ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

AI WRITING TOOLS AND STUDENTS' CRITICAL THINKING IN ENGLISH WRITING

Omer Ayub Khan[1], Rimsha Imran[2], Dr. Nadia Majid[3]
[1] Department of English Language Teaching & Linguistics,
IER, University of the Punjab, Lahore, Pakistan
Email: omerwarsi666@gmailcom
[2] Department of English Language Teaching & Linguistics,
IER, University of the Punjab, Lahore, Pakistan
Email: irimshah677@gmail.com
[3] Lecturer, Department of English Language Teaching & Linguistics,
IER, University of the Punjab, Lahore, Pakistan
Email: nadia.majid29@hotmail.com
***Corresponding Author**
Dr. Nadia Majid

Abstract
Nowadays, English-language learners use AI tools like ChatGPT, Grammarly, and QuillBot for their daily writing work, but it remains uncertain whether these tools lead to better or weaker critical thinking. The present study explored the correlation between the use of AI writing tools and critical thinking of 200 undergraduate students from universities in Pakistan. AI writing tool usage was measured using a cross-sectional online survey with five-point Likert scales on the following five items (Cronbach's α = .61) and a critical thinking scale (7 items, α = .75). Students reported use (M = 3.96, SD = 0.43) and critical thinking (M = 3.78, SD = 0.51) significantly above the scale midpoint (both p < .001). The five levels of frequency of AI use were found to significantly differ in terms of critical thinking, F (4, 195) = 6.19, p < .001, η² = .11. The use of AI tools was positively and highly significantly associated with critical thinking (r(198) = .76, p < .001), and was a good predictor of critical thinking in a simple regression model (β = .76, R² = .58). Low-to-moderate uncritical acceptance of AI suggestions was also reported by students (M = 3.07). The results show that the more students use AI writing tools, the more critical they are in their engagement with them, suggesting an assistive (not substitutive) pattern of use. The findings are a convenience sample and self-reported cross-sectional data, which means that they are descriptive, not causal.

Keywords: *AI writing tools; critical thinking; cognitive skills; cognitive offloading; English language learning; higher education; Pakistan.*

1.Introduction

Artificial intelligence (AI) has swiftly entered the realm of English language teaching and learning. Within seconds, ChatGPT and other tools like Grammarly and QuillBot can provide you with ideas, grammar corrections, paraphrasing, and restructuring arguments. ChatGPT and similar tools can suggest ideas, correct grammar, and paraphrase and restructure arguments in mere seconds (Roe et al., 2023). Research into Pakistani undergraduates reveals the use of these tools by students to produce their English writing (Khalid & Ahmed, 2025; Shahzad & Zaman, 2025). While there are many great advantages to accuracy and efficiency, the education question is moving to what students' thinking will look like when they can have a tool do much of the drafting for them. Critical thinking, defined as the ability to analyze, evaluate and create arguments on their own, is one of the key cognitive skills aimed for by higher education and one of the key areas in which it is developed: academic writing. There are two conflicting stories. One believes that AI can serve as a cognitive helper that relieves learners from mundane linguistic tasks, allowing them to focus on more complex thinking skills (Kalyuga & Plass, 2025). The other warns that the transfer of the generation of ideas and the building of arguments results in cognitive offloading, cognitive passivity and less

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

independent thinking (Deep & Chen, 2025; Gerlich, 2025).

The current evidence has three drawbacks that warrant the present study. First, many papers focus on the term "AI use" without differentiating between assistive use and substitutive use of AI. Secondly, numerous studies focus on self-reported attitudes rather than on how students express their use of the AI's outputs. Third, there is limited evidence, especially in the context of higher education in developing countries, and particularly Pakistan, about how students use AI tools, as exam-oriented evaluation and large class sizes can have an impact on this.This research is aimed to fill these gaps, where the study explores the relationship between the use of AI writing tools and critical thinking in English writing within a sample of undergraduate students of Pakistan. It poses two questions. RQ1: How is AI writing tool use associated with undergraduates' cognitive skills (critical thinking) in English writing? RQ2: How often do students critically assess the result of AI when using it? Based on RQ1, we put forward that the use of AI writing tools would be significantly related to critical thinking (H1).

2.Literature Review

2.1 AI writing tools as cognitive and pedagogical mediators

AI writing tools leverage natural language processing and machine learning to create, edit, and enhance content, ensuring it remains relevant and compelling (Rahman, 2026; Deep & Chen, 2025). Adiguzel et al. (2023) define them as "mediational artifacts" which assist the development of the learner's language from the sociocultural point of view. Kalyuga and Plass (2025) take it one step further and state that AI can be a "cognitive partner" that takes over mundane language tasks, freeing up time and mental bandwidth for more complex processing. This positive outlook is challenged. Students' acceptance of AI corrections in Amaria (2025) tends to be without learning the rules, meaning that AI is taking the place of the cognitive engagement, rather than supporting it. Thus, the role of AI seems to be dependent on either its use by students or the design of the teaching and learning process, as well as learner autonomy.

