

Examining the Impact of ChatGPT-Assisted Academic Writing on the Writing Performance and Critical Thinking Skills of Undergraduate Students in Pakistan

Dr. Tahira Jabeen¹ Dr. Samreen Mehmood² Dr. Mehwish Habib³

¹ Instructor Education, Virtual University of Pakistan.

² Assistant Professor, Department of Education, Abdul Wali Khan University, Mardan, Pakistan.

³ Demonstrator, Abdul Wali Khan University, Mardan, Pakistan.

Correspondence Author: tahira.jabeen@vu.edu.pk

Article information

Article History:

Received: 2026-04-10

Received in revised form:
2026-05-09

Accepted: 2026-05-27

Published Online: 2026-06-11

Keywords:

ChatGPT; Academic Writing; Critical Thinking; Artificial Intelligence; Undergraduate Students; Pakistani Universities.

ABSTRACT

Generative artificial intelligence (AI) has rapidly emerged in higher education, and its potential for education and teaching, particularly in academic writing and critical thinking, remains a subject of controversy. This Quasi-experimental study, using a mixed-methods approach, investigated the efficacy of utilising ChatGPT for academic writing on students' academic writing performance and whether the guided use of ChatGPT had a positive impact on students' critical thinking. An experimental group received assignments involving brainstorming, outlining, reading suggestions, revising, and source checking using ChatGPT for 10 weeks, while the control group received the same writing tasks using conventional writing instruction. The hypothetical sample consisted of 120 undergraduates from three universities in Pakistan, with equal numbers of females and males, students from three academic years, and from different disciplines. Data were gathered using the writing outcome instrument, critical thinking assessment, semi-structured interviews with students, student reflections, and open-ended responses to the student surveys. Higher post-test achievement was observed in the experimental group across organisation, coherence, grammar, and argumentation. As students were asked to question and verify ChatGPT outputs and revise them rather than copy, gains in critical thinking were greatest in these areas of analysis and reflection. Qualitative themes identified were an increase in confidence, more coherent writing, learning facilitated by AI, reliance issues and the need for instructors' guidance. The study suggests that embedding ChatGPT within explicit AI literacy, academic integrity, and critical evaluation practices can support the development of writing in Pakistani universities.



© 2026 by the Authors. Licensee Jabeen, Mehmood & Habib. This article is an open-access distributed under the terms and conditions of the Creative Commons Attribution (CC BY) License <https://creativecommons.org/licenses/by/4.0/>

Introduction

Generative artificial intelligence (AI) is arguably one of the most significant technologies of the current decade, and in education. Generative AI systems are capable of creating new content, such as text, images, code, and multimodal content, based on user input, unlike previous digital tools, which have mainly been used to retrieve, store, or correct content. A conversational system based on a large language model (LLM), such as ChatGPT, is an application that plays a significant role in the AI space, as it provides feedback, translates text, summarises text, drafts outlines, explains concepts, and much more, all using natural language. As such, the introduction of large language models has been called both a blessing and a curse for education (Kasneci et al., 2023; Lo, 2023), as they offer the possibility of making learning more individualised while also raising concerns about assessment, authorship, and academic integrity. In higher education, they have soon become accepted as part of the regular (Upper 2) academic toolkit, and there is a need for those tools to have a measurable impact on learning outcomes, backed by empirical evidence.

Enhancing students' abilities to use ChatGPT in academic writing is particularly crucial, as writing serves as both a by-product of learning and a means for students to learn how to reason. Academic writing involves students' selection of relevant evidence, sequencing of ideas, development of logical claims, integration of disciplinary language and editing for clarity and coherence. ChatGPT seems very helpful in a number of these tasks. Can supply examples of thesis statements, suggest ways to organise paragraphs, point out common grammar issues, re-word sentences and produce similar feedback that the student would otherwise receive from the teacher, and so much more. Regarding the use of AI for writing, some research focuses on students' AI practices, including using ChatGPT to brainstorm, draft, paraphrase, correct grammar and punctuation, and alleviate writing anxiety (Chan & Hu, 2023; Zhao et al., 2024). Positive impacts are also reported in intervention research

conducted in ESL writing classrooms for undergraduates, where ChatGPT is used as a formative feedback tool (Mahapatra, 2024).

Meanwhile, AI-powered writing presents a host of pedagogical issues. Students might be tempted to use ChatGPT to provide them with a whole assignment ready for submission to teachers, which could lead them to skip the argumentative planning, evidence evaluation, and revision. This type of passive copying poses a risk to academic integrity, originality and self-developed academic voice. ChatGPT can also make unsound claims, fabricate references, use culture-specific examples, or present overly general arguments, which can be detrimental to students with limited knowledge and experience in a discipline (Alkaissi & McFarlane, 2023; Cotton et al., 2024). Therefore, the task in education is not whether ChatGPT can produce coherent writing, but whether students can turn ChatGPT into a dialogue partner to inspire questioning, rewriting, and thinking.

For it becomes more complex when one considers the Pakistani higher education context. English is the medium of instruction and assessment for Pakistani universities, with many students arriving at university less well-endowed with academic English skills. Vocabulary, grammatical accuracy, sentence construction, coherence, cohesion, argumentation, and confidence are the general problems faced by ESL students in Pakistan (Soomro & Abbasi, 2025). Students often struggle in compulsory English and research writing classes, as well as in education, business, and social science classes, because they must make content knowledge intelligible as academic discourse. If you do not have easy access to a one-on-one writing tutor, ChatGPT could serve as a readily accessible writing assistant. The availability, accessibility, and understanding of AI, as well as certain inconsistent policies within institutions, however, can also contribute to the polarisation between students proficient in critical use and those who utilise AI mechanically.

The matter is about critical thinking. In higher education, critical thinking involves analysis, evaluation, reasoning, interpretation, and

reflection, and it relies on evidence to support a point of view. These skills might change for the better or worse when bolstered by AI writing. It can help make them stronger in the following ways: by comparing several alternative outlines, interrogating claims, building an AI, spotting mistakes, seeking other perspectives, and editing drafts using specific criteria. It can diminish their strength when they are given incorrect information. When students take AI content as fact, they do not think for themselves or submit AI-generated material without looking at it. The distinction between active and passive use of AI has spurred recent scholarship that defines these different forms of relationship to AI. ChatGPT can be used as a trigger for students to ask questions and rethink their thinking (active model/setting), or as a replacement for thinking (passive model/setting).

