.procir.2023.09.168>.
11. Garrison DR. Self-directed learning: toward a comprehensive model. *Adult Educ Q.* 1997;48(1):18–33. <https://doi.org/10.1177/074171369704800103>.
12. Geldhof GJ, Warner DA, Finders JK, Thogmartin AA, Clark A, Longway KA. Revisiting the utility of retrospective pre-post designs: the need for mixed-method pilot data. *Eval Program Plan.* 2018;70:83–9. <https://doi.org/10.1016/j.evalprogplan.2018.05.002>.
13. Gurjar NR. Effect of educational program on knowledge and self-care behavior among arthritis patients: Pre-experimental research design. *Int J Orthop Trauma Nurs.* 2024;52:101038. <https://doi.org/10.1016/j.ijotn.2023.101038>.
14. Hakimah N. Assessing the impact of Project-Based learning on students' writing skills: a Pre-Experimental study. *Acitya J Teach Educ.* 2023;5(2):434–48. <https://doi.org/10.30650/ajte.v5i2.3723>.
15. Jabeur SB, Ballouk H, Mefteh-Wali S, Omri A. Forecasting the macrolevel determinants of entrepreneurial opportunities using artificial intelligence models. *Technol Forecast Soc Change.* 2022;175:121353. <https://doi.org/10.1016/j.techfore.2021.121353>.
16. Kim J. Within-subjects design. In: Salkind NJ, editor. Encyclopedia of research design. SAGE Publications Inc; 2010. pp. 1639–44. <https://doi.org/10.4135/9781412961288.n503>.
17. Lee C, Yeung AS, Ip T. University english Language learners' readiness to use computer technology for self-directed learning. *System.* 2017;67:99–110. <https://doi.org/10.1016/j.system.2017.05.001>.
18. Lee E, Hannafin MJ. A design framework for enhancing engagement in student centered learning: own it, learn it, and share it. *Education Tech Research Dev.* 2016;64:707–34. <https://doi.org/10.1007/s11423-015-9422-5>.

19. Lee HY, Chen PH, Wang WS, Huang YM, Wu TT. Empowering ChatGPT with guidance mechanism in blended learning: effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *Int J Educational Technol High Educ*. 2024;21(1):16. <https://doi.org/10.1186/s41239-024-00447-4>.
20. Lee JY, Hwang Y. A meta-analysis of the effects of using AI chatbot in Korean EFL education. *Stud Engl Lang Literature*. 2022;48(1):213–43.
21. Levine S, Beck SW, Mah C, Phalen L, Pittman J. How do students use ChatGPT as a writing support? *J Adolesc Adult Lit*. 2024;00:1–13. <https://doi.org/10.1002/jaal.1373>.
22. Lin X. Exploring the role of ChatGPT as a facilitator for motivating self-directed learning among adult learners. *Adult Learn*. 2023. <https://doi.org/10.1177/10451595231184928>.
23. Link M. New data strategies: nonprobability sampling, mobile, big data. *Qual Assur Educ*. 2018;26(2):303–14.
24. Man D, Chau MH. Learning to evaluate through that-clauses: evidence from a longitudinal learner corpus. *J Engl Acad Purp*. 2019;37:22–33. <https://doi.org/10.1016/j.jeap.2018.11.007>.
25. Mogavi R, Deng C, Kim J, Zhou J, Kwon P, Hosny Saleh Metwally YD, Tlili A, Bassanelli A, Bucchiarone S, Gujar A, Nacke S, L. E, Hui P. ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers Hum Behavior: Artif Hum*. 2024;2(1):100027. <https://doi.org/10.1016/j.chbah.2023.100027>.
26. Mohamed A, Shaaban T. Investigating college students' metacognitive awareness in enhancing ESP writing proficiency. *J Teach Engl Specif Acad Purposes*. 2023;11(2):387–401. <https://doi.org/10.22190/JTESAP230420029M>.
27. Mohamed A. Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing english as a foreign Language (EFL) teaching: perceptions of EFL faculty members. *Educ Inform Technol*. 2024;1–22. <https://doi.org/10.1007/s10639-023-11917-z>.
28. Mohamed AM, Shaaban TS, Bakry SH, Guillén-Gámez FD, Strzelecki A. Empowering the faculty of education students: applying ai's potential for motivating and enhancing learning. *Innov High Educ*. 2024;1–23. <https://doi.org/10.1007/s10755-024-09747-z>.