2.2 Cognitive Offloading and Cognitive Load

Cognitive Load Theory suggests that removing extraneous (unnecessary) load in learning would enable the working-memory to allocate resources for deeper processing of the material (Sweller et al., 2011). This makes AI writing tools a perfect fit, as they can handle repetitive tasks like grammar corrections and sentence structure restructuring. The theory's weakness in this context is that the lower the load, the less the deep learning must be. Learners who offload tasks, such as idea generation and argument-building and use language as a thinking tool, can successfully perform tasks and think little, resulting in a gap between task performance and actual learning (Kalyuga & Plass, 2025). AI can also remove the "desirable difficulty" that deep learning requires by supplying quick answers. The theory, therefore, accounting for immediate changes in performance is not sufficient to guarantee either an increase or a decrease in critical thinking.

2.3 AI Writing Tools and Critical Thinking

Most of the recent evidence is an alarming warning. For example, in the field of analytical ability, Amaria (2025) reported that frequent AI users evidenced lower skills in evaluating arguments and creating new interpretations. Pakistanis' students faced a negative trend when they depended on AI tools in their study, Khan 2024 said, indicating a reduction in their ability to think critically and independently. Deep and Chen (2025) suggest that AI fosters cognitive passiveness, where students accept the content created by AI without questioning it. Gerlich (2025) also uncovered a significant negative correlation between the use of an AI tool and critical thinking, with this relationship being mediated by the act of cognitive offloading, and a stronger association with critical thinking scores among younger participants, who were more likely to depend on the tool.

The other work is more positive. While interacting with AI, students may be encouraged to adjust, reflect and reconsider their arguments, thereby enhancing their critical thinking and

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

problem-solving abilities.AI feedback can stimulate critical thinking by encouraging students to adjust, reflect and reconsider their arguments, thereby improving their critical thinking and problem-solving skills. In sum, these studies indicate that the impact of AI on critical thinking is not necessarily positive or negative, but can be either substitutive (AI replacing thinking) or augmentative (AI supplementing thinking). Context also matters. AI can enhance the coherence and structure of undergraduate writing, but overreliance can lead to loss of independence in writing (Hussain, 2025), and AI cannot guarantee learning outcomes (Sadia, 2025) are noted in Pakistani studies. Because of the nature of the exam, the format of the lessons and the few opportunities for students to write on their own, students might seek to use AI to speed up the process and gain efficiency. Students may be driven towards AI for efficiency, as exam-oriented assessment, large classes, and limited individual writing instruction can push students towards this option for speed.

2.4 The Present Study

Existing studies have found conflicting directions of the relationship, used different measures and rarely focused on Pakistani undergraduates. The current study thus examines the extent of use of AI writing tools and the self-reported engagement in critical thinking with AI-generated content, as well as investigates the link between the two methods.

2.5 Design and Participants

The study adopted of quantitative and cross-sectional survey design. The target population consisted of undergraduate students, who are taking academic writing assignments in English at the Universities in Pakistan. Participants were recruited through convenience sampling by sharing an online questionnaire via WhatsApp groups, university student organizations and informal networks. The inclusion criteria included: (a) current undergraduate students in a Pakistani university, (b) knowledge of AI writing tools like ChatGPT, Grammarly, or QuillBot, and (c) experience with academic writing in English. A total of 200 students did the final sample. The characteristics of these are listed in Table 1.

Most were female (73.0%), aged 21–23 (60.0%), in their seventh or eighth semester (54.5%), and studying English or Linguistics (69.0%). The most used was ChatGPT (83.5%).

Table 1: Characteristics of the Participants (N = 200)

Variable	Category	*N*	%
Gender	Male	54	27.0
	Female	146	73.0
Age (years)	Below 18	2	1.0
	18–20	51	25.5
	21–23	120	60.0
	24–26	18	9.0
	Above 26	9	4.5
Semester	1–2	25	12.5
	3–4	29	14.5
	5–6	37	18.5
	7–8	109	54.5
Field of study	English/Linguistics	138	69.0
	Social Sciences	8	4.0
	Natural Sciences	7	3.5
	Other	47	23.5
Main AI tool used	ChatGPT	167	83.5

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

	Grammarly	4	2.0
	QuillBot	3	1.5
	Other	26	13.0
Self-rated AI-use frequency	1 (Least)	6	3.0
	2	20	10.0
	3	74	37.0
	4	60	30.0
	5 (Extreme)	40	20.0

Note. Percentages are calculated on N = 200.

