

Effects of Varying LLM Access on Essay Writing Behavior

Julia Christenson, Karin de Langis, Shirley Anugrah Hayati, Dongyeop Kang

University of Minnesota

{chri5306, dento019, hayat023, dongyeop}@umn.edu

Abstract

Investigating the degree to which large language models (LLMs) affect teaching and learning in universities can help identify strategies for integrating LLMs in a way that supports, rather than undermines, student learning outcomes. This study examined how varying levels of LLM assistance affect writing performance, engagement, and perceived authorship. We report a pilot study in which 24 college students were randomly assigned to write a short essay with no LLM access, limited access (≤ 3 prompts, responses capped at 100 words), or unlimited access. Overall essay quality was statistically indistinguishable across groups. Yet writing behavior and perceived authorship diverged sharply: students with limited access reported higher ownership (62.5% would submit the essay as independent work, vs. 25% in the unlimited group), stronger organizational gains, and more strategic, revision-focused prompting. The unlimited group spent more time writing, produced essays more similar to LLM output, and reported reduced creative expression. Our findings suggest that constraining, rather than banning, LLM access may preserve authorship confidence while retaining the scaffolding benefits of AI assistance.

1 Introduction

LLMs have become increasingly embedded in academic writing practices, reshaping how students plan, compose, and revise text. LLM-based tools such as ChatGPT and Grammarly now offer real-time feedback and idea generation, lowering barriers to writing while simultaneously raising questions about authorship, learning, and skill development. As these systems become normalized in educational contexts, it is important to understand their influence on student writing behavior.

Prior work primarily frames LLM-assisted writing as a productivity and accessibility aid, reporting gains in fluency, error reduction, and brainstorming support across literacy levels, as well as tools

that scaffold multiple stages of the writing process (Weijers et al., 2024; Lee et al., 2024; Zhang, 2025; Xu, 2025). At the same time, researchers caution that sustained reliance on LLM may undermine independent writing, creativity, and critical thinking (Rohilla, 2025; Gerlich, 2025a). Others find no significant effect on overall academic achievement (Yoo et al., 2025). Students often report reduced confidence and creativity with prolonged use, and prior work also highlights risks related to dependency, ethical ambiguity, and homogenized or biased outputs (Zhai et al., 2024; Chakrabarty et al., 2025). Human-AI interaction research further shows that AI involvement can negatively influence perceptions of authorship and writing quality, particularly when disclosed (Draxler et al., 2024; Li et al., 2024; Mansour et al., 2024; Joshi and Vogel, 2025; Gerlich, 2025b).

Existing research has shown that the use of LLMs has negative cognitive effects (Kosmyrna et al., 2025), but we hypothesize that the degree of LLM access may play an important role in the efficacy of LLM assistants in educational contexts. This study addresses the gap by examining how different levels of LLM assistance influence students' writing by formulating three main research questions:

- **RQ1:** How does varied LLM assistance influence students' writing behavior, including time allocation, writing strategies, and prompting practices?
- **RQ2:** How do students perceive authorship, ownership, and acceptability of LLM-assisted writing?
- **RQ3:** How does the level of LLM assistance affect writing quality, originality, and lexical diversity in student essays?

To address these questions, we conduct a study in which twenty-four college students wrote an es-

say in response to a prompt. By comparing writing produced under different levels of LLM assistance, this work aims to clarify when LLM assistance enhances writing efficiency and when it begins to erode core literacy skills.

We found that both LLM-assisted groups engaged in similar write-prompt-edit cycles, but with different strategic orientations: limited-access participants focused prompts on major revisions and argument development, while unlimited-access participants concentrated on minor mechanical edits. These behavioral differences mapped onto perceived authorship: limited-access participants reported stronger feelings of ownership, greater organizational improvement, and higher willingness to submit the essay as their own work. Essay grades and lexical diversity were marginally higher in the unlimited group, though differences were not statistically significant. Taken together, the findings suggest that constraining LLM access, rather than eliminating it, may offer a middle path between the cognitive risks of over-reliance and the scaffolding benefits of AI assistance.