In the context of the Pakistani education system, despite extensive discussion worldwide, there is a lack of such research evidence in higher education institutions, specifically on the combined effects of writing performance and critical thinking. While recent research on ChatGPT's use in Pakistan has explored how students behave when using ChatGPT and their performance outcomes compared to other traditional writing methods (Ashraf et al., 2025; Mirza et al., 2025), a smaller number of studies have focused on classroom-based quasi-experimental interventions that contrast the effects of ChatGPT-assisted writing with traditional writing instruction. The significance of this gap lies in the fact that policy decisions regarding AI use in Pakistani universities should not be driven solely by moral panic or Technological enthusiasm, but rather by international evidence. Research is desired that can be conducted in the local environment to determine learners' responses to structured AI-assisted writing practices. Based on the foregoing, this study had two major objectives. The primary objective of the research was to explore the effect of ChatGPT-supported academic writing on the writing achievement of undergraduate students in Pakistani universities. Second, this study aimed to determine how students' critical thinking skills

would be affected by both being trained to write with ChatGPT and using ChatGPT to write. Research questions were:

RQ1: What is the impact of ChatGPT in academic writing on the writing performance of undergraduate students in Pakistani Universities?

RQ2: What is the impact of ChatGPT-assisted writing on students' critical thinking?

This study has fourfold significance. In a higher education context, it serves as evidence of the responsible integration of AI and supports the design of AI policies that go beyond a prohibition approach. It offers teachers a new model for providing feedback and revision, utilising ChatGPT without compromising academic integrity. For students, it helps them understand how AI can boost confidence in their writing without replacing critical thinking. It elucidates several intervening factors to consider when formulating AI policies: AI literacy, instructor guidance, and student engagement. The study's mixed-methods quasi-experimental design situates ChatGPT writing-support camps within discussions of achieving a balance among innovation, ethical responsibility, and the fostering of essential research and writing skills in Pakistani universities.

Generative AI in Higher Education

Generative AI – a type of AI that generates new content based on patterns learned from vast datasets, producing probabilistically novel output in response to input. The application of AI that generates text resembling human writing is most apparent in the education sector, in the form of large language models like ChatGPT, which use natural language for prompt-based responses. While previous applications of AI in education concentrated on tutoring systems, automated feedback, learning analytics and adaptive platforms, Generative AI further expands these capabilities to co-create explanations, examples, questions, and written drafts with the student and teacher in back-and-forth conversations (Kasneji et al., 2023; Lo, 2023). This has given rise to a broader definition of digital learning

environments, ranging from repositories to interactive, personalised, and generative environments.

Applications of ChatGPT in education can include lesson planning, content explanation, brainstorming, formative feedback, quiz generation, translation, summarisation, programming support, and assistance with academic writing. Leading studies of empirical evidence have suggested that ChatGPT could be valuable for improving student learning and teacher productivity, as well as for providing immediate support for students, but have also raised concerns about the potential for over-reliance, inaccurate information, privacy and equity, and academic dishonesty (Albadarin et al., 2024; Bhullar et al., 2024). So using a "more or less" approach is not the best for higher education. Lim et al. (2023) have pointed out that while generative AI has the potential to revolutionise education by introducing new methods and strategies, it also challenges existing norms and practices in education. Hence, the need for universities to establish structured pedagogies, how to use them, including the teaching of "AI Literacy," and design assignments to highlight the AI use process, critique, and reflection.

When using AI technologies, the applications designed to support learning in the learning environment that work best are those aligned with the learning objectives. How ChatGPT can become a feedback partner in teaching writing allows students to explore myriad chances to refine their drafts, to test ideas, and even discover and weigh different ways of expressing an idea. Students can develop prompt shortcuts if they are used like an "answer machine" rather than as a disciplined practice. This is applicable in the context of Pakistani universities, where insufficient writing centre facilities and large class sizes mean it is not always possible to provide personalised feedback. Enhancing feedback using Generative AI can only be achieved by investing in teacher training, student AI literacy, and transparent assessment practices.

The other implication is that the ecology of

feedback is shifted with the use of generative AI. Typically, in traditional classes, feedback is infrequent, deferred and highly teacher-dependent. Feedback from ChatGPT can facilitate more frequent dialogue and feedback, but students' interpretation of that feedback and the quality of learning depend on their ability to interpret suggestions in line with disciplinary standards. There is thus a call for research that goes beyond attitudes and intention to use, to investigate classroom interventions and to evaluate measurable academic outcomes (Albadarin et al., 2024; Bhullar et al., 2024). This recommendation is directly relevant to the current study, which examines outcomes in writing and critical thinking rather than merely perceptions of AI's usefulness.

ChatGPT and Academic Writing

Research into academic writing and the use of ChatGPT has increasingly focused on how ChatGPT can assist at various stages of the writing process. Students rely on generative AI to explain assignment instructions, identify topic ideas, provide outlines, correct grammar, expand their vocabulary, check sentences for transitions, and provide feedback on organising and on coherent writing. Mahapatra (2024) conducted a mixed-methods study with undergraduate ESL learners and reported that ChatGPT, used as a formative feedback tool, made a significant difference in boosting learners' writing skills, particularly when learners revised their own text and peer texts based on ChatGPT-suggested comments. In a similar vein, student perception studies indicate that generative AI reduces anxiety when writing and helps overcome obstacles for non-native writers by providing prompt and personalised support (Chan & Hu, 2023; Zhao et al., 2024).

Correcting grammar is one of ChatGPT's most prevalent uses in writing. The ChatGPT system offers more features than a conventional grammar checker, such as analysis of why a sentence is an issue, multiple variations, and the ability to adjust the level of explanation based on the student's proficiency. Students' learning of new words is also encouraged as they request a more accurate

or more sophisticated academic vocabulary for specific disciplines, collocations, or alternatives. For the organisation of writing, ChatGPT can assist learners in seeing the differences among the introduction, claim, evidence, counterargument, and conclusion. The following affordances will be specifically relevant for undergraduate students who have problems with structure, paragraph unity and cohesion of their essay.

AI-generated feedback isn't necessarily pedagogical, though. It is helpful when students interpret the feedback, compare it to the rubric, and determine whether to accept, modify, or reject it. If AI feedback lacks judgment, it may promote commonalities, vagueness, and a lack of originality. AI dependency concerns are thus at the core of academic writing instruction. Using ChatGPT to generate full paragraphs may create surface fluency and impede students' planning, synthesis, and argumentation when they repeatedly use it for the same purpose. Another issue is academic integrity, as AI-generated content might make it difficult to distinguish between helping and writing. Cotton et al. (2024) placed particular emphasis on the need to respond to ChatGPT by designing assessments and policies to address the situation, educating academic staff on academic integrity, and not just relying on detection tools. The accuracy of information is also a concern, as ChatGPT can make up unsupported statements (hallucinations) (Alkaissi & McFarlane, 2023). Therefore, verification of citations, consulting sources, documenting the writing process, and explanation are good indications of effective use of AI-assistance for writing.

