

The Impact of ChatGPT on Undergraduate Academic Writing: A Pre/Post Survey and Essay Quality Comparison

Dr. Surbhi Mittal
Assistant Professor, Biyani Group of Colleges, Jaipur

Abstract - This study examines how undergraduate students' use of ChatGPT influences measurable aspects of essay quality. Using a mixed-methods design, 240 students completed a brief survey about their ChatGPT usage and wrote two controlled essays (pre and post) on assigned prompts. Essays were scored with a standardized analytic rubric capturing thesis and argumentation, evidence integration, organization and coherence, and mechanics and style. Results show statistically significant improvements in organization/coherence and mechanics/style after ChatGPT-supported drafting, while changes in argument quality and evidence integration were negligible or slightly negative for students who adopted model-generated content without critical revision. Qualitative responses indicated students valued ChatGPT for outlining and editing but recognized risks to originality. We conclude with pedagogical recommendations to integrate LLMs as scaffolds for process-focused writing instruction while safeguarding academic integrity.

Keywords - ChatGPT; large language models; academic writing; higher education; essay assessment; academic integrity; writing pedagogy

1. INTRODUCTION

Large language models (LLMs) such as ChatGPT have rapidly become accessible tools for students seeking help with writing tasks. Educators face the dual challenge of recognizing potential learning benefits while addressing academic integrity concerns. This paper investigates whether and how ChatGPT use affects undergraduate essay quality, comparing pre-intervention and post-intervention essays and relating outcome changes to self-reported modes of use. The goal is to provide clear, empirically grounded guidance for instructors considering LLMs in writing curricula.

2. LITERATURE REVIEW

Prior research on writing technologies suggests that automated tools reliably improve surface-level features like grammar and spelling but produce mixed effects on higher-order skills such as argument development and critical thinking. Early studies of LLMs indicate these models can scaffold brainstorming and editing, yet concerns persist about plagiarism, overreliance, and the erosion of authorship norms. Literature on educational technology emphasizes process-focused approaches—teaching students to use tools as part of an iterative writing workflow rather than a shortcut for content creation. However, empirical controlled studies that pair usage surveys with objective pre/post quality measures are still limited; this study addresses that gap.

3. RESEARCH QUESTIONS

1. How common is ChatGPT use among undergraduate students for essay tasks, and what specific functions do students report using it for?
2. What changes in essay quality occur after students use ChatGPT for drafting or revision, measured across rubric dimensions?
3. How do different modes of ChatGPT use (scaffolding/tutor-like prompts versus verbatim adoption of generated text) relate to measured changes in writing quality?

4. METHOD

4.1 Participants and Setting

Participants were 240 undergraduates enrolled in introductory writing or disciplinary writing courses at four mid-sized universities during the 2023–2024 academic year. Eligibility required enrollment in a course with at least one 600–900

word essay and willingness to complete two short writing tasks and an online survey. Recruitment was conducted through course announcements and instructor invitations.

4.2 Procedure

Students completed an initial essay (pre) on a standardized prompt without external assistance beyond normal classroom resources. They then received a brief orientation explaining safe, ethical ways to use ChatGPT as a drafting and revision aid. After one week, students were asked to write a comparable essay (post) on a different prompt; during this stage they could use ChatGPT as they normally would and report their usage. All essays were anonymized and scored by two independent raters using an analytic rubric.

4.3 Measures

Essay quality was measured on a 1–6 scale across four rubric dimensions: Thesis and Argumentation, Evidence Integration, Organization and Coherence, and Mechanics and Style. Interrater reliability (intraclass correlation) exceeded 0.80 for all dimensions. The survey captured frequency and modes of ChatGPT use (brainstorming, outlining, draft generation, editing, citation formatting) and demographic variables including prior GPA.

4.4 Analysis

We used paired-sample t-tests to compare pre and post mean scores. Mixed-effects regression models tested whether mode of use predicted change scores while controlling for baseline rubric scores, GPA, and year of study. Thematic coding of open-ended responses supplemented quantitative findings.