2 Related Work

Prior research has examined human-AI interactions for scholastic writing needs through engagement with LLMs and behavioral outcomes. Several studies on college students' use of LLMs find that learners primarily rely on LLMs for brainstorming, revision, and information retrieval (Jelson et al., 2025; Kondoro, 2025; Wang et al., 2024). Although LLM assistants effectively correct surface-level language errors, they have been shown to reduce syntactic complexity and linguistic diversity with sustained use (Wang and Spitz, 2025; Bui and Kim, 2025; Pan et al., 2025). Some studies found that students with higher AI literacy and more strategic prompting practices tend to produce stronger writing, while students with lower performance often under-utilize or misuse AI tools (Nguyen et al., 2024; Kim et al., 2025; Joshi and Vogel, 2025). A few studies observed modest improvements in students' independent learning and English writing ability when such tools are used alongside instruction (Wang, 2020; Seo, 2024; Xiao, 2024; Shen, 2025). However, these prior works share limitations. Most studies employ binary comparisons—AI versus no-AI—rather than examining a response relationship to varied levels of AI access. This framing obscures whether the negative effects

of LLM use are properties of AI assistance itself or the degree to which it is provided.

Growing concerns about dependence on LLMs negatively affecting critical thinking skills are prevalent. Survey research reveal many students and teachers are aware of the over-reliance on LLM writing assistants and concerns over academic dishonesty (Aliakbari et al., 2026; Joshi and Vogel, 2025; Kondoro, 2025; Khalid et al., 2025). Several studies show lowered critical thinking abilities for those who exhibit higher dependence on LLMs (Kosmyrna et al., 2025; Gerlich, 2025a).

Recent studies pivoted to researching methods to increase perceived ownership and transparency through tailored LLM assistants. One study has found that greater AI transparency and interpretability are associated with more positive student experiences and reduced misuse, including academic dishonesty and diminished learning outcomes (Cui, 2025). Several works have studied human-AI collaboration patterns and suggest writer-centered approaches to LLM-based writing assistants, (Mysore et al., 2025; Guo et al., 2024; Liu and August, 2025; Gerlich, 2025b; Shibani et al., 2024), but largely emphasize interaction strategies and design recommendations without measuring whether such interventions actually improve writing outcomes or reduce problematic dependency patterns.

Empirical studies demonstrate the benefits of LLM assistance, and the detrimental effects of AI on cognition and critical thinking skills, but the impact in real-world academic settings remains unclear (Yoo et al., 2025). Together, these gaps motivate our study focused on how constraining LLM access can impact how students navigate a real-world academic situation.

3 Approach

We aimed to investigate the extent to which students' writing experiences are impacted by the assistance LLM provides for them. The primary concern of the study was whether limiting LLM use provides a viable middle ground between the negative cognitive effects of LLM use and the benefits.¹

3.1 Participants

We recruited a total of 24 undergraduate and graduate students with 18 different majors through email recruitment. Details about participant demographic

¹The study was approved by the IRB.

Category	Percentage
Race	
White	41.7%
Asian	16.6%
Black or African-American	8.3%
Hispanic	4.2%
Two or more races	25.0%
Prefer not to disclose	4.2%
Gender	
Female	75.0%
Male	25.0%
Major	
STEM	41.7%
Health and Medicine	29.2%
Arts and Humanities	16.7%
Business	8.3%
Social Sciences	4.2%
Academic Level	
Senior	66.7%
Junior	16.7%
Sophomore	4.2%
Freshman	8.3%
Graduate	4.2%

Table 1: Participant demographics (N=24).

information are provided in Table 1. 71% of the participants rated themselves as strong writers by selecting either a 4 or 5 on a Likert scale out of 5. Participants, on average, have used LLM-based tools in some way in their previous 5 out of 10 writing assignments. Most students who use LLM tools in coursework reported using ChatGPT and built-in tools such as Grammarly. Participants who regularly use LLM-based tools typically employ them to generate ideas, seek assistance when they get stuck, and understand the material better. The participants who used LLM tools less frequently described the writing output as repetitive and superficial, and noted that they avoid LLM tools due to environmental concerns and academic integrity.