2.6 Instrument

A structured questionnaire was designed based on themes in the literature to gather data on the themes of cognitive enhancement versus dependency, critical thinking and analytical ability, and patterns of AI usage. Items were based on previous theory and research and thus content validity was supported. Here two scales are examined. The items were reviewed for clarity and relevance by the supervisor and subject experts, and checked the relevance by a small pilot group who informed the wording changes. Answers were provided using a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). AI writing tool uses initially had six items. Two were removed due to their negative impact on internal consistency, leaving four items (e.g., use AI when having trouble generating ideas, use AI to review writing draft, use AI as a part of the writing process, compare writing to AI alternatives). The critical thinking (the cognitive-skills outcome) had seven items: Evaluating AI support for arguments, exploring alternate ways of presenting an idea, checking the accuracy and relevance of AI-generated content, re-evaluating the logical flow, identifying weaknesses in their arguments, and uncritically accepting AI suggestions (which is negatively worded—item 13). Participants also rated on a five-point item how often they use AI (1 = least to 5 = extreme) which is used for forming groups. Internal consistency was evaluated using the traditional criterion of $\alpha \geq .70$ (Nunnally, 1978) and the results are presented in Table 2.

2.7 Procedure and Ethics

The questionnaire was sent to the respondents via a web-based platform for about a week in May, 2026. It included an information and consent statement that explained the study, how participation was voluntary, how the subjects would be anonymous, and that they had the right to withdraw from the study. It began with an information and consent statement, which explained the purpose of the study, voluntary participation, anonymity and the right to withdraw. There was no information gathered that could identify the individual and the length of time it took to complete was approximately 8–10 minutes and was used only for research. This study was conducted under the ethical protocols of University of the Punjab, Institute of Education and Research. Incomplete responses and respondents which did not meet the inclusion criteria were excluded.

2.7 Data Analysis

The data were analyzed using SPSS. Cronbach's alpha and mean inter-item correlations assessed reliability. Each scale and item were described by means of descriptive statistics. One-sample t-tests were used to compare scale means with neutrality (3) in a 3-point scale. The results of the one-way ANOVA were used to compare critical thinking across the five levels of frequency of self-reported AI use using eta squared (η^2) as effective size. The relationship between AI writing tool use (predictor) and critical thinking (outcome) were analyzed using Pearson's correlation and simple linear regression. The level of significance was set at $\alpha = .05$. Cohen's (1988) benchmarks are used to interpret effect sizes.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

2.8 Reliability

The internal consistency of critical thinking was acceptable (α = .75, mean inter item r = .35). Even after the removal of 2 items, the AI writing tool use scale did not meet the traditional cutoff (α = .61, mean inter item r = .28), but it is retained due to its theoretical importance, and the results should be interpreted with caution (Table 2).

Table 2

Reliability, Descriptive Statistics and One-Sample t-Tests against the Midpoint (Test Value = 3)

Scale	Items	α	*M*	*SD*	*t*(199)	*p*	95% CI of mean diff.	Cohen's *d*
AI writing tool use	4	.61	3.96	0.43	31.46	< .001	0.90, 1.02	2.22
Critical thinking	7	.75	3.78	0.51	21.62	< .001	0.71, 0.85	1.53

Note. Scales range from 1 to 5. Mean differences from the midpoint were 0.96 (AI writing tool use) and 0.78 (critical thinking). Cohen's *d* is the mean difference divided by the sample *SD*.

***2.9* Extent of AI Writing Tool Use**

The mean score for the students' use of AI writing tools was 3.96 (SD = 0.43), well above neutral (Table 2). The means for the four items were similar, ranging from 3.90 to 4.00 (Table 3). Dropping off ideas when writing had the highest mean-value, and using it to review organization had the second highest mean value.

Table 3

Item-Level Descriptive Statistics for AI Writing Tool Use

Item	*N*	*M*	*SD*
Consulting AI for idea generation	200	4.00	0.63
Reviewing organization of written work	200	3.97	0.60
Regular part of the writing process	200	3.97	0.65
Comparing own writing with alternatives	200	3.90	0.65

Note. Items rated 1 (strongly disagree) to 5 (strongly agree).

2.10Critical Thinking

Still, students' critical thinking while using AI was significantly higher than the midpoint (M= 3.78, SD = 0.51; Table 2), but the lower of the two-scale means. The mean degree of agreement was between 3.84 and 3.97 for all but one of the seven items, which included re-evaluating logical flow, considering different ways to present an idea, checking accuracy of AI content, checking for relevance to an assignment, recognizing weaknesses in arguments, checking for consistency of a concept, and checking for bias. The item with the lowest mean score and the widest SD was the item regarding "uncritically accepting AI suggestions" (M = 3.07, SD = 1.11), meaning that students on average neither endorsed nor rejected this item and their responses were divided. That is, while the majority of students described an active assessment

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

of the output of AI, a substantial minority did not.

2.11 Critical Thinking by Frequency of AI Use

A one-way ANOVA revealed significant differences between the five levels of self-rated frequency for critical thinking, $F(4, 195) = 6.19$, $p < .001$, $\eta^2 = .11$; this was a medium sized effect (Table 4). The mean scores for the least frequent group ($M = 3.28$) were lowest and the mean scores for the extreme group ($M = 4.01$) were highest, with a slight drop at level 3 ($M = 3.63$) compared to level 2 ($M = 3.77$). The omnibus test does not specify which pair(s) differ, and the smallest group ($n = 6$) was very small, so comparisons among groups should be interpreted as tentative.

Table 4

Critical Thinking by Self-Rated Frequency of AI Use

AI-use level	***N***	***M***	***SD***
1 (Least)	6	3.28	0.36
2	20	3.77	0.42
3	74	3.63	0.46
4	60	3.86	0.56
5 (Extreme)	40	4.01	0.43

Note. One-way ANOVA: $F(4, 195) = 6.19$, $p < .001$, $\eta^2 = .11$.

***2.*12 Association Between AI Writing Tool Use and Critical Thinking**

The use of AI writing tools is significantly associated with critical thinking. There is a strong correlation between the use of AI writing tools and critical thinking.

There was strong and positive correlation between AI writing tool use and critical thinking, $r(198) = .76$, $p < .001$. The simple linear regression, with AI use of writing tool as the predictor, and critical thinking as the outcome, was significant, $F(1, 198) = 266.6$, $p < .001$, $R^2 = .575$, explaining 57.5% of the variance. Each one-point higher score on AI writing tool use was associated with a 0.88-point higher critical-thinking score ($B = 0.88$, $SE = 0.05$, $\beta = .76$, $t = 16.33$, $p < .001$; Table 5). This was to support H1.

Table 5

Simple Linear Regression Predicting Critical Thinking from AI Writing Tool Use

Predictor	***B***	***SE***	**β**	***t***	***p***
Constant	0.29	0.21		1.37	.172
AI writing tool use	0.88	0.05	.76	16.33	< .001

Note. $N = 200$. $R^2 = .575$. Dependent variable: critical thinking.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

3 Discussion

The present study explored the relationship between the use of AI writing tools and critical thinking in English writing among undergraduate students in Pakistan. Three conclusions are noted. First, AI is already commonplace: students reported using AI to brainstorm ideas, to verify the organization and to compare alternative wordings, at scores significantly higher than the mid-point. Second, students reported that they critically used AI-generated content, verifying the accuracy and relevance of what they were reading, and challenging their own thoughts, with a significant proportion saying that they would take suggestions from AI without questioning them. Third, the higher the frequency of AI use, the more extensive the use, and the more the higher the reported levels of critical thinking, both between frequency groups ($\eta^2 = .11$) and in a strong correlation ($r = .76$).

This positive correlation aligns with the idea that AI can serve as a cognitive partner, assisting learners in tasks that are repetitive and allow them to focus on evaluating and arguing their thoughts (Kalyuga & Plass, 2025), as well as Graham and Milan's (2025) perspective that AI feedback can encourage students to reflect on their reasoning. Students who heavily rely on AI might be students who work to reduce extraneous load and then spend the extra time checking and refining (Sweller et al., 2011). The item-level results align with this assistive pattern, as students predominantly mentioned using AI for generating ideas and structuring ideas, and checking the output of AI.

The findings, however, run counter to findings of associations between higher levels of AI use and lower critical thinking (Amaria, 2025; Gerlich, 2025; Khan, 2024), and cognitive passivity (Deep & Chen, 2025). There are multiple possible explanations. This critical thinking scale measures how students interact with AI-generated content, not their actual critical thinking abilities, and could be a self-perceived measure. Respondents were predominantly English or Linguistics students (69%) and might be more mindful about language and writing than other undergraduates. Some respondents may have provided socially desirable responses. Furthermore, the correlation is likely to be inflated by common-method variance since both the scales were self-reported in the same survey (Podsakoff et al., 2003). The correlation might also be in the direction of students who already think critically,using AI more purposefully. The split on the uncritical-acceptance item, as well as the lower score for critical thinking than use, indicates that the picture is not altogether bright. The lowest frequency group was the lowest on critical thinking, but the "gap" at level 3 indicates that the relationship is not linear. Results have implications with care. To support the use of AI as an augmentative tool, instructors may develop writing assignments that involve justifying, verifying and revising AI-generated text. AI-literacy training can be created to make the difference between helping and replacing thinking explicit, especially for the small proportion of students who report taking advice without question.