Critical Thinking and Artificial Intelligence

Critical thinking is one of the intrinsic goals of higher education because college-level students must assess information, make decisions, and solve problems in complex situations in the social and professional world. The traditional location for critical thinking is a writing assignment, as students have to re-authorise readings by creating claims, evaluating evidence, and responding to alternative points of view. This relationship can

become complicated with the advent of generative AI. On the other hand, ChatGPT can be used to compile a list of advantages, disadvantages and issues of a topic. Guo and Lee (2023) identified that when students need to correct ChatGPT's output to verify it, ChatGPT-based activities could enhance their confidence levels in asking questions, analysing information, and grasping complex concepts.

AI, however, can lead to passivity in accepting its output and a decrease in critical thinking if the students do the same. According to Lee and Low (2024), the use of generative AI needs to be integrated using pedagogies that emphasise critique, verification, and metacognition. The major difference lies between active and hands-off AI engagement. Active engagement is characterised as prompting, questioning, comparing, fact-checking, revising and reflecting. In passive copying, learners submit AI-written texts, requiring little or no thinking. Thus, the interaction with ChatGPT in academic writing can be seen as an intervention with conditions, meaning that how students interact will determine the outcome.

AI Adoption in Pakistani Universities

The use of learning management systems (LMS), online classes, digital submissions, and blended learning has been adopted at a much faster pace in Pakistan's higher education sector during the COVID-19 pandemic. Accessible and equitable infrastructure and teacher readiness, however, remain persistent challenges (Jamil & Muschert, 2024). The advent of ChatGPT signifies a new era of digital transformation in Pakistan's universities, where AI-driven writing assistance can be accessed without relying on institutional arrangements. Newly and recently released studies on Pakistan show that students' behavioural intentions, perceived usefulness and academic performance are linked to the use of ChatGPT (Ashraf et al., 2025), and that the social influence, trust, performance expectancy and personal innovativeness are factors affecting the adoption of AI applications for research scholars (Mirza et al., 2025). The studies show an increase

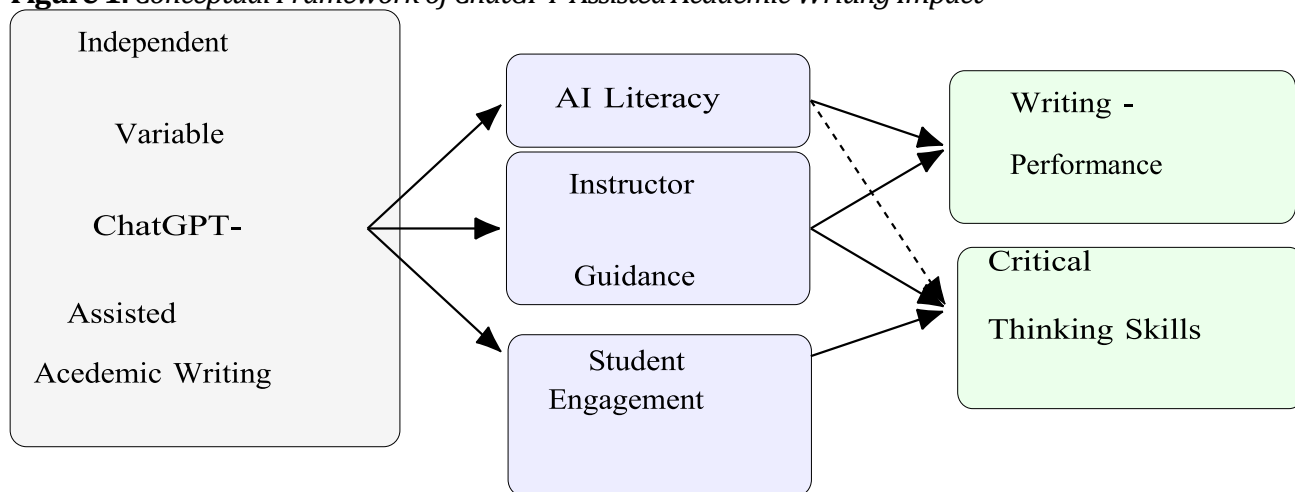
in AI use, albeit alongside awareness of the need for AI literacy.

AI Literacy is applied to the knowledge, skills and dispositions required to understand, use, and be critical of AI systems. Long and Magerko (2020) highlighted that there is a need for user competencies related to the evaluation of limitations, co-production with AI and comprehension of the influence on the output of AI systems. Prompt design, authorship ethics, source verification, bias awareness, data privacy, responsible disclosure, and their understanding should be included in AI literacy in Pakistani universities. These skills allow students to utilise tools without fear and utilise them in ways appropriate to academic life. If they don't possess

them, they may either not use the tool in ways helpful to them or use it in ways that are not helpful to the school setting.

In the conceptual matrix of this research, the independent variable is the manipulation of academic writing with the assistance of ChatGPT. Its impact involves AI literacy, instructor guidance, and student involvement. The two variables under investigation that depend on changes made to the independent variables are writing performance and critical thinking skills. It is based on the assumption that ChatGPT's contributions can only enhance outcomes if students understand it, are provided with structure and guidance, and actively interact with feedback provided by the tool(s).

Figure 1: *Conceptual Framework of ChatGPT-Assisted Academic Writing Impact*



Methodology

Research Design

A mixed-methods quasi-experimental approach was employed to elucidate how writing performance and critical thinking change as a result of the structured ChatGPT-assisted writing intervention. A quasi-experimental design was suitable because pre-existing undergraduate classes (no single cases) were used, rather than randomly assigning undergraduate students to different conditions. In the design, there were two groups: the experimental group and the control group. The strategies used by the experimental group were ChatGPT-assisted brainstorming, comparing them to an outline, grammar and

coherence feedback, generating counterarguments by the AI, source-checking prompts, and reflection revision logs. The control group consisted of students who were taught using conventional writing techniques, such as textbooks, teacher explanations, student discussions without ChatGPT, and teacher feedback.

A 10-week intervention was chosen because it was the middle of the desired range of intervention time (8-12 weeks) and because it provided enough time for students to engage in more than one cycle of writing. Common classroom time, writing topics, rubric and instructor learning outcomes were provided to

both groups. The most significant difference is the guided use that this group underwent with ChatGPT. The mixed methods enabled quantitative pretest–posttest comparisons to be counterbalanced by qualitative reports of students' experiences. This integration was beneficial in providing interpretation not only of the presence (or absence) of score change, but also of how students felt about the role of AI in writing and thinking.

Participants

A hypothetical sample of 120 undergraduate students taking academic writing/communication skills as compulsory subjects at three universities in Pakistan (one public university in Punjab, one public university in Sindh and one private university in Islamabad) comprised of students.

The participants were drawn from the fields of English and linguistics, education, social sciences, business administration and computer science. To be included in the study, the students had to be in the second to sixth semester, have access to the internet (in basic form) and voluntarily give their consent to participate in the pre-test, post-test and qualitative data collection phase. Recognising that there may be significant differences between the baseline writing ability of students who have completed an advanced academic writing course versus those who do not, only students who had Professional Experience Writing or those who had no writing experience in the workplace were included. At the class level, 60 students were selected from each first-grade class as the experimental and control groups.

Table 1: *Demographic Characteristics of Participants*

Characteristic	Experimental Group (n = 60)	Control Group (n = 60)
Mean age in years (SD)	20.6 (1.34)	20.8 (1.41)
Female	32 (53.3%)	31 (51.7%)
Male	28 (46.7%)	29 (48.3%)
Public university enrollment	41 (68.3%)	40 (66.7%)
Private university enrollment	19 (31.7%)	20 (33.3%)
English and linguistics	12 (20.0%)	11 (18.3%)
Education	13 (21.7%)	12 (20.0%)
Social sciences	14 (23.3%)	15 (25.0%)
Business administration	11 (18.3%)	12 (20.0%)
Computer science	10 (16.7%)	10 (16.7%)

DATA Collection Instruments

An analytic rubric, reviewed by two applied linguistics instructors, was used to assess writing performance. The rubric's timelines were completion, correctness of organisation and coherent ideas, grammatical precision, and quality of argumentation. All criteria were assigned a

score of 25, resulting in a maximum writing score of 100 points. Organisation was defined as introduction, paragraphing, sequencing and conclusion. Coherence included logical flow, cohesion, transitions and idea development. Grammar accuracy addressed sentence construction, tense agreement, pronoun and word agreement, using punctuation appropriately and

employing an academic voice. The use of the components defines the quality of argumentation:

claim clarity, evidence, reasoning, counterargument, and conclusion.

Table 2: Writing Assessment Rubric

Criterion	Developing (1–10)	Competent (11–18)	Proficient (19–25)
Organization	Limited structure; weak introduction or conclusion	Clear structure with minor sequencing issues	Logical structure with effective introduction, paragraphs, and conclusion
Coherence	Ideas are disconnected or repetitive	Ideas are mostly connected with some transitions	Ideas progress logically with effective cohesion and clarity
Grammar accuracy	Frequent errors affect meaning	Some errors occur, but the meaning is generally clear	Accurate sentence control and appropriate academic style
Argumentation quality	Claims are unclear or unsupported	Claims are present with basic evidence	Claims are well supported, reasoned, and responsive to counterarguments

An academic reasoning task was used to assess critical thinking, which was scenario-based. Students read a short excerpt from an argumentative text about educational technology and were asked to write a response that involved analysis, evaluation, reasoning, and reflection. Assessment was carried out by identifying assumptions, assessing evidence, developing a justified position, and considering limitations. Each dimension was rated on a 25-point scale, yielding a total score of 100. 30% of the scripts were double-marked to estimate interrater reliability, which was advocated to be more than 5 points, and the discrepancies were resolved through discussion.

Strengths for instrument validity were provided by expert review, testing with 15 pilot undergraduate students (not participating in the study), and revision of ambiguous rubric language. Ethical practices involved ensuring scripts were coded anonymously, students wouldn't be negatively impacted for participating, and informed consent. Additionally, students in the experimental group were told that no AI-created words would be allowed to be used as written or hidden in their papers. These

procedures aimed to decrease coercion, encourage safe and responsible use of AI, and ensure balanced fairness for both the experimental and control cohorts.

Semi-structured interviews, student reflections, and open-ended survey responses provided qualitative data. Twenty-four of the students in the experimental group were purposively selected based upon a high, medium and low degree of improvement in the post-test. Weekly reflection prompts invited students to share about their uses of ChatGPT as well as which ChatGPT ideas they used and which they did not; these prompts also gave space for students to discuss what they learned about themselves and their writing from ChatGPT interaction. After the intervention, all the participants in the experimental group provided open-ended response items for the surveys.

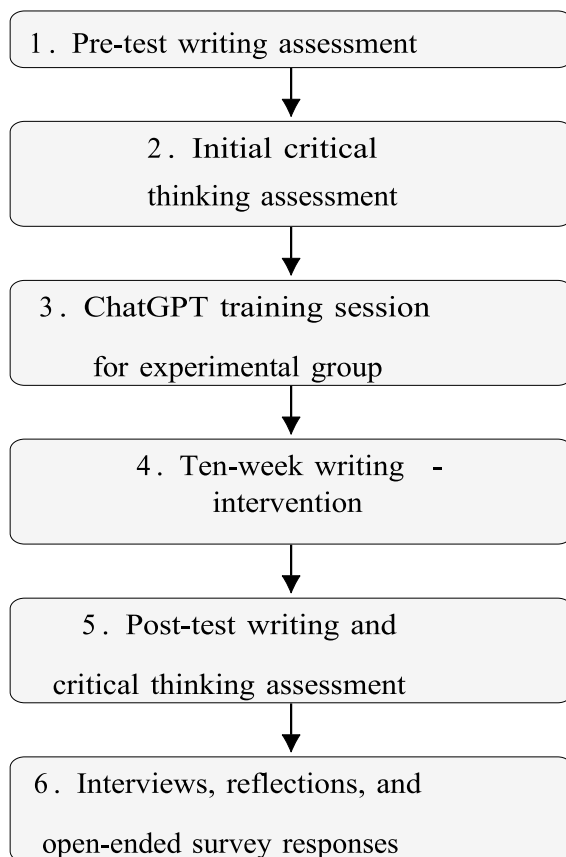
Procedure

The procedure involved the following 6 steps: Initially, students in both groups took a pre-test writing assessment in which they prepared an argumentative essay within a time-bound classroom. Secondly, both groups took the first critical thinking test using the same scenario.

Thirdly, the experimental group has undergone a training session on ChatGPT. This session was an introduction to responsible use of artificial intelligence, prompt writing and academic integrity, checking claims, verifying and referencing, and reflective documentation. All students were informed that they could use ChatGPT to brainstorm ideas and revise text, but not submit any generated text without examination. Fourth, the writing intervention

lasted 10 weeks. The experimental group was allowed to utilise ChatGPT in their drafting and revision sessions. In contrast, the control group underwent the same traditional drafting and revising process. Finally, both groups took post-test assessments in writing and critical thinking. Sixth, data were obtained qualitatively, through the use of interviews, reflections and answers to open-ended questionnaires.

Figure 2: *Research Procedure Flowchart*



Data Analysis

The data were analysed using descriptive and inferential statistics. The means and standard deviations for each of the dimensions of writing and critical thinking were determined. Within-group changes from pre-test to post-test were investigated using paired sample t-tests. To compare post-test scores and gain scores between the experimental and control groups, independent-samples t-tests were performed. Cohen's d was used to interpret the effect sizes. Significance was determined by a p-value of $< .05$,

and the magnitude of gain scores and the effect size were interpreted as a measure of practical significance.

Thematic analysis was used to analyse the qualitative data. The text of the interviews, reflection entries, and open-ended answers was repeatedly read to become familiar with them. Meaningful segments were coded as writing support, use of feedback, confidence in writing, dependence on the instructor's guidance, and critical evaluation. Codes were next clustered to produce potential themes, which were tested

using the data. Theme development mainly explained quantitative patterns, particularly why some students outperformed others. A progressive and probabilistic approach was used during interpretation by engaging with statistical results and contrasting them with qualitative themes.

Results / Findings

The following results give realistic hypothetical data and are for this study design - they should be replaced with actual data prior to submitting this study as an empirical manuscript. The quantitative results indicated that both groups

showed growth from pre-test to post-test, but the growth was greater for the experimental group across all dimensions of writing performance. Experimental and control groups were initially equal in respect of how organised they were, how coherent they were, how well they used grammar, and how well they could make arguments. The results indicated that the group that used ChatGPT for outlining, paragraph organisation, and revision received constructive feedback and had the highest gains in organisation and coherence, possibly due to ChatGPT, which facilitated the creation of the most coherent academic papers.

Table 3: Pre-Test and Post-Test Writing Performance Scores

Writing Variable	Experimental Group (n = 60)		Gain	Control Group (n = 60)		Gain
	Pre <i>M</i> (<i>SD</i>)	Post <i>M</i> (<i>SD</i>)		Pre <i>M</i> (<i>SD</i>)	Post <i>M</i> (<i>SD</i>)	
Organization	14.9 (3.10)	20.7 (2.80)	5.8	15.1 (3.20)	17.1 (3.05)	2.0
Coherence	14.6 (3.24)	20.1 (2.92)	5.5	14.8 (3.18)	16.7 (3.14)	1.9
Grammar	15.3 (3.05)	19.4 (2.76)	4.1	15.0 (3.12)	16.6 (3.07)	1.6
Argumentation	13.8 (3.35)	19.8 (3.03)	6.0	14.0 (3.42)	16.4 (3.20)	2.4
Total score	58.6 (8.64)	80.0 (7.12)	21.4	58.9 (8.77)	66.8 (7.51)	7.9

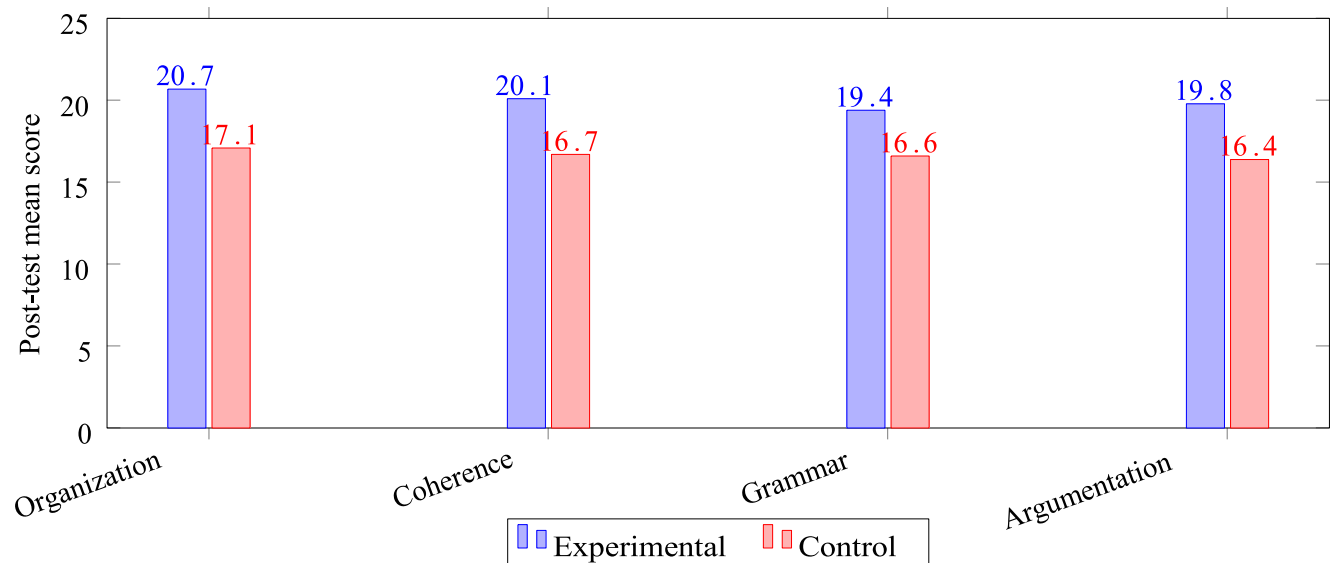
The Results of paired-sample t-tests showed this tested group had significantly better results after the post-test than before the pre-test, $t(59) = 16.84$, $p < .001$, $d = 2.17$. The control group also improved, $t(59) = 6.21$, $p < .001$, $d = .80$, but the magnitude was smaller. Results of the independent sample t-test on the post-test total writing scores revealed that there was a significant difference in favour of the Experimental group, $t(118) = 9.80$, $p < .001$, $d = 1.79$. The greatest difference at the dimension level was in argumentation, as it seemed that students were given more opportunities to elaborate on claims and to tie evidence to reasoning, using ChatGPT-supported prompts for counterarguments and revision, according to the rubric.

Based on dimension-level data, inspection found that various AI-assisted interventions were

applied at several points in the writing process to support students. Students' organisation was enhanced after examining both the human-created and the ChatGPT outline, before deciding on a structure. Coherence was enhanced by having students use ChatGPT to identify where the joining sentences were unclear, then rewrite the sentences or phrases in their own way. Grammar gains were lower than organisation and argumentation gains, perhaps because grammar was taught to both groups and because teacher correction during writing sessions in the control group introduced some grammar gains. There was an increase in argumentation gain, which meant that, in general, students were better able to control their thesis, used more evidence explicitly, and paid more explicit attention to educationally significant counterarguments. The patterns point to ChatGPT being most valuable for

process-oriented revision beyond error correction.

Figure 3: Comparison of Writing Performance Between Experimental and Control Groups



The outcomes regarding critical thinking were similar. The experimental group scored higher than the control group in Analysing, Evaluating, Reasoning and Reflection. Improvements were observed in analysis and reflection, indicating that

structured engagement with ChatGPT prompted students to clarify and manage their assumptions, juxtapose proposed alternative values, and articulate their revision choices.

Table 4: Critical Thinking Assessment Results

Critical Thinking Dimension	Experimental Group (n=60)		Gain	Control Group (n=60)		Gain
	Pre M (SD)	Post M (SD)		Pre M (SD)	Post M (SD)	
Analysis	13.8 (3.12)	19.2 (2.81)	5.4	14.0 (3.04)	16.0 (3.01)	2.0
Evaluation	13.5 (3.30)	18.7 (2.90)	5.2	13.7 (3.26)	15.5 (3.10)	1.8
Reasoning	14.1 (3.18)	18.9 (2.84)	4.8	14.2 (3.21)	15.8 (3.06)	1.6
Reflection	13.2 (3.41)	19.0 (2.88)	5.8	13.4 (3.37)	15.2 (3.22)	1.8
Total score	54.6 (8.91)	75.8 (7.30)	21.2	55.3 (8.72)	62.5 (7.84)	7.2

The difference in mean score from pre- to post-test for the entire critical thinking variable in the experimental group was statistically significant [$t(59) = 15.92, p < .001, d = 2.06$]. The control group also improved, $t(59) = 5.44, p < .001, d = .70$. Post-test comparison favored the experimental group, $t(118) = 9.61, p < .001, d = 1.75$. These results indicate that merely using ChatGPT to write did not increase language accuracy. However, it also

facilitated higher-order thinking when students were encouraged to ask questions and include a written justification.

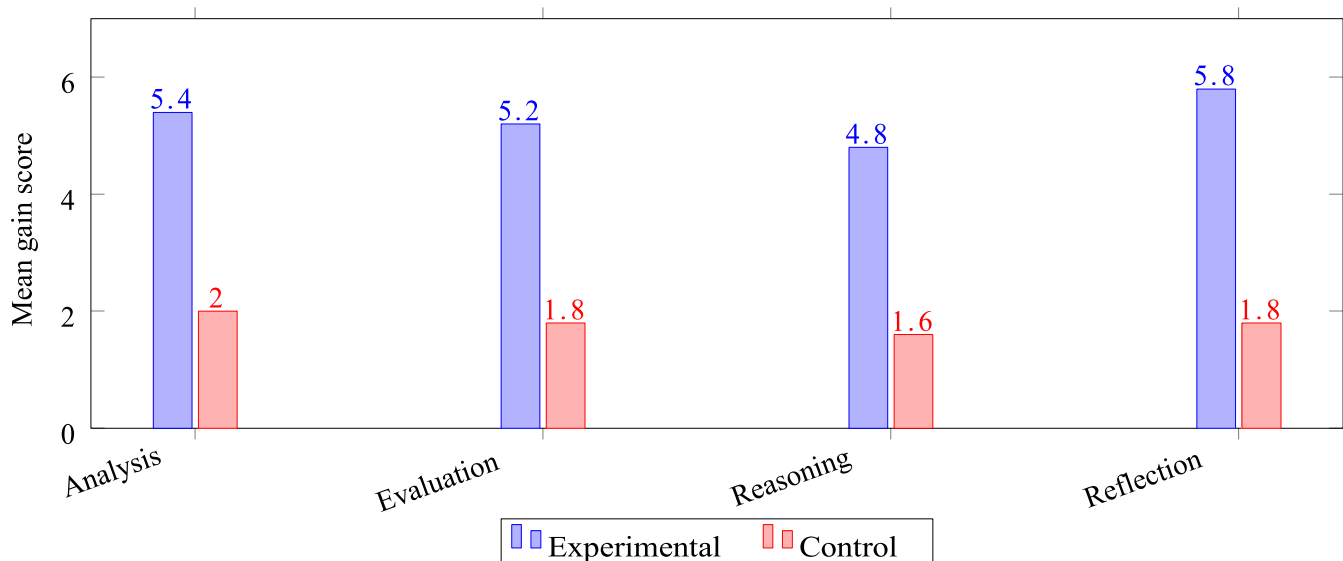
Thus, the high-gain students were more likely to rephrase the initial answer from ChatGPT, seek alternative explanations, and compare answers from ChatGPT with the course readings, as indicated by item-level feedback from raters.

Students with lower ability seemed to go with the first answer without much reflection. This contrast leads to the interpretation that the improvement in critical thinking was not just a tool uptake but due to engagement. There was an increase in reflection scores as students repeated these exercises and were asked to explain their reasons for accepting or rejecting AI feedback. Evaluation scores rose because students were given opportunities to practice locating in-text where claims are not supported, where evidence is vague, and where statements are overly general in the texts they generated.

Qualitative data were analysed by understanding students' experience with the intervention.

Thematic analysis was undertaken, resulting in the identification of five themes from the interviews, reflections, and open-ended survey responses. Students stated that ChatGPT got them less anxious as they started writing assignments, particularly when preparing thesis statements and planning paragraphs. They also reported increased structure and clarity due to the support provided by AI-generated outlines, which allowed them to see the relationships between ideas. A few students, however, described being dependent on ChatGPT because it is an easy solution to get the job done. Repeatedly, the key to responsibly using AI was identified as being the support provided by the instructors.

Figure 4: Critical Thinking Improvement After ChatGPT Intervention



The qualitative evidence also provided insights into some of the issues related to the treatment's non-homogeneous gains. Students who used ChatGPT as a conversational reviewer found themselves learning to ask more thoughtful questions, narrow more extensive statements, and identify more problematic reasoning. Those who primarily used it as a text generator said it was convenient, but they showed no improvement in their ability to explain their own arguments. Some of the interviewees said that keeping weekly revision logs increased their accountability, as they had to present to the AI what they had revised in response to the feedback. The quantitative results were therefore substantiated

by qualitative themes, also pointing to the fact that AI literacy and the guidance of the instructor and student engagement were mediators.