3.2 Study Procedure

First, participants completed a preliminary survey to determine eligibility. The eligibility requirements were native-level English fluency and identifying as an undergraduate or graduate student. Participants were randomly placed into one of three groups: control, limited access to LLM, or unlimited access. Participants then completed an approximately 30-minute in-person writing session. All participants provided informed consent under an IRB-approved protocol. Participants were given instructions to respond to an essay prompt in at least

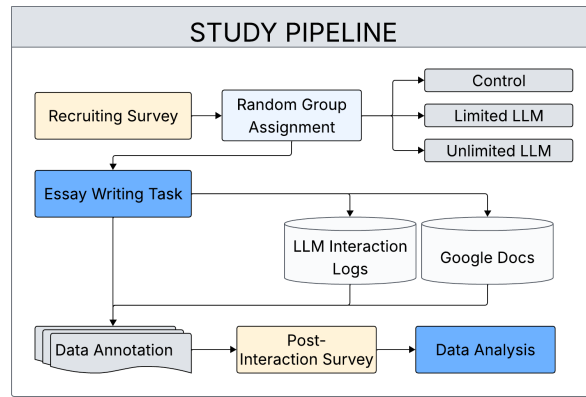


Figure 1: Study pipeline.

300 words, using a text document. Instructions can be found in Appendix A.1. Each participant received the same prompt, shown below:

Essay Prompt

Write a response in which you discuss the extent to which you agree or disagree with the recommendation and explain your reasoning for the position you take.

College students should be encouraged to pursue subjects that interest them rather than the courses that seem most likely to lead to jobs.

Since participants completed the writing task on their personal computers, we also instructed them to disable all AI writing extensions (e.g., Grammarly) on their browsers.

Participants in the control group did not receive any help from LLM when writing their essays, while participants in other groups have access to chat with LLM-based chatbots during their writing process. To access the LLM-based chatbot on the website, we implemented a backend application using Flask and Python, deployed with Vercel. We used the Gemini-2.5-Flash model (Gemini Team, 2025) for our chatbot because of our institution’s license.

Participants in the limited LLM access and unlimited LLM access groups interacted with their respective chatbots through a provided link. The writing assistant for the limited LLM group operated with a system prompt that restricted its responses to general guidance. The full system prompt is in Appendix A.2. This group was also given a maximum of three prompts to ask the LLM-based writing as-

Example Essay #1

Many people state that college is the best year of your life. I believe that is because it is a time for exploration and opportunities. This is the time to take new risks and try everything that interests you. However, students in college are at varying places in the world. Some have the privilege of getting financial help, whereas others may need to work multiple jobs to stay stable while getting their degree. While career readiness is important in a selection of classes, students should balance these with classes that also interest them.

Taking classes based on interest can fuel one's desire to learn and boost motivation. When students are actually curious about a topic, they will participate more in class and engage more deeply with the material. Connecting this to my own experience, I took a class on rock music, even though I am not studying music. I found myself enjoying the class a lot more because I enjoyed the topic and learning about music history.

Apart from student success in college, post-college happiness is often tied to fulfillment. Research shows that while income certainly matters up to a certain point (often cited around \$75,000 to alleviate daily stress), long-term job and life satisfaction are more closely tied to intrinsic factors like meaning and interest. These classes you can take in college are much more worth exploring if it could mean you find a new passion or even a new hobby.

LLM-assisted portion

Human-written portion

Figure 2: Example annotated essay with highlighted LLM-assisted portions for the limited group. This student interweaves responses from LLM to support their main claim in this essay.

sistant, whereas the unlimited access group had no such limit. The number of prompts was selected due to the shortened length of the writing task. The limited LLM group had a maximum token output and a system prompt instruction to provide short responses that reflect general guidance and encourage students to develop their own ideas, rather than longer responses that directly answer the prompt. Meanwhile, the unlimited LLM group did not have a restricted token output or special instructions.

Upon completion of the essay, participants were given a post-interaction survey to assess their perception of the writing process and their essay. Participants received a \$10 gift card as compensation.

4 Methodology

4.1 Data Collection & Annotation

We collected the interactions with the LLM via our web application and annotated the writing actions and interactions. Due to the limited number of participants and short writing task given, we developed a simplified taxonomy inspired by previous research (Du et al., 2022; Nguyen et al., 2024) with the following four writing intentions: writing, editing, structuring, and prompting.

We collected interaction data through a custom web application that logged participants' exchanges with the LLM, along with their writing activity in Google Docs. These logs enabled us to analyze both writing actions and prompting behavior throughout the essay-writing process.