4. Limitations and Future Research

Several limitations apply in this paper. The type of design is a cross section; therefore, no causal or long-term statements can be made; the use of the word impact should not be interpreted as associations reported. The convenience sample is not representative of Pakistani undergraduate students as it is mainly composed of female and English/Linguistics students, and is made up of ChatGPT users. All the measures are self-reported and hence there are inherent perception and social-desirability bias and common method variance. AI writing tool use scale demonstrated a low reliability ($\alpha = .61$) that can weaken or skew relationships. Small usage group ($n = 6$) was also lowest and pairwise comparisons were not reported. The study did not differentiate between substitutive and assistive use, and it did not differentiate between tools; instead, it was measures of perceived critical thinking rather than actual critical-thinking task.

5. Conclusion

The majority of 200 Pakistani undergraduates used AI writing tools and there was a strong and positive correlation between using these tools and students' critical thinking regarding English writing. The students discussed using AI primarily for ideas and organization, and the majority of them analyzed the results, while some few took what they were given without questioning it. The results are in line with assistive use and do not indicate that AI can enhance critical thinking skills. More robust designs are required to set direction and to distinguish the impact of the use of AI from its quantity.

Declarations

Ethics approval and consent: Informed consent was obtained from participants before they answered the questionnaire via electronic means.

References

Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology, 15*(3), *Article ep429*. https://doi.org/10.30935/cedtech/13152

Amaria, F. (2025). The impact of AI writing tools on academic writing skills in English learning classroom. *International Journal of Language and Education Research, 7*(3), *159–177*. https://doi.org/10.5281/zenodo.23677039

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Deep, P. D., & Chen, Y. (2025). The role of AI in academic writing: Impacts on writing skills, critical thinking, and integrity in higher education. *Societies, 15*(9), *Article 247*. https://doi.org/10.3390/soc15090247

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), *Article 6*. https://doi.org/10.3390/soc15010006

Graham, O., & Milan, Y. (2025). *The impact of ChatGPT reliance on the development of student critical thinking skills* [Preprint]. Preprints.org. https://doi.org/10.20944/preprints202506.0861.v1

Hussain, S. A., Faizan, K., Ishtiaq, M., & Bibi, L. (2025). Overreliance on artificial intelligence in education and its effects on students' creativity and analytical abilities. *Review of Applied Management and Social Sciences, 8*(4), *1843–1857*. https://doi.org/10.47067/ramss.v8i4.655

Kalyuga, S., & Plass, J. L. (2025). *Rethinking cognitive load theory*. Oxford University Press. https://doi.org/10.1093/9780190078539.001.0001

Khalid, S., & Ahmed, H. N. (2025). Using AI to develop English writing skills: Pakistani undergraduate students' beliefs and practices. *Social Sciences & Humanity Research Review, 3*(4). https://doi.org/10.63468/sshrr.273

Khan, W. M. (2024). Analyzing the AI tools' impact on critical thinking in BS English students at Pakistani universities. *Journal of Applied Linguistics and TESOL, 7*(4), *1232–1238*.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), *879–903*. https://doi.org/10.1037/0021-9010.88.5.879

Rahman, W. Z. (2026). Artificial intelligence (AI) in English writing skills learning: A systematic review. *Indonesian Journal of Research and Educational Review, 5*(2), *981–994*. https://doi.org/10.51574/ijrer.v5i2.4727

Roe, J., Renandya, W. A., & Jacobs, G. M. (2023). A review of AI-powered writing tools and their implications for academic integrity in the language classroom. *Journal of English and Applied Linguistics, 2*(1), *Article 3*. https://doi.org/10.59588/2961-3094.1035

Sadia, S., Ahmad, K., & Ali, Z. (2025). The role of AI tools in enhancing English language proficiency: A case study of Pakistani undergraduate students. *Journal of Applied Linguistics and TESOL, 8*(2), *40–46*.

Shahzad, S., & Zaman, N. (2025). Using AI for academic writing: Perceptions of students from universities of Lahore. *Journal of Arts and Linguistics Studies, 3*(4), *6253–6284*. https://doi.org/10.71281/jals.v3i4.533

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4