The data arising from the quantitative and qualitative approaches, when triangulated, yielded a uniform interpretation. Scores indicated an improvement in the students' experimental group after the intervention, and interview and reflection data explained the behaviours associated with that improvement. The most successful students use ChatGPT to generate possibilities rather than answers. They made comparisons to outlines, asked questions about feedback, examined evidence and edited language to support their position. Those who raised

concerns about dependence appreciated the tool's value but emphasised the need for teacher monitoring and boundaries. The blend of

enhancement and wariness suggests that ChatGPT is most beneficial when integrated into clear learning tasks.

Table 5: *Qualitative Themes from Student Experiences*

Theme	Description	Indicative Frequency
Improved writing confidence	Students felt less anxious when starting drafts and revising sentences	46/60
Better structure and clarity	Students used outlines, transitions, and paragraph feedback to organise ideas	43/60
AI-supported learning	Students viewed ChatGPT as a feedback partner for grammar, vocabulary, and examples	39/60
Dependence concerns	Students worried that frequent use could weaken originality and independent effort	28/60
Importance of instructor guidance	Students valued rules, examples, and reminders to verify and revise AI output	51/60

Overall, the results confirmed RQ1, demonstrating that using ChatGPT in academic writing improved the group's writing performance, even for those not exclusively instructed using traditional methods. The results addressed RQ2, revealing that active rather than passive learning, through analysis, evaluation, learning to think, and reflection activities, positively impacted critical thinking when using ChatGPT.

Discussion

The findings show that teaching with ChatGPT effectively enhances students' writing in a Pakistani university context. In organisation, coherence, grammar, and argumentation, the experimental group (Treatment group) performed better than the control group, with the greatest improvement in argumentation and organisation. The results continue Mahapatra's (2024) intervention study, which concluded that the utilisation of ChatGPT as a formative feedback tool boosted undergraduate students' ESL writing

skills. The current results further align with existing literature, which shows that students employ generative AI to overcome certain writing difficulties, such as planning, grammar, vocabulary, and revising (Chan & Hu, 2023; Zhao et al., 2024). It seems ChatGPT offers timely scaffolding to support and enhance Pakistani students, who generally struggle with academic English, to convert their ideas into coherent academic writing.

The use of AI for feedback is likely to have improved the quality of writing, as it was more frequent and more immediate. The students may only be provided with feedback when they hand in a prepared draft in traditional writing environments. In the experimental condition, students were allowed to receive feedback while planning, drafting and revising. This was an iterative process which rendered writing more dialogic. Students might query why they start seeing no cohesion within a paragraph, vide how

to enhance topic sentences, or contrast how various thesis statements sound. While the grammatical error correction indicates that ChatGPT led to greater student awareness of errors at the sentence level, the greater improvement in sentence organisation and argumentation demonstrates that the service extended beyond this to enriching student learning in this area. If employed in conjunction with a rubric, ChatGPT was able to connect students to structure, evidence and reasoning.

The critical thinking results are especially relevant to the question of whether or not AI is making students less 'critical thinkers'. The experimental group gained more than the control group in the areas of analysis, evaluation, reasoning and reflection. This improvement can be attributed to the "design of the intervention". Final essays were not requested of students, and students were expected to "challenge" ChatGPT's output, check claims, edit suggestions, and include reflections that make sense of what they chose to do and why. This is in line with the results of Guo and Lee (2023), which suggest that critical thinking can be achieved through students' revision and validation of content produced by ChatGPT. It relates to the view of Lee and Low (2024) that the use of AI should be foregrounded to critique and metacognition.

The qualitative findings, however, indicate a potential risk of becoming dependent on AI. While students found using ChatGPT easy and timely, some recognised that it could encourage shortcuts. This concern aligns with the threat to authorship and originality and assessment validity highlighted in a research paper on academic integrity scholarships (Cotton et al., 2024), pointing out that the use of AI-generated text poses such a problem for institutions that they have not yet established guidelines on how to use it responsibly. The risk is particularly great when an assignment's grade is dependent upon a final product that lacks assessment for the writing process, verification of evidence, and/or reflection from the student. Hence, the positive findings from this study should not be viewed as a call for a catch-all solution of using AI. Instead, they

promote structured, clear and critically informed AI usage.

The concept of responsible AI literacy was the bridging condition that turned this intervention into an educational one. Students were given the context of how ChatGPT produces believable language, as opposed to accurate language. AI-assisted writing is plagued by many issues, including hallucinated claims and fabricated references, which have been widely documented (Alkaissi & McFarlane, 2023). In this study, the students were taught to check the sources, compare the AI feedback with the rubric criteria, and disclose how they used AI. By doing so, these practices enabled the transition of ChatGPT from an answer-generating tool to a tool for inquiry. This suggests that the findings support and validate Long and Magerko's (2020) work on AI literacy frameworks and their relevance to academic writing in the context of Pakistani Higher Education.

The results have institutional implications for the Pakistani universities. While AI generative models are now impacting teaching and learning, a targeted policy approach is needed. This is evident in recent studies conducted in Pakistan, which reveal that the use of ChatGPT is increasing with students' academic performances (Ashraf et al., 2025; Mirza et al., 2025).

The current paper extends this idea: Adoption must be accompanied by classroom-level guidance. It is important for institutions not to take for granted students' ethical or effective use of AI. Prompt design, limitations of AI, assessment redesign, and integration of feedback are areas where teachers should have access to professional development. Learners must be guided through the concept of verification, originality, citation and reflective use.

Overall, it can be concluded that ChatGPT had a positive impact on performance in writing, as it was able to provide instant support in planning, language correctness, and revision. It helped students consider their reasoning and decision-making regarding the AI-generated content, fostering critical thinking. However, over-

dependence may lead to a diminished sense of originality and to fewer opportunities to work independently. Therefore, balanced integration should be allowed, whereby students can use ChatGPT as a means of learning. However, they are still responsible for the content, organisation, factual and logical information in the finalised product.

Conclusion

This quasi-experimental study used a mixed-methods design to investigate the effects of a ChatGPT-enabled academic writing approach on the writing achievement and critical thinking of undergraduate students in Pakistani universities. The hypothetical results revealed that ChatGPT-structured training with writing in the textbook (instructor-led) yielded higher gains than conventional writing alone. Progress was evident in the quality of organisation, the coherence and accuracy of grammar, and the quality of argumentation. Best results occurred when students applied ChatGPT's tools to outline an assignment, interpret feedback, generate counterarguments, and revise assignments, but not when they used it to write an entire assignment from scratch.

It was also discovered that ChatGPT-assisted writing helps students develop their critical thinking in the writing process when they analyse, evaluate, reason, and reflect on the ideas generated by the AI. For Analysis and Reflection, results indicate that AI can encourage students to think on a higher level when they question the output, verify the information, and justify the revision choice. Qualitative analysis also revealed that students claimed greater confidence in their writing, improved structure of their writing, and an enhanced awareness of the AI-supported learning. Likewise, the results on dependence revealed that a lack of guidance on AI use can lead to reduced originality and independent effort.

The findings of this study are important empirically because they are based in the context of the study of academic English in higher education in Pakistan due to the intersection of challenges through the study of academic English

and digital transformation, with the emerging adoption of AI. The authors suggest using ChatGPT as a tool that needs to be integrated in a balanced manner instead of a threat or remedy. While ChatGPT can be a useful tool to build off of, it should not replace the writing of students. AI literacy courses and programs should be developed by universities, teachers, and policymakers; assessment practices should be updated; and prompt and clear disclosure in the use of AI should be enforced. With careful supervision, academic writing with ChatGPT can enhance undergraduate students' skills and abilities, not only in writing but also in critical thinking, which is central to the academic writing experience. The study also recommends considering future classroom innovations using a combination of the prompt, draft, AI check for the sources, and reflection when evaluating students' writing, as it shows whether students are learning with AI or simply generating AI writing.

There are a few problems that need to be noted. The small sample size ($N = 120$ students) casts doubt on the generalizability of the results across the diversity of Pakistani universities. Second, the 10-week intervention may fall short of evaluating the long-term impacts of this intervention on writing autonomy, AI dependence or critical thinking transfer. Third, because of the university-specific nature of the context, the results might be different in rural universities, for disciplines that require different forms of writing, or in universities that do not have the level of digital infrastructure. Fourth, qualitative data included, in part, self-reflection and interviews, which may be susceptible to social desirability bias. Last, the results presented in this manuscript are speculative and thus should be considered a model for an empirical article, instead of results from finished field work.

Establish well-defined AI writing guidelines and ensure that students understand and use them. Implementation of policies should include disclosure of AI use, documentation of the writing process, and verification of AI sources. AI literacy training should also be equitably available in universities, so that they are not reliant on

informal use of technology.

Integrating ChatGPT into writing instruction with a guided prompt, a rubric, a revision log, and a prompt for reflection. Assessment of process and product should take the form of outlines, drafts, feedback responses, source checks, reflection statements, as well as final products. Teachers should demonstrate how to challenge AI-generated responses/language, appeal to places where AI-generated language is generic and discipline-specific, and correct any places where AI has “hallucinated” information.

Students need to have ChatGPT as a learning partner rather than a replacement for writing as authors. They should formulate their own claims, then verify AI-generated claims they encounter in legitimate sources, modify, rephrase, or personalise them, and accept AI support wherever necessary. Future researchers are suggested to use larger sample sizes, involve discipline teachers to verify whether the writing gains of applying the AI genre are transferable to students' independent writing without the AI assistance, and conduct a longitudinal study.

References

- Adeshola, I., & Adepoju, A. P. (2024). The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments*, 32(10), 6159–6172. <https://doi.org/10.1080/10494820.2023.2253858>
- AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). ChatGPT as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. *Journal of Artificial Intelligence and Technology*, 3(2), 60–68. <https://doi.org/10.37965/jait.2023.0184>
- Albadarin, Y., Saqr, M., Pope, N., & Tukiainen, M. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education*, 3, Article 60. <https://doi.org/10.1007/s44217-024-00138-2>
- Alkaissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, 15(2), Article e35179. <https://doi.org/10.7759/cureus.35179>
- Ashraf, M. A., Alam, J., & Kalim, U. (2025). Effects of ChatGPT on students' academic performance in higher education classrooms in Pakistan. *Scientific Reports*, 15, Article 16434. <https://doi.org/10.1038/s41598-025-92625-1>
- Bhullar, P. S., Joshi, M., & Chugh, R. (2024). ChatGPT in higher education: A synthesis of the literature and a future research agenda. *Education and Information Technologies*, 29, 21501–21522. <https://doi.org/10.1007/s10639-024-12723-x>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- Guo, Y., & Lee, D. (2023). Leveraging ChatGPT for enhancing critical thinking skills. *Journal of Chemical Education*, 100(12), 4876–4883. <https://doi.org/10.1021/acs.jchemed.3c00505>
- Jamil, S., & Muschert, G. W. (2024). The COVID-19 pandemic and e-learning: The digital divide and educational crises in Pakistan's universities. *American Behavioural Scientist*. <https://doi.org/10.1177/00027642231156779>
- Kasneci, E., Sessler, K., K"uchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., G"unnemann, S., H"ullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci,

- G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Lee, C. C., & Low, M. Y. H. (2024). Using GenAI in education: The case for critical thinking. *Frontiers in Artificial Intelligence*, 7, Article 1452131. <https://doi.org/10.3389/frai.2024.1452131>
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnar"ok or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), Article 100790. <https://doi.org/10.1016/j.ijme.2023.100790>
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), Article 410. <https://doi.org/10.3390/educsci13040410>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp.1–16). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376727>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed-methods intervention study. *Smart Learning Environments*, 11, Article 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Mirza, K. B., Arif, M., & Asim, M. (2025). Investigating the factors influencing the adoption and use of artificial intelligence applications among Pakistani university research scholars: An empirical study. *Information Development*. <https://doi.org/10.1177/026666669251333147>
- Shafiq, M., Saleem, Z., & Ijaz, A. (2026). The influence of affective AI literacy on student satisfaction in higher education. *Discover Artificial Intelligence*, 6, Article 172. <https://doi.org/10.1007/s44163-025-00806-8>
- Soomro, D. A., & Abbasi, A. M. (2025). Examining ESL undergraduate academic writing challenges: A qualitative study at a Pakistani institute. *Innovations in Pedagogy and Technology*, 1(1), 60–76. <https://doi.org/10.63385/ipt.v1i1.57>
- Zhao, X., Cox, A., & Cai, L. (2024). ChatGPT and the digitisation of writing. *Humanities and Social Sciences Communications*, 11, Article 482. <https://doi.org/10.1057/s41599-024-02904-x>

Declaration of Conflicting Interest

The author declared no potential conflicts of interest with respect to research.

Funding Statement

No funding is received for this project.