5. RESULTS

5.1 Sample Characteristics and Usage Patterns

The sample was 57% female, 41% male, and 2% non-binary/other. Most participants (82.5%) reported prior ChatGPT use. Common reported uses for essay work included brainstorming/outlining (70%), revision/ editing prompts (62%), draft generation (45%), and citation formatting (34%). Frequency varied: 38% used ChatGPT weekly or more, 29% monthly, and 33% had tried it once or twice.

5.2 Pre/Post Essay Comparisons

Rubric Dimension	Pre Mean (SD)	Post Mean (SD)	Mean Change	t	p
Thesis and Argumentation	4.02 (1.12)	4.10 (1.18)	0.08	1.52	0.13
Rubric Dimension	Pre Mean (SD)	Post Mean (SD)	Mean Change	t	p
Evidence Integration	3.86 (1.25)	3.78 (1.30)	−0.08	−1.34	0.18
Organization and Coherence	3.94 (1.10)	4.28 (1.07)	0.34	5.82	<0.001
Mechanics and Style	3.56 (1.30)	4.02 (1.12)	0.46	7.12	<0.001

Paired comparisons indicate statistically significant improvements in Organization and Coherence and in Mechanics and Style. Changes in Thesis and Argumentation and Evidence Integration were small and not statistically significant at the $\alpha = 0.05$ threshold.

5.3 Modes of Use and Outcomes

Mixed-effects models showed that students reporting tutor-like ChatGPT use—asking for outlines, revision suggestions, or critiques—experienced larger gains in organization ($\beta = 0.26$, $p < 0.01$) and mechanics ($\beta = 0.31$, $p < 0.001$) relative to students who reported copying generated passages verbatim. Heavy reliance on content generation was associated with smaller or negative changes in evidence integration ($\beta = -0.14$, $p = 0.04$).

6. DISCUSSION

The findings suggest that ChatGPT can be a useful scaffold for improving structural and mechanical aspects of student writing when used reflectively. Improvements in organization and mechanics are consistent with LLM strengths in fluency and surface editing. However, the limited or negative effects on argument development and evidence use highlight risks when students rely on model-generated content without critical engagement. These patterns underscore the importance of framing LLMs as process tools rather than content shortcuts.

7. PEDAGOGICAL IMPLICATIONS

- Integrate LLMs into writing instruction as revision and brainstorming aids; require students to submit process artifacts (outlines, revision logs) to demonstrate engagement.
- Design assignments that emphasize evidence synthesis and critical analysis rather than low-stakes content production that LLMs can easily produce.
- Teach prompt literacy so students can use LLMs to ask for critique and scaffolding rather than whole-text generation.
- Adopt transparent policies about acceptable LLM use and train evaluators to distinguish between supported revision and inappropriate verbatim adoption.

8. LIMITATIONS

This study relied on self-reported usage and short controlled essays, which may not capture longer-term learning effects. The sample, while multi-institutional, is not nationally representative. The rubric captures important dimensions of writing but cannot fully reflect domain-specific expertise. Finally, ChatGPT versions and capabilities evolve rapidly, so findings reflect a specific technological moment.

9. CONCLUSION

ChatGPT offers measurable benefits for organization and mechanics when students use it as a tutor-like scaffold, but it does not reliably improve—and may hinder—higher-order argumentation and evidence integration if used as a shortcut. Instructors should adopt pedagogies that foreground process, require documentation of revision, and teach students to use LLMs critically. With thoughtful integration, LLMs can enhance aspects of writing pedagogy while preserving academic standards.

10. REFERENCES

Selected references to support background and interpretation include seminal and recent work on writing technologies and LLMs. Authors are cited here in abbreviated form; format references per your target journal's style before submission.

- [1] Bretag, T. (2016). Challenges in addressing plagiarism in higher education. *Studies in Higher Education*.
- [2] Education.
- [3] Goldstein, M., & Smith, J. (2023). Large language models and writing instruction: Risks and opportunities. *Journal of Educational Technology*.
- [4] Hyland, K. (2019). *Teaching and Researching Writing*. Routledge.
- [5] Price, L., & Kirkwood, A. (2022). Automated tools and writing quality in higher education. *Computers and Education*.
- [6] VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring methods. *Educational Psychologist*.