To understand participants' writing behavior, we developed a taxonomy of four activities during the writing process: *writing*, *editing*, *structuring*, and *prompting*, informed by prior work on writing processes and revision behaviors. Writing refers to generating new text in full sentences, editing captures revisions to previously written content, and structuring involves organizing ideas or outlining the essay. Prompting means that the user is interacting with LLM to support writing. Two members of the research team manually annotated participants' writing intentions using this taxonomy, based on the Google Docs revision history and logged LLM interactions. Inter-annotator agreement for the writing intentions was assessed using Cohen's κ , yielding substantial agreement ($\kappa = 0.71$).

4.2 Metrics

We established a set of metrics designed to capture the nuances of writing behavior, essay quality, lexical/semantic overlap with LLM output, and vocabulary richness, allowing us to correlate them with survey responses later in the analysis. Given our limited statistical power to detect all but very large effects on writing quality, we present descriptive statistics without asserting statistical significance.

Rubric-Based Evaluation Essays were evaluated using a custom rubric derived from standard English composition criteria, as shown in A.2. The rubric assesses seven distinct dimensions: Focus and Details, Idea Support and Development, Organization, Voice, Vocabulary and Word Choice, Sentence Structure and Grammar, and Reads as human-written², in which a higher score meant more human-like quality. Each category was scored on a five-point scale and accompanied by a brief justification. This rubric was inspired by standard English essay rubrics (of Teachers of English).

To ensure reliability, grading was conducted independently by two university writing consultants employed at our institution's writing center, with the final rubric scores determined by averaging

²This metric functions as a holistic AI-detection proxy rather than a compositional quality measure

their two grades.³

Writing Time and Volume We tracked the total time each participant spent writing the essay (in minutes) to capture differences in writing effort and process efficiency across conditions. We also recorded the final word count of each essay as a measure of output length and productivity.

Lexical Diversity We measured lexical diversity using the Root Type-Token Ratio (Root TTR) (Guiraud, 1954). This metric provides a stable estimate of lexical diversity across essays of varying lengths by calculating the ratio of unique words (types) to the square root of the total words (tokens). Higher values indicate a broader, more diverse vocabulary usage relative to the total length of the text.

Prompt Type Prompts were categorized into four types: brainstorming/developing arguments, structure/organization, major revisions, and minor revisions. Developing arguments represented prompting to brainstorm talking points and specific facts or statistics. Structure/organization prompts focused on structuring the essay and making it cohesive. Major revisions included large revisions to the text, such as prompting to rewrite the entire essay, fixing grammatical mistakes, or modifying the essay stylistically. Lastly, minor revisions focused on rephrasing, sentence-level grammar checks, and asking for synonyms. Given the accuracy rate of annotations done by ChatGPT (Churchill et al., 2026), and due to the large number of prompts, the prompts were initially organized using GPT-4 (OpenAI et al., 2024) and all verified by one member of the research team with 94% accuracy.

AI Similarity Scores To assess lexical overlap with LLM output and the degree to which participants incorporated LLM-generated text, we computed cosine similarity scores between the final student essays and the raw LLM responses generated during their specific writing session. Sentence embeddings were generated using the all-MiniLM-L6-v2 model (Reimers and Gurevych, 2019). This resulted in a similarity score ranging from 0 to 1, where higher scores indicate a greater direct incorporation of LLM-generated phrasing.

³The graders were recruited via email and provided informed consent under an IRB-approved protocol.

5 Findings

As we attempt to understand how different levels of LLM assistance influence writing outcomes and writing behaviors, we organize our analysis around these main research questions: [RQ1]: How does varied LLM assistance influence students' writing behavior, including time allocation, writing strategies, and prompting practices? (Section 5.1) [RQ2]: How do students perceive authorship, ownership, and acceptability of LLM-assisted writing? (Section 5.2) [RQ3]: How does the level of LLM assistance affect writing quality, originality, and lexical diversity in student essays? (Section 5.3)

To address these questions, we combine rubric-based essay evaluation, lexical and similarity metrics, interaction logs, and survey responses. Each subsection below explicitly corresponds to one or more research questions.

5.1 Writing Behavior and Interaction Patterns

To examine how varied LLM assistance shapes writing behavior, we analyzed writing time, prompt usage patterns, and survey results across groups.

Writing Time Average writing time increased with greater LLM access. The control group completed essays in an average of 22.5 minutes, the limited LLM group in 27.8 minutes, and the unlimited LLM group in 30.4 minutes, as summarized in Table 5. Notably, variance also differed across groups: the unlimited group showed the least spread in completion time (most participants clustered near the 30-minute mark), while the limited group showed the greatest variability (Figure 3). One interpretation is that the three-prompt cap introduced individual differences in pacing: students who front-loaded prompts early finished faster, while those who saved prompts for revision extended their sessions. The unlimited group, facing no strategic constraint, may have converged on a similar write-and-revise rhythm regardless of individual differences. No correlation between time and essay grade was observed.

Writing Process We analyzed temporal writing intentions by normalizing each participant's writing session from start to finish. The limited group averaged 2.5 prompts, compared to 4.5 prompts for the unlimited group. Figures 5a, 5b, and 5c illustrate how participants distributed effort across planning, writing, prompting, and editing.

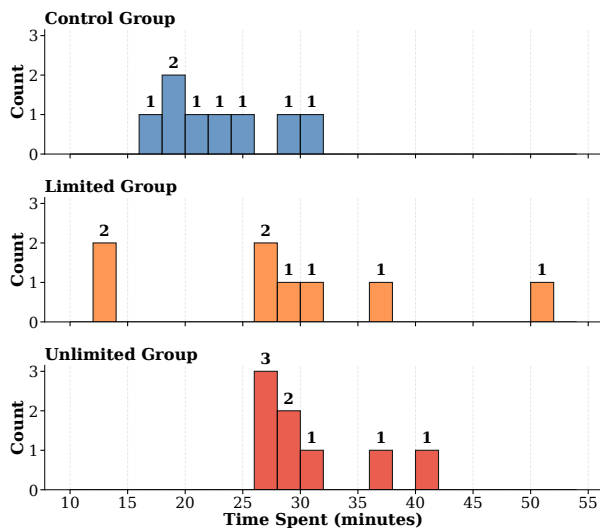


Figure 3: Distribution of writing time (in minutes) by user group.

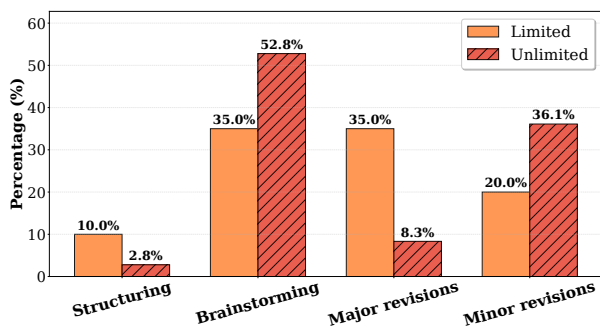


Figure 4: Distribution of the type of prompts asked by the LLM groups.

As shown in Figure 5a, two participants in the control group spent time on structuring the essay, and many participants spent most time writing sequentially, with additional time spent on editing the essay. Figure 5b shows the writing strategies of the limited LLM group. Six users demonstrated their collaboration with LLM, with writing themselves first, then prompting, and continuing this process until they were out of prompts. Despite having unlimited access to LLM, figure 5c shows that six of eight unlimited users still started the essay themselves, then prompted LLM for additional help. A majority of users in both groups with access to LLM utilized LLM in the latter part of their writing process (0.8-1).

Across all groups, most participants began by writing independently before seeking assistance. Editing behaviors were equally prevalent across all groups, while explicit structuring activities were relatively rare across all conditions. The increased LLM access was not associated with a difference

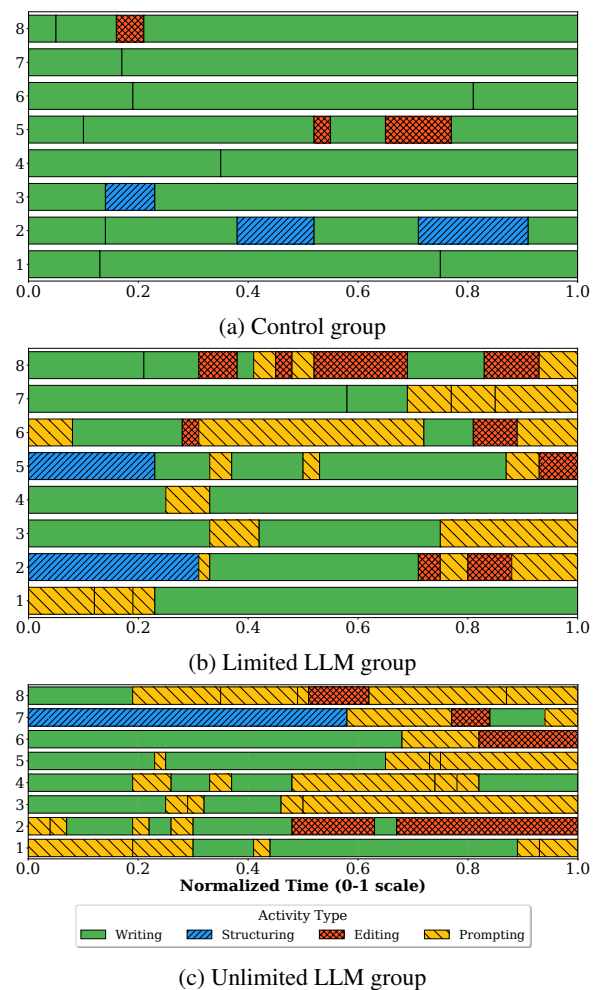


Figure 5: Distribution of writing intentions across the writing process for all user groups (8 participants per group). Each row represents a participant's timeline, with colored segments indicating their activities across normalized time.

in the writing and prompting process.

Prompt Types Figure 4 shows that brainstorming/argument development dominated prompts in both LLM assisted groups. Following argument-focused prompts, the unlimited group spent 36% of their prompts on minor revisions to the text. The limited users distributed prompts differently, with 35% on major revisions, and 20% on minor revisions.

Both groups showed low proportions of prompts directed toward structuring or organizing the essay. This pattern may reflect the possibility that participants may have completed structural planning independently, or the constrained length of the writing task may have reduced the concern over organization. The observed difference in major versus minor revision allocation between groups

Table 2: Limited group correlation matrix between pre-survey and post-session measures ($*p < 0.05$).

	Acceptance	Dev. Arguments	Time
Crit. thinking eval.	0.92*	–	–
Prompt literate	-0.80*	0.71*	–
Relied on AI	–	–	0.78*

Table 3: Unlimited group correlation matrix between pre-survey and post-session measures ($*p < 0.05$).

	Acceptance	Similarity	Prompt #
Crit. thinking eval.	–	-0.79*	–
Saved time	–	–	-0.71*
Comfortable with AI	-0.91*	–	–
AI use frequency	-0.80*	–	–
Difficulty	0.81*	–	–

aligns with expectations that the restricted number of prompts encouraged the students to prioritize higher-impact revisions.

5.2 Perceptions of Ownership and Acceptability

To understand students’ perceptions of authorship when using LLM to help their writing process, we analyzed responses from post-interaction surveys. We report the behavioral metrics as descriptive statistics due to low statistical power.

Post-interaction survey Post interaction responses revealed ambivalence toward LLM-assisted authorship.

First, participants in the limited group expressed more ownership towards their essays. When asked if they would submit it as their own work, 62.5% responded yes, 25% were unsure, and 12.5% declined. Most of the participants also perceived an increased ability to organize ideas (62.5%).

Correlational analysis within the limited group revealed several patterns (Table 2). Participants who used LLM frequently in writing assignments reported greater perceived need for critical thinking to decide what to accept or reject from the LLM ($r = 0.92, p < 0.05$). Those with self-reported prompt literacy accepted fewer prompts ($r = -0.80, p < 0.05$), but more frequently directed prompts towards developing arguments ($r = 0.71, p < 0.05$). This pattern suggests that the increased critical engagement with LLM outputs and prompting style may reflect an adjustment to the constrained nature of the system’s responses compared to their previous exposure to LLM tools, such as ChatGPT. Participants who reported spend-

ing more time writing the essay also revealed feeling greater reliance on LLM ($r = 0.78, p < 0.05$), suggesting more time was spent on the LLM interface, prompting, and incorporating suggestions.

Conversely, **the unlimited group demonstrated less ownership over their work**. Only 25% of these participants would submit their essay as their own work, and 75% claimed unsure. This suggests that this group had increased concerns over academic integrity and perceived originality. However, alternative explanations, such as low confidence in content or structure, cannot be ruled out.

Correlational analysis for the unlimited group (3) revealed several associations. A moderate negative correlation ($r = -0.79, p < 0.05$) between the cosine similarity score and the perceived critical thinking demands was observed. This implies that **students who actively evaluated and rejected LLM output produced more original work**, although causality cannot be determined.

Students in the unlimited group who believed they saved time using LLM used fewer prompts ($r = -0.71, p < 0.05$). Conversely, comfort with LLM tools ($r = -0.91, p < 0.05$) and higher baseline LLM use frequency ($r = -0.80, p < 0.05$) were associated with lower prompt acceptance rates. Additionally, perceived task difficulty showed a positive relationship with prompt acceptance ($r = 0.81, p < 0.05$), suggesting that students reporting higher difficulty more readily incorporated suggestions.

Notably, across both groups with access to LLM, students with stronger self-reported prompting skill accepted fewer suggestions overall (limited: $r = -0.80$; unlimited: $r = -0.91$). This pattern could indicate that **students with stronger prompting literacy applied more critical judgment in evaluating suggestions**. These results support prior research that prompt-literate individuals are more likely to think more critically of the output and engage in iterative interactions with LLMs (Kim et al., 2025).

Participants also expressed nuanced views on LLM use in coursework, aligning with previous studies (Aliakbari et al., 2026; Joshi and Vogel, 2025). Most responses indicated that LLM should be permitted for tasks such as brainstorming, source discovery, and concept clarification, but not for completing assignments or copying generated text. A small number of participants favored stricter limitations or unrestricted use, reflecting ongoing tension between perceived educational value and

Table 4: Average essay grades by user group, with highest scores in bold. Overall, differences between groups in grading outcomes are minor.

Rubric Item	Control	Limited	Unlimited
Focus + Details	4.3	4.5	4.6
Idea Support	4.6	4.5	4.7
Organization	4.3	4.3	4.7
Voice	4.9	4.7	4.5
Vocabulary	4.1	4.4	4.4
Sentence/Grammar	3.9	4.9	4.7
Reads as human-written	4.9	4.6	4.6
Total Score	88.66%	88.39%	89.38%

concerns about learning and academic integrity.

5.3 Writing Quality and Originality

To evaluate writing quality and originality, we analyzed essays using a rubric-based assessment, lexical diversity metrics, and similarity comparisons between student essays and LLM-generated responses.

Rubric Based Evaluation Table 4 shows the average rubric scores, which were largely stable between groups. The statistical power of our study only allows us to detect very large effects of group on grading, and we do not find any significant differences between groups.

Lexical Diversity As shown in Table 5, the unlimited group exhibited the highest average Root TTR (9.555), followed by the limited group (9.4365) and the control group (8.963). These results suggest that **access to LLMs may encourage broader vocabulary use**. However, we again cannot definitively determine any statistical relationships given our limited power.

AI Similarity Scores The unlimited group exhibited a higher average similarity score (0.552) compared to the limited group (0.444), suggesting marginally greater incorporation of LLM output (See Table 5). Although this is consistent with expectations, the difference was small and not statistically significant (ANOVA $p = 0.21$), preventing definitive conclusions regarding the dependence of LLM based only on similarity.

6 Discussion

Contrary to initial concerns, unlimited LLM access did not produce lower-quality essays than limited access or no LLM baselines. Only marginal differences in lexical diversity and similarity between

Table 5: Average Metric Evaluations by User Group

Metric	Control	Limited	Unlimited
Essay Grade	88.66%	88.39%	89.38%
Cosine Similarity	–	0.444	0.552
Root TTR	8.963	9.4365	9.555
Writing Time (mins)	22.5	27.8	30.4
Word Count	329	439	383

LLM suggestions and essay text were exhibited. However, the unlimited group spent more time writing and exhibited a higher similarity to LLM output (0.552 vs. 0.444), suggesting greater LLM reliance or slower, more iterative integration of LLM text. This aligns with prior work on cognitive offloading, where increased AI reliance may reduce rather than increase self-directed cognitive engagement (Gerlich, 2025a; Kosmyna et al., 2025).

The LLM groups demonstrated varied prompting styles. Both groups exhibited human-AI collaboration through iterating, prompting, and editing the essay, but the limited group prioritized major revisions, whereas the unlimited group targeted minor mechanical revisions. This strategic reallocation may have been correlated with measurable ownership benefits. 62.5% of limited participants reported willingness to submit their essay as independent work, compared to 25% in the unlimited group. Similarly, 62.5% of the limited group reported improved organizational ability.

Correlation analysis between pre- and post-interaction survey responses exhibited differences in how users interacted with the AI. As shown in Table 3, limited group participants who are comfortable with AI tools reported elevated critical thinking demands ($r = 0.92, p < 0.01$) and focused prompts on argument-development help ($r = 0.71, p < 0.05$). This pattern suggests that AI-literate students were more cautious and critical of the output. The unlimited LLM group exhibited a strong negative correlation between perceived need for critical thinking and LLM similarity scores. This suggests that students who evaluated LLM responses more rigorously produced more original work.

While constrained LLM access appeared to encourage strategic use and increase feelings of authorship, the study’s small sample size and simplified writing task limit generalizability.

7 Conclusion

This study examined how varied levels of LLM assistance influence college students' writing quality, writing behavior, and perceptions of authorship. In an IRB-approved study with 24 participants, results showed that overall essay quality showed no statistical significance, and systematic patterns in writing behavior, cognitive engagement, and ownership views revealed nuances about how LLMs impact student writing.

These findings motivate future research on LLM-assisted writing in more realistic academic settings, with larger samples and extended writing tasks, to better understand how students balance AI support with independent authorship.

8 Limitations

This study has several limitations. The data were collected over a two-week period from native English speaking college students in authors' university, which limits generalizability despite efforts to recruit participants from diverse backgrounds. The sample size was small, with 24 participants, due to time and resource constraints. As a result, the study focused on a single opinion-based essay prompt, which may not reflect writing behaviors or LLM interactions in other contexts, such as research-driven or evidence-based assignments. Longer writing tasks may reveal different patterns. Future work should explicitly examine EFL/ESL learners, who represent a significant and potentially more LLM-reliant group.

The model also lacked built-in conversational memory. Participants assumed that the LLM would retain context across prompts, leading to confusion when follow-up prompts were interpreted in isolation. This issue was only identified after receiving feedback from the initial participants.

Additional limitations arose in data collection and annotation. Google Docs editing histories only captured major revisions, preventing fine-grained analysis of sentence-level edits and micro revisions during continuous writing. Manual annotation of writing intentions, while validated through inter-rater agreement ($\kappa = 0.71$), introduced potential bias. Additionally, we used Gemini-2.5-Flash rather than ChatGPT, which may limit comparison to students' real-world AI experiences. We did not observe any clear attempts to circumvent the system prompt constraints. Alternative explanations for correlational outcomes (e.g., low confidence

rather than AI-induced ambiguity) cannot be fully ruled out.

We also used cosine similarity to determine an AI similarity score, which captures surface-level semantic overlap but does not directly measure originality in a deeper cognitive sense.

Future work should expand participant pools across institutions and regions, include multiple writing tasks of varying genres and lengths, incorporate memory-enabled frameworks to better reflect real-world AI writing tools, and employ more detailed logging mechanisms to capture granular writing behavior. Longitudinal designs examining skill development would also help capture long term effects of varied LLM access on writing ability.

9 Ethical Considerations

This study does have ethical considerations related to participant privacy, informed consent, and responsible use of generative AI. All participants were fully informed of the purpose of this study and the structure of the writing tasks. Participation was voluntary, and students were free to withdraw at any time without penalty. Incentives were given to participants as per gift cards in return for their time and effort. The study also considered the ethical implications of exposing participants to biased, inaccurate, or misleading AI-assisted output. The LLM was constrained to provide short, general writing suggestions rather than fully generated paragraphs or factual claims. Our study does have potential harm if there is a misinterpretation of our findings, which could encourage the misuse or altogether elimination of AI tools in a school environment. To address this, we emphasized the balance, acknowledgment, and responsibility of AI use, especially in an academic setting.

References

- Mohammad Aliakbari, Pooria Barzan, and Seyyed Pