



Introducing Large Language Models in Communication and Public Relations Education: A Mixed-Methods Pilot Study

Sebastian Fitzek^{1,2} · Alina Bărgăoanu¹

Accepted: 8 April 2025

© The Author(s) 2025

Abstract

This mixed-methods pilot study investigates the potential of large language models (LLMs), specifically ChatGPT 3.5, to assist communication and public relations students in improving academic writing, focusing on clarity, conciseness, and coherence—skills aligned with cognitive load theory. An analysis of 60 abstracts (30 student-drafted abstracts and 30 AI-augmented abstracts) through qualitative and quantitative methods revealed modest enhancements. Two blinded expert evaluators identified statistically significant improvements in clarity and conciseness (Evaluator 1: $p=0.0024$; Evaluator 2: $p=0.0462$), although evaluator variability highlighted assessment subjectivity. Without a control group, causal claims are limited, and training sessions may have enhanced students' proficiency with ChatGPT, complicating attribution. The students perceived greater benefits (mean improvement=1.2) than the evaluators confirmed (mean=0.52), suggesting potential overconfidence. Risks, including overreliance on AI and threats to originality, underscore the need for ethical guidelines, such as critical evaluation training. Owing to its small, homogeneous sample size and short duration, this study lacks demographic diversity and longitudinal insight, thus necessitating broader, long-term research.

Keywords Large language models · ChatGPT · Academic writing · Higher education · AI in education · Ethical AI use

Abbreviations

AI	Artificial intelligence
CAQDAS	Computer-assisted Qualitative Data Analysis Software
CHERRIES	Checklist for Reporting Results of internet E-Surveys

✉ Sebastian Fitzek
sebastian.fitzek@comunicare.ro

¹ Faculty of Communication and Public Relations, National University of Political Studies and Public Administration, Bucharest, Romania

² Institute for Quality of Life Research, Romanian Academy, Bucharest, Romania

CPE	Certificate of Proficiency in English
LLMs	Large Language Models
NLP	Natural Language Processing

Introduction

Large language models (LLMs), such as ChatGPT, have recently emerged as powerful text-generation tools with potential implications for education (Lydia et al., 2023; Miguel et al., 2023; Montenegro-Rueda et al., 2023; Tan, 2023). Unlike intelligent tutoring systems or adaptive learning platforms, ChatGPT is a pattern-based generative model that relies on extensive training data rather than learner-specific instructions (Introducing ChatGPT, 2024; Montenegro-Rueda et al., 2023; Roll & Wylie, 2016). While some studies indicate that LLMs can enhance certain writing tasks—offering rapid linguistic refinements and potentially supporting iterative improvement (Imran & Almusharraf, 2023; Marzuki et al., 2023)—their effectiveness in fostering deeper learning outcomes, such as critical thinking or argument development, remains less clear (Li et al., 2024; Shibani et al., 2024). Current evidence often focuses on surface-level text features, leaving a research gap regarding LLMs’ broader impact on academic writing quality and their role in shaping students’ engagement with the writing process (Kouam & Muchowe, 2024; Li et al., 2024; Shibani et al., 2024). Moreover, research typically lacks explicit attention to demographic and linguistic diversity, further constraining generalizability (Li et al., 2024).

To address this gap, the present pilot study investigates how revisions using ChatGPT 3.5—a widely accessible LLM at the study’s outset (January 2024)—influence academic writing quality in higher education communication and public relation programs. In communication and public relations, students often struggle to distill complex ideas into concise, audience-focused narratives while upholding professional writing standards. ChatGPT’s capacity to process and rephrase extensive data offers distinct support for these tasks. We focus on three widely recognized attributes of high-quality academic abstracts—clarity, conciseness, and coherence (Bayat, 2014; Fischer, 2013)—because they represent foundational attributes of effective academic communication, directly relating to cognitive load theory (Sweller, 1988), which posits that learning efficiency improves when unnecessary cognitive load is minimized. By reducing extraneous cognitive effort, LLMs may enable students to focus on higher-order skills such as argumentation. Although recent research suggests that LLMs can improve such features (Imran & Almusharraf, 2023; Marzuki et al., 2023; Montenegro-Rueda et al., 2023), questions remain about authenticity, intellectual ownership, and the balance between AI-generated enhancements and a student’s own voice (Castro, 2023; Cotton et al., 2024; Ippolito et al., 2022). Despite their advantages, LLMs carry risks, including overreliance on AI, potentially stifling critical thinking, and the dilution of individual writing voice—issues this study explores alongside its benefits.

Our approach draws on two theoretical frameworks to situate the integration of AI tools in writing instruction. The cognitive theory of multimedia learning (Mayer, 2005), which asserts that well-designed multimedia environments enhance learning

by effectively managing multiple information sources, posits that combining multiple sources of input can enhance learning if managed properly. Moreover, cognitive load theory highlights that reducing extraneous cognitive effort (e.g., basic language polishing) may free cognitive resources for higher-order thinking tasks (Sweller, 1988). By automating certain low-level writing tasks, LLMs might allow students to focus more on content and critical analysis, thereby supporting reconstructive methods in the social sciences with the aid of artificial intelligence (Dergaa et al., 2023; Elder & Paul, 2020; Jeon & Lee, 2023; Schaffler et al., 2022; Кочубей, 2021). These theories not only provide a conceptual foundation for our study but also directly inform our methodological choices. For example, cognitive load theory has guided our decision to implement a mixed-methods approach, combining computational evaluations with qualitative feedback, to assess both the reduction in cognitive load and the enhancement of higher-order writing skills.

Although prior work by Habib et al. (2024) examined AI's role in nurturing creativity, our study focuses on whether these tools can support essential writing competencies without undermining the learner's confidence in their original contributions.

Additionally, while academic writing is fundamental across disciplines, many students struggle to meet university-level standards in structure, coherence, and engagement with evidence (Esman et al., 2021; Kellogg & Whiteford, 2009; Mayer et al., 2023). The incorporation of LLMs may help address such challenges by offering immediate suggestions that improve text organization and reduce writing anxiety (Кочубей, 2021); however, this must be done ethically and responsibly (Azer, 2017; Castro, 2023; Cotton et al., 2024). Given that writing quality is subjective and context dependent (Charney, 1984) and that students' readiness to critically evaluate AI outputs varies (Ippolito et al., 2022), empirically testing these interventions in realistic academic tasks is essential.

Our mixed-methods study combines “computational evaluations”—expert rubric-based assessments of original and AI-revised abstracts—with student perceptions gathered via surveys. Here, “computational evaluations” refer to standardized, systematic ratings by qualified evaluators who apply uniform criteria to gauge improvements in clarity, conciseness, and coherence. This standardized evaluation approach aligns with our theoretical framework by providing a systematic measurement of writing improvements while maintaining consistent assessment criteria. We utilized prompt-based learning techniques (Mayer et al., 2023) to instruct ChatGPT to refine drafts for linguistic clarity and structural organization but not to generate content independently, thus testing ChatGPT's potential to improve writing features rather than fully produce academic text. Given our small, volunteer-based sample, we position this research as exploratory. Nevertheless, by focusing on a well-defined academic genre—scientific abstracts—we aim to offer preliminary insights into how LLM assistance might influence fundamental writing quality attributes and inform pedagogical practices (Elder & Paul, 2020; Roll & Wylie, 2016).

In summary, this pilot study examines how LLM-driven revisions affect the academic writing quality of communication students, investigating both measurable improvements and students' subjective experiences. While ChatGPT's capacity to refine drafts may hold promise, further research is needed to determine how such

tools can best support genuine skill development, maintain academic integrity, and ultimately contribute to more effective and ethically grounded writing instruction.

Method

Study Design

This pilot study employed a single-center, single-arm, nonblinded cohort design with a mixed-methods approach to explore the impact of large language models (LLMs), specifically ChatGPT, on the quality of academic writing among communication and public relations students. The design was chosen to examine both the immediate effects of AI assistance on writing quality and the broader implications for learning and skill development, addressing a key gap in current AI education research (Plano Clark, 2017; Tashakkori & Teddlie, 2010). The study was conducted from 10 January to 5 February 2024, enabling the assessment of both performance improvements and learning outcomes through qualitative and quantitative methods. Despite the lack of a control group, this study provides exploratory data on AI-assisted writing, laying a foundation for hypothesis generation and future controlled studies.

Participants

The study involved 30 volunteer students from the Faculty of Communication and Public Relations at the National University of Political Studies and Public Administration (SNSPA), encompassing both the bachelor's and master's levels. All participants were enrolled in communication and public relations courses and had previously completed an introductory data science course focused on the fundamentals of using ChatGPT 3.5, which was verified through English proficiency certificates (e.g., CPE). This foundational exposure ensured baseline familiarity with LLM capabilities and limitations and that both original and revised abstracts were created and evaluated under conditions minimizing language-related variability. The subsequent, more intensive data science course—during which this study took place—emphasized ethical considerations, prompted optimization, and critical evaluation of AI outputs. Participation was voluntary, and informed consent was obtained after a briefing on the study's objectives, procedures, and the option to withdraw at any point without penalty. Owing to the homogeneous demographic and linguistic background of the participants, the findings may not be fully generalizable beyond this context. Although the sample size ($n=30$) is modest, it is sufficient for a pilot study aimed at generating preliminary insights and refining methodologies for future research. This allows for detailed qualitative analysis and initial statistical trends, balancing feasibility and exploratory depth.

Evaluation Framework

The study utilized a comprehensive framework, informed by mixed methods genre analysis (Hashemi & Gohari Moghaddam, 2019), to assess three dimensions of writ-

ing quality: clarity (logical flow and effective communication), conciseness (economy of language), and coherence (structural unity). Expert evaluators employed a standardized five-point scale complemented by qualitative feedback, documenting specific improvements in each dimension. The student reflection logs provided additional data on tool usage patterns and decision-making processes, enabling analysis of both writing outcomes and learning development.

A rigorous calibration process was integrated into the evaluation framework to ensure consistency and reliability. This included an intensive workshop where evaluators analyzed sample abstracts together and discussed their interpretations of clarity, conciseness, and coherence criteria. Calibration continued through practice evaluation sessions using test abstracts not included in the study sample, enabling evaluators to align their assessment approaches and establish consistent scoring patterns. The formal assessment began only after strong interrater agreement was achieved (Cohen's $\kappa = 0.666$ for overall interrater reliability). Future studies should complement expert evaluations with automated textual analysis tools (e.g., readability indices, lexical diversity measures) to increase objectivity and reduce evaluator subjectivity.

Procedure

To ensure transparency and maintain academic integrity, the students received clear instructions on ChatGPT usage. The participants were informed explicitly that ChatGPT 3.5 is a generative AI tool that provides suggestions on the basis of patterns learned from training data and does not serve as an adaptive tutor or authoritative academic source. ChatGPT 3.5 was used exclusively to enhance linguistic clarity, conciseness, and coherence of the student-written abstracts; the AI did not autonomously generate original content. Students revised their initial drafts via ChatGPT prompts specifically designed to improve readability and organization.

Phase 1: Initial Independent Abstract Writing

The participants were provided with a deidentified academic article ("The role of artificial intelligence (AI) in public relations and product marketing in modern organizations" by Alawaad, 2021), stripped of title, abstract, and authorship details (Alawaad, 2021). They were instructed to write an abstract independently, without any AI assistance, within a fixed time window of one academic hour. This activity was conducted as a monitored in-class task to ensure compliance with the instructions.

Phase 2: ChatGPT Training

The participants attended three structured training sessions (each lasting 1.5 h) focused on the effective and ethical use of ChatGPT 3.5. These sessions defined ethical AI usage, emphasizing proper attribution, avoidance of verbatim copying, and the importance of preserving academic integrity. Students received written guidelines and example scenarios illustrating acceptable and unacceptable practices. The training encompassed clarifications on what constitutes acceptable assistance and

guidelines to maintain originality, alongside explicit warnings against plagiarism. Additionally, participants were shown how to craft effective prompts to elicit useful and context-relevant suggestions from ChatGPT and were trained to interpret, verify, and refine these suggestions critically. The sessions also emphasized the importance of cross-checking AI-proposed text against the original source and their own understanding to prevent misinformation or irrelevant additions. Students were introduced to “responsible AI usage” principles, including verifying factual claims presented by ChatGPT, preserving their personal writing style, and ensuring that the final text represents their own scholarly voice rather than a direct copy-paste of AI-generated content.

GPT-3.5 was selected owing to its widespread availability and established use in educational contexts at the time of the study (January 2024), ensuring relevance to institutions with similar resource constraints and facilitating replicability. The training sessions may introduce learning effects, potentially improving students’ proficiency with ChatGPT rather than solely reflecting tool efficacy. Future studies could incorporate a longitudinal design or control group to distinguish AI assistance from user skill development.

Phase 3: ChatGPT-assisted Revision

The participants used ChatGPT 3.5 to refine their initial abstracts following a structured revision protocol. The process progressed from general refinement requests to specific improvements, with participants documenting their interactions and decision-making rationale throughout. Under light supervision, participants maintained detailed logs of their prompts and responses, submitting brief reflective statements about their integration process to ensure originality. For example, one participant’s text evolved from ‘The paper discusses AI tools in marketing but lacks detail on implementation’ to ‘This study explores how AI-driven tools integrate into marketing strategies, offering insights into their role in product promotion and consumer engagement.’ This example demonstrates how students maintained their interpretive voice while incorporating AI suggestions. The revision process emphasized critical evaluation via three criteria: accuracy of content interpretation, preservation of intended meaning, and maintenance of academic voice. This structured approach ensured the thoughtful integration of AI assistance rather than the wholesale adoption of generated content.

Phase 4: Blind Expert Evaluation

Two independent experts in communication sciences and data science—unaware whose abstracts were original and were AI assisted—rated 60 abstracts (30 original and 30 revised). The scoring rubric (1 to 5) assesses clarity, conciseness, and coherence. A brief calibration session was conducted with the experts prior to evaluation to ensure consistent application of the rubric. “Computational evaluations” refer to these standardized, systematic ratings by qualified evaluators who applied uniform criteria to gauge improvements in clarity, conciseness, and coherence. To acknowledge overlapping content between versions, the focus was placed on identifying improvements

in text organization, expression, and logical flow. The goal was to isolate the effect of ChatGPT assistance on text quality rather than on content originality per se.

Survey and Validation

The participants completed an online survey via Google Forms (Appendix 2), which captured their perceptions of differences between self-written abstracts and ChatGPT-assisted abstracts (Workspace, n.d.). The survey included Likert-scale questions and open-ended items, allowing students to express views on efficiency, clarity, and challenges in maintaining academic standards. The survey design followed a structured progression, moving from general perceptions of AI assistance to specific experiences with the revision process. Open-ended questions were strategically placed to capture detailed reflections on the integration of AI suggestions and perceived changes in the writing approach. Prior to full deployment, the survey underwent pilot testing, expert review, and cognitive interviews to ensure item clarity and relevance (Eysenbach, 2004). A Cronbach's alpha of ~ 0.847 indicated strong internal consistency. The validation process included multiple iterations of refinement on the basis of pilot participant feedback and expert reviewer suggestions, ensuring that the questions effectively captured both the technical and pedagogical aspects of AI-assisted writing. This rigorous validation process aimed to reduce measurement error and ensure that the survey results accurately reflected students' experiences and perceptions.

Ethical Considerations

This study adhered to strict ethical guidelines. The participants provided informed consent, were assured of confidentiality, and were free to withdraw without any academic repercussions. They were explicitly informed that they remained fully responsible for the authenticity and originality of their final submissions, despite using ChatGPT. The ethical framework specifically addressed the unique challenges of AI integration in academic writing, emphasizing the distinction between appropriate AI assistance and academic misconduct. The participants received detailed guidelines outlining the acceptable use of AI-generated suggestions, including specific examples of proper and improper implementation. No personally identifying data were collected or stored. Additionally, participants were reminded that ChatGPT's suggestions should be critically evaluated and not blindly accepted, reinforcing the principle that AI is a supplemental tool rather than a definitive authority.

The ethical protocol also included ongoing monitoring mechanisms throughout the study. Instructors maintained observation logs during supervised sessions, documenting any concerns or questions about AI usage that arose. These observations provide real-time guidance and help refine the ethical guidelines for future research. To further emphasize academic integrity, participants were advised to employ plagiarism-checking tools and to properly attribute any nonoriginal content. To ensure adherence to ethical guidelines, the study incorporated measures such as supervised revision sessions and periodic reviews of participant logs to monitor compliance with responsible AI usage principles. While these measures were not quantifiably enforced during this pilot, future studies might implement systematic originality verification

methods to ensure that responsible AI use is maintained at all stages of the academic writing process.

Results

This section presents the findings of our mixed-methods pilot study, detailing participant characteristics, statistical analyses, dimension-specific results integrated with interrater reliability, comparisons between student perceptions and expert evaluations, and a summary of the key outcomes.

Participant Characteristics and Context

This study involved thirty volunteer students from the Faculty of Communication and Public Relations (SNSPA). The participants' ages ranged from 20 to 47 years (mean = 23.47, SD = 6.27), indicating a relatively diverse age distribution. Among the 30 participants, 18 were female and 12 were male, indicating varied gender representations. Academic enrollment included 21 Bachelor's and 9 Master's level students. Further analysis revealed that Master's students (mean age = 27.3 years) had more prior writing experience than Bachelor's students did (mean age = 21.8 years), providing an opportunity to examine how academic experience might influence AI tool utilization. Owing to the homogeneous demographic and linguistic background of the participants, the findings may not be fully generalizable beyond this Romanian institutional context.

Statistical Analysis

Our analysis employed both parametric and nonparametric tests to assess improvements in writing quality, with a focus on key findings. Paired *t* tests comparing original and ChatGPT-assisted abstracts revealed significant improvements (Evaluator 1: $p = 0.0024$; Evaluator 2: $p = 0.0462$). Cohen's kappa (0.666) indicated substantial interrater reliability, with separate kappa values for text origin identification (0.68) and quality rating (0.65).

Key Findings

Writing quality improvements varied systematically with initial proficiency level ($r = -0.42$, $p < 0.05$), indicating that students with lower initial proficiency experienced greater enhancements. Clarity showed the strongest enhancement, with a mean improvement of 0.57 points for Evaluator 1 and 0.48 points for Evaluator 2. Conciseness demonstrated moderate but significant gains, particularly among master's-level students, who showed a mean improvement of 0.58 points compared with 0.43 points for bachelor's-level students. Coherence improvements, while present, exhibited greater variability between evaluators, with Evaluator 1 noting a mean improvement of 0.50 points and Evaluator 2 noting 0.37 points, suggesting that this dimension may be more resistant to AI-assisted enhancement.

Table 1 Descriptive statistics of the writing quality assessments (scores after ChatGPT assistance; $n=30$)

Statistic	Evaluator 1	Evaluator 2
Mean Score	3.15	3.20
Median Score	3.23	3.22
Standard Deviation	0.87	0.96
Variance	0.74	0.84
T-Statistic	-3.32	-2.08
P Value	0.0024	0.0462

Note: Mean scores represent the quality ratings after ChatGPT assistance, with paired t tests comparing these against original abstracts. Both evaluators reported statistically significant improvements ($p<0.05$), although the magnitude of improvement varied

Table 2 Descriptive statistics by dimension ($n=30$)

Dimension	Evaluator	Original Mean (SD)	Revised Mean (SD)
Clarity	Evaluator 1	2.85 (0.82)	3.42 (0.91)
Clarity	Evaluator 2	2.90 (0.88)	3.38 (0.94)
Conciseness	Evaluator 1	2.78 (0.85)	3.35 (0.89)
Conciseness	Evaluator 2	2.82 (0.90)	3.29 (0.92)
Coherence	Evaluator 1	2.81 (0.84)	3.31 (0.87)
Coherence	Evaluator 2	2.88 (0.92)	3.25 (0.95)

Note: Scores show consistent improvement across all dimensions after ChatGPT assistance, with clarity showing the highest gains

Dimension-specific Results and Interrater Reliability

The dimension-specific analyses revealed several key patterns. Clarity improvements were most pronounced and consistent across both evaluators, with Evaluator 1 observing a mean improvement of 0.57 points and Evaluator 2 observing 0.48 points. Conciseness demonstrated moderate but significant gains, particularly among master's-level students, who showed a mean improvement of 0.58 points compared with 0.43 points for bachelor's-level students. Coherence improvements, while present, exhibited greater variability between evaluators, with Evaluator 1 noting a mean improvement of 0.50 points and Evaluator 2 noting 0.37 points. This suggests that coherence may be more resistant to AI-assisted enhancement (See Tables 1 and 2).

Cohen's kappa (0.666) indicated substantial agreement between the two independent evaluators, confirming the reliability of the assessments. Separate kappa values for text origin identification (0.68) and quality rating (0.65) further corroborated this reliability. Additional analysis of the evaluator comments revealed high agreement on the technical aspects of writing improvement ($\kappa=0.72$) but lower agreement on stylistic elements ($\kappa=0.58$), suggesting that certain aspects of writing quality remain inherently subjective even with standardized evaluation criteria. The variation in effect sizes (Cohen's $d=0.693$ for R1 vs. 0.426 for R2) highlights that individual evaluator perspectives can influence the assessment of writing improvements.

Visual Representations of Findings

While these visuals demonstrate a generally positive shift in writing quality, the spread of data points and the presence of error bars indicate that improvements were not uniform across all participants (See Fig. 1). Specifically, the error bars in Fig. 2 illustrate the variability in the evaluators' ratings, highlighting that while clarity and conciseness consistently improved, coherence fluctuated more.

Student-evaluator Comparisons

Comparing Student Perceptions with Expert Evaluations

Integrating qualitative feedback from surveys with quantitative data revealed a nuanced picture. The students generally perceived greater enhancements in their writing (mean perceived improvement = 1.2 points) than the evaluators' ratings reflected (mean evaluated improvement = 0.52 points). The discrepancy between students' perceived AI benefits and evaluator ratings suggests potential student overconfidence, possibly fostering an undue reliance on AI-generated content. This highlights a need for training students in critical evaluation and self-assessment skills. This discrepancy varied by writing dimension, with the closest alignment between student and evaluator assessments occurring in clarity improvements and the largest gaps appearing in coherence evaluations.

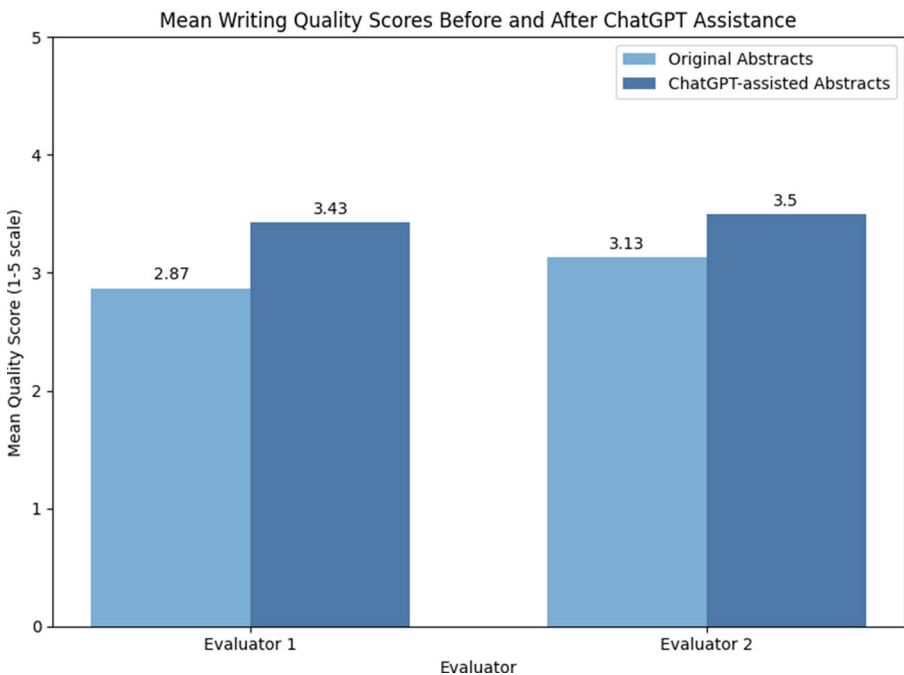


Fig. 1 Representation of perceived improvements in clarity and conciseness

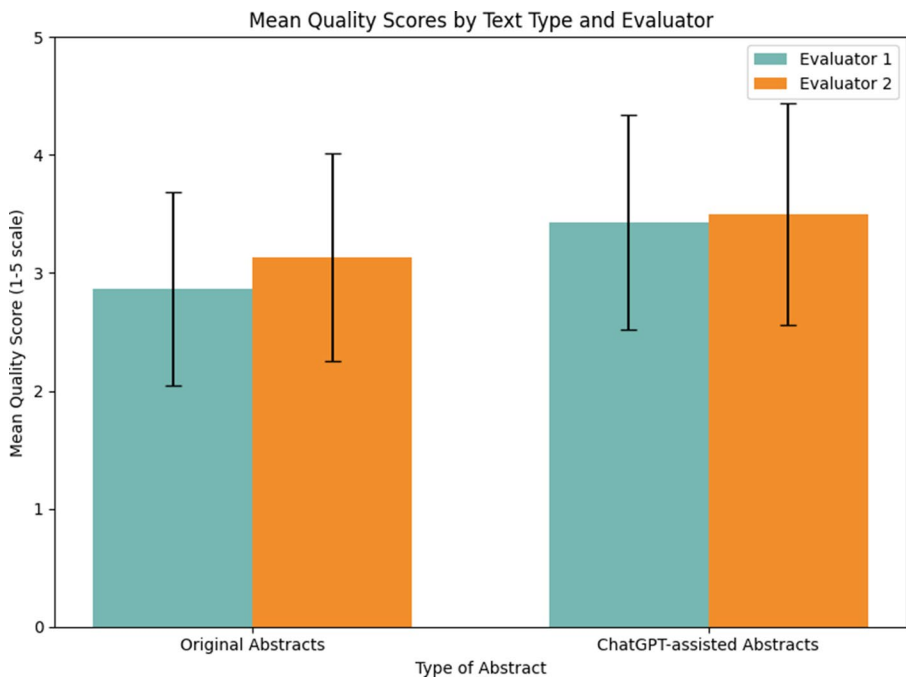


Fig. 2 Mean quality scores by text type and evaluator with error bars

Alignment with Student Feedback

The student response patterns revealed distinct differences in AI tool utilization across academic levels. Bachelor's students (78%) employed ChatGPT primarily for structural improvements and organizational clarity, whereas Master's students (65%) focused on stylistic refinements and nuanced language enhancement. A temporal evolution in usage patterns emerged, progressing from basic corrections to sophisticated applications, such as testing argument clarity from multiple perspectives. For example, one Master's student stated, "ChatGPT helped me organize my thoughts and present them more clearly," demonstrating how the tool supported both structural and stylistic improvements while maintaining academic voice. Additionally, another student remarked that ChatGPT allowed them to explore alternative phrasings that enriched their arguments without diluting their original intent.

The interaction with AI evolved from initial surface-level engagement to more strategic usage. Students reported increasingly sophisticated applications, developing from basic corrections to testing argument clarity from multiple angles. This progression included exploring alternative phrasings while maintaining original meaning and using AI suggestions as catalysts for further refinement rather than direct solutions. This evolution in tool usage suggests a developing capacity for critical engagement with AI-generated content, particularly among more experienced writers.

Summary and Preliminary Conclusions

Overall, the results indicate that ChatGPT can improve certain elements of academic writing, particularly clarity and conciseness, under controlled conditions. Analysis revealed distinct patterns of improvement across writing quality dimensions, with clarity showing the most consistent gains (mean improvement=0.53 points) compared with more variable coherence enhancement (mean improvement=0.44 points), suggesting differential effectiveness of AI assistance across writing aspects.

Our findings revealed significant correlations between student characteristics and improvement patterns. The effectiveness of AI assistance varies notably with academic experience and initial writing proficiency. Master's level students demonstrated more strategic and selective use of AI suggestions, whereas bachelor's level students derived greater benefit from structural improvement recommendations. This variation suggests that AI assistance in academic writing requires tailored implementation strategies on the basis of user experience and proficiency levels.

The notable discrepancy between student perceptions and evaluator assessments may suggest positive engagement with the writing process. The students' heightened awareness of potential improvements could indicate increased metacognitive engagement with their writing development.

The observed enhancements, while promising, should be viewed as indicative rather than definitive, given this pilot study's limited scope and sample size. These preliminary findings establish a foundation for future research that should examine not only quantifiable improvements in writing quality but also the complex dynamics of student-AI interactions in academic writing development. Understanding these dynamics will be crucial for developing effective pedagogical strategies for AI integration in writing instruction.

Discussion

This pilot study provides timely insights into the potential of large language models (LLMs) in enhancing academic writing quality in higher education. Our mixed-methods approach revealed both quantitative improvements in clarity and conciseness and qualitative insights into the benefits and challenges of AI integration in academic writing (Imran & Almusharraf, 2023; Marzuki et al., 2023).

Our findings align with and extend current pedagogical frameworks, particularly the cognitive theory of multimedia learning (Mayer, 2005) and the cognitive load theory (Sweller, 1988). These frameworks suggest that AI assistance may fundamentally alter how students approach writing by reducing their cognitive load during basic tasks, enabling deeper engagement with higher-order aspects of academic writing. The systematic variation in writing quality improvement across proficiency levels supports Montenegro-Rueda et al.'s (2023) assertion that AI tools require differentiated implementation strategies. Specifically, bachelor's-level students benefited the most from structured scaffolding approaches, whereas master's-level students demonstrated greater success with autonomous AI interaction strategies. The statistically significant improvements observed by both evaluators reinforce prior research

suggesting that, when used judiciously, LLMs can enhance certain text quality attributes, such as clarity and conciseness (Montenegro-Rueda et al., 2023). However, developing higher-order writing skills, such as argumentation and critical analysis, still requires traditional pedagogical interventions, aligning with findings from Imran and Almusharraf (2023). Recent research by Marzuki et al. (2023) supports this interpretation, demonstrating how a reduced cognitive load in the mechanical aspects of writing can enhance the focus on content development and argumentation.

Students' tendency to overestimate AI-generated content may undermine critical self-assessment, highlighting a pedagogical concern warranting further exploration. Therefore, instructors should explicitly address these perceptions in training sessions. The practical implementation of these findings points toward an integrated approach to AI in academic writing instruction. Effective implementation begins with comprehensive training in the critical evaluation of AI suggestions, followed by guided practice with differentiated support, and culminates in the development of autonomous AI utilization strategies. Using LLMs introduces risks such as overdependence, which may diminish critical thinking, encourage uniform writing styles, and spark ethical debates about academic integrity. This progression addresses the gap identified by Cotton et al. (2023) between technological capability and pedagogical effectiveness.

Instructors should integrate AI tools gradually, initially using low-stakes assignments (e.g., draft revisions), and provide explicit instruction on critically evaluating AI suggestions. Workshops could teach cross-checking AI outputs with original sources, ensuring alignment with scholarly norms. Furthermore, writing instructors should develop clear guidelines for AI tool usage that account for student proficiency levels and learning objectives. These guidelines emphasize critical engagement with AI suggestions rather than passive acceptance, aligning with Montenegro-Rueda et al.'s (2023) recommendations for responsible AI integration in education. Additionally, assessment criteria should be adapted to acknowledge the role of AI assistance while maintaining high standards for original thinking and academic integrity.

Students often perceived greater enhancements than evaluators recognized, underscoring the importance of aligning educational tools with clearly articulated academic standards and evaluation criteria. This perception gap may reflect the immediate cognitive benefits students experience when using AI assistance, such as reduced mental effort in sentence construction and organization. These observations align with Kellogg and Whiteford's (2009) findings about the development of advanced writing skills, suggesting that AI tools could serve as scaffolding in the progression toward writing expertise (Kouam & Muchowe, 2024).

The study's qualitative insights revealed that beyond improved clarity, students valued ChatGPT for its efficiency and capacity to spark new ideas. The relationship between AI assistance and creativity emerged as particularly nuanced, with students reporting that AI suggestions often prompted them to consider alternative perspectives and phrasings they might not have otherwise explored. This finding extends Habib et al.'s (2024) work on AI's role in fostering creativity while also addressing concerns raised by Cotton et al. (2023) about maintaining academic integrity.

In essence, this pilot study presents a cautiously optimistic view of LLM integration in academic writing instruction. The observed improvements in writing quality, while significant, appear to be mediated by factors such as student expertise,

task type, and implementation strategy. This suggests that successful AI integration requires a nuanced understanding of both the technology's capabilities and the pedagogical context in which it is deployed. As emphasized by Jeon and Lee (2023), along with Lee and Lee (2023) and Marzuki et al. (2023), developing standardized rubrics, fostering academic honesty, and maintaining inclusivity will be crucial as we continue to refine our understanding of AI's role in academia.

Study Limitations

While our study provides valuable insights into AI-assisted academic writing, several important limitations warrant consideration. The sample size ($n=30$) and focus on communicating with students at a single institution limit the generalizability of our findings. Additionally, volunteer-based recruitment likely attracted participants with a preexisting interest in AI technology, potentially introducing selection bias that may have influenced the outcomes. The study's duration, while sufficient for observing immediate effects, does not address long-term learning impacts or sustained writing improvement.

Methodological constraints also merit acknowledgment. The absence of a control group makes it difficult to isolate the specific effects of AI assistance from other factors that might influence writing improvement. The absence of demographic and linguistic diversity limits our ability to generalize findings across various educational contexts. Furthermore, while our evaluation criteria were carefully developed, the inherent subjectivity in assessing writing quality introduces unavoidable variability in our measurements.

This study did not investigate the influence of demographic factors such as native language or prior writing experience on AI effectiveness. The short study duration also prevents the assessment of long-term impacts on writing skill development. Future research should explore these variables for equitable outcomes.

Additionally, while our theoretical framework combining cognitive load theory and multimedia learning theory provides valuable insights, future studies might benefit from exploring additional theoretical perspectives that could better explain the interaction between students and AI tools in academic writing.

Conclusions

Our investigation revealed that LLMs, particularly ChatGPT, can offer notable improvements in specific aspects of academic writing quality, such as clarity and conciseness. These enhancements varied across writing dimensions and student populations, indicating that the impact of AI on academic writing is influenced by factors such as student expertise, task complexity, and implementation methods. The observed discrepancy between student perceptions and expert evaluations provides valuable insights for educational practice. While students reported substantial benefits from AI assistance, expert assessments indicated more moderate improvements, highlighting the need for refined evaluation frameworks capable of measuring both the immediate and long-term impacts of AI integration in academic writing.

Future Directions

Future research should expand investigations across diverse academic disciplines and student populations to increase the generalizability of the findings. Future research should incorporate demographic and linguistic diversity to ensure equitable outcomes. Longitudinal studies are essential to determine whether AI assistance sustains writing skill development or merely enhances short-term revisions. The development of comprehensive pedagogical frameworks for AI integration, addressing both the maintenance of academic integrity and the optimization of learning outcomes, is essential. This study demonstrates the need to integrate AI tools into educational practice by adhering to strict ethical standards and informed consent procedures. As AI capabilities continue to evolve, ongoing collaboration among educators, students, and policymakers is necessary to ensure that AI integration supports core educational objectives while maintaining academic standards. This balanced approach will guide the development of AI applications in higher education. Additionally, future research could investigate how different pedagogical methods mediate the effectiveness of AI assistance.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40593-025-00477-7>.

Author Contributions S.F. (Sebastian Fitzek) conceptualized the study, designed the research methodology, and conducted the statistical analysis. A.B. (Alina Bărgăoanu) contributed to the literature review, supervised the data collection process, and provided critical revisions to the manuscript. S.F. wrote the main manuscript text, and A.B. assisted in refining the analysis and discussion sections. Both authors reviewed and approved the final manuscript.

Funding This research utilized internal university resources. The authors were involved in all the study aspects, from conceptualization to manuscript preparation.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing Interests The authors declare no competing interests.

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/>.

References

- Alawaad, H. (2021). The role of artificial intelligence (AI) in public relations and product marketing in modern organizations. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12, 3180–3187.
- Azer, S. A. (2017). Social media channels in health care research and rising ethical issues. *AMA Journal of Ethics*, 19(11), 1061–1069. <https://journalofethics.ama-assn.org/article/social-media-channels-health-care-research-and-rising-ethical-issues/2017-11>
- Bayat, N. (2014). The effect of the process writing approach on writing success and anxiety. *Educational Sciences: Theory and Practice*, 14(3), 1133–1141. <https://eric.ed.gov/?id=EJ1034097>
- Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English*, 18(1), 65–81. <https://doi.org/10.58680/rte198415687>
- 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>
- de Castro, C. A. (2023). A discussion about the impact of ChatGPT in education: Benefits and concerns. *Journal of Business Theory and Practice*, 11(2). <https://doi.org/10.22158/jbtp.v11n2p28>. Article 2.
- Dergaa, I., Chamari, K., Zmijewski, P., & Saad, H. B. (2023). From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 40(2), 615–622. <https://doi.org/10.5114/biolsport.2023.125623>
- Elder, L., & Paul, R. (2020). *Critical thinking: Tools for taking charge of your learning and your life*. Rowman & Littlefield.
- Esman, E. N., Madrigal, D. V., & Mascuñana, C. G. (2021). Social media exposure and english writing proficiency of grade 11 students in a Philippine public high school. *Technium Social Sciences Journal*, 20, 212–232. <https://doi.org/10.47577/tssj.v20i1.3001>
- Eysenbach, G. (2004). Improving the quality of web surveys: The checklist for reporting results of internet E-Surveys (CHERRIES). *Journal of Medical Internet Research*, 6(3), e132. <https://doi.org/10.2196/jmir.6.3.e34>
- Fischer, J. D. (2013). A contemporary take on strunk and white for legal writers. *Scribes Journal of Legal Writing*, 15, 127. <https://heinonline.org/HOL/Page?handle=hein.journals/scrib15&id=133÷=&collection=>
- Hashemi, M. R., & Gohari Moghaddam, I. (2019). A mixed methods genre analysis of the discussion section of MMR articles in applied linguistics. *Journal of Mixed Methods Research*, 13(2), 242–260. <https://doi.org/10.1177/1558689816674626>
- Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology*, 15(4), ep464. <https://doi.org/10.30935/cedtech/13605>
- Introducing ChatGPT (2024). March 13). <https://openai.com/index/chatgpt/>
- Ippolito, D., Yuan, A., Coenen, A., & Burnam, S. (2022). Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers (No. arXiv:2211.05030). arXiv. <https://doi.org/10.48550/arXiv.2211.05030>
- Jeon, J., & Lee, S. (2023). Large Language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28(12), 15873–15892. <https://doi.org/10.1007/s10639-023-11834-1>
- Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist*, 44(4), 250–266. <https://doi.org/10.1080/00461520903213600>
- Kouam, A. W. F., & Muchowe, R. M. (2024). Exploring graduate students' perception and adoption of AI chatbots in Zimbabwe: Balancing pedagogical innovation and development of higher-order cognitive skills. *Journal of Applied Learning and Teaching*, 7(1). <https://doi.org/10.37074/jalt.2024.7.1.12>
- Li, Q., Fu, L., Zhang, W., Chen, X., Yu, J., Xia, W., Zhang, W., Tang, R., & Yu, Y. (2024). Adapting large Language models for education: Foundational capabilities, potentials, and challenges. *ArXiv*. <https://doi.org/10.48550/ArXiv.2401.08664>. No. ArXiv:2401.08664.
- Lydia, E. G., Vidhyavathi, P., Malathi, P., & AI IN EDUCATION: OPPORTUNITIES AND CHALLENGES FOR PERSONALIZED LEARNING. (2023). A STUDY ON. *Industrial Engineering Journal*, 52(05), 750–759. <https://doi.org/10.36893/IEJ.2023.V52I05.750-759>

- Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2), 2236469. <https://doi.org/10.1080/2331186X.2023.2236469>
- Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press.
- Mayer, C. W. F., Ludwig, S., & Brandt, S. (2023). Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large Language models. *Journal of Research on Technology in Education*, 55(1), 125–141. <https://doi.org/10.1080/15391523.2022.2142872>
- Miguel, A., Cardona, E. D., Rodríguez, R. J., & Ishmael, K. (2023). Artificial Intelligence and the Future of Teaching and Learning: Insights and Recommendations. <https://policycommons.net/artifacts/3854312/ai-report/4660267/>
- Montenegro-Rueda, M., Fernández-Cerero, J., Fernández-Batanero, J. M., & López-Meneses, E. (2023). Impact of the implementation of ChatGPT in education. *A Systematic Review Computers*, 12(8). <https://doi.org/10.3390/computers12080153>. Article 8.
- Plano Clark, V. L. (2017). Mixed methods research. *The Journal of Positive Psychology*, 12(3), 305–306. <https://doi.org/10.1080/17439760.2016.1262619>
- Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599. <https://doi.org/10.1007/s40593-016-0110-3>
- Schaffler, Y., Probst, T., Jesser, A., Humer, E., Pieh, C., Stippl, P., Haid, B., & Schigl, B. (2022). Perceived barriers and facilitators to psychotherapy utilization and how they relate to patient's psychotherapeutic goals. *Healthcare*, 10(11), 2228. <https://doi.org/10.3390/healthcare10112228>
- Shibani, A., Knight, S., Kitto, K., Karunanayake, A., & Buckingham Shum, S. (2024). Untangling Critical Interaction with AI in Students' Written Assessment. Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, 1–6. <https://doi.org/10.1145/3613905.3651083>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- Tan, X. (2023). The impact of ChatGPT on education and future prospects. Highlights in science. *Engineering and Technology*, 61, 138–143. <https://doi.org/10.54097/hset.v6i1.10285>
- Tashakkori, A., & Teddlie, C. (2010). *SAGE handbook of mixed methods in social & behavioral research*. SAGE.
- Workspace, G. (n.d.) (Ed.). (2025). Google Forms: Online Form Builder. Google Workspace. Retrieved March 15, from <https://workspace.google.com/products/forms/>
- Кочубей, В. (2021). Труднощі Написання академічного Есе Та Шляхи Їх Подолання. Науковий Вісник Чернівецького Національного Університету Імені Юрія Федьковича. Серія: Германська Філологія, 831–832, Article 831–832. <https://doi.org/10.31861/gph2021.831-832.147-156>

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