29. Morris TH. Adaptivity through self-directed learning to Meet the challenges of our ever-changing world. *Adult Learn*. 2019;30(2):56–66. <https://doi.org/10.1177/1045159518814486>.
30. Nasim SM, Mohamed AM, Mujeeba S, Shaaban TS. Examining gender and major differences in college students' metacognitive awareness of ESP writing skills. *Lang Teach Res Q*. 2024;42:278–94. <https://doi.org/10.32038/ltrq.2024.42.16>.
31. Rizzi F. Education levels as a dependent and independent variable: an analysis of the relationships between social origins, education attained, and employment outputs in the Italian case (15/23 Classification-JEL codes: I20; I21; J24; J62; C42). 2023.
32. Ruiz-Rojas LI, Acosta-Vargas P, De-Moreta-Llovet J, Gonzalez-Rodríguez M. Empowering education with generative artificial intelligence tools: approach with an instructional design matrix. *Sustainability*. 2023;15(15):11524. <https://doi.org/10.3390/su151511524>.
33. Shaaban T, Mohamed A. Exploring the effectiveness of augmented reality technology on reading comprehension skills among early childhood pupils with learning disabilities. *J Computers Educ*. 2024;11(2):423–44. <https://doi.org/10.1007/s40692-023-00269-9>.
34. Strzelecki A, ElArabawy S. Investigation of the moderation effect of gender and study level on the acceptance and use of generative AI by higher education students: comparative evidence from Poland and Egypt. *Br J Educ Technol*. 2024;55:1209–30. <https://doi.org/10.1111/bjet.13425>.
35. Supriyatno T, Susilawati S, Ahdi H. E-learning development in improving students' critical thinking ability. *Cypriot J Educational Sci*. 2020;15(5):1099–106. <https://doi.org/10.18844/cjes.v15i5.5154>.
36. Tough A. The adult's learning projects. Toronto: Ontario Institute for Studies in Education; 1971.
37. Wang C, Li Z, Bonk C. Understanding Self-Directed learning in AI-Assisted writing: A mixed methods study of postsecondary learners. *Computers Education: Artif Intell*. 2024;100247. <https://doi.org/10.1016/j.caeai.2024.100247>.
38. Warschauer M, Xu Y. Generative AI for Language learning: entering a new era. *Lang Learn Technol*. 2024;28(2):1–4. <https://hdl.handle.net/10125/73569>.
39. Wu D, Zhang S, Ma Z, Yue X-G, Dong RK. Unlocking potential: key factors shaping undergraduate Self-Directed learning in AI-Enhanced educational environments. *Systems*. 2024;12(9):332. <https://doi.org/10.3390/systems12090332>.
40. Xiao Y, Zhi Y. An exploratory study of EFL learners' use of ChatGPT for Language learning tasks: experience and perceptions. *Languages*. 2023;8(3):212. <https://doi.org/10.3390/languages8030212>.
41. Yan D. Impact of ChatGPT on learners in a L2 writing practicum: an exploratory investigation. *Educ Inform Technol*. 2023;28(11):13943–67.
42. Yasuda S. Exploring changes in FL writers' meaning-making choices in summary writing: a systemic functional approach. *J Second Lang Writ*. 2015;27:105–21. <https://doi.org/10.1016/j.jslw.2014.09.008>.
43. Yu H. Reflection on whether chat GPT should be banned by academia from the perspective of education and teaching. *Front Psychol*. 2023;14:1181712. <https://doi.org/10.3389/fpsyg.2023.1181712>.
44. Zhang P, Tur G. A systematic review of ChatGPT use in K-12 education. *Eur J Educ*. 2024;59(2):e12599. <https://doi.org/10.1111/ejed.12599>.
45. Zhu M, Doo MY. The relationship among motivation, self-monitoring, self-management, and learning strategies of MOOC learners. *J Comput High Educ*. 2022;34(2):321–42. <https://doi.org/10.1007/s12528-021-09301-2>.
46. Zimmerman BJ. Self-regulated learning and academic achievement: an overview. *Educational Psychol*. 1990;25(1):3–17.
47. Zimmerman BJ. Investigating self-regulation and motivation: historical background, methodological developments, and future prospects. *Am Educ Res J*. 2008;45(1):166–83. <https://doi.org/10.3102/0002831207312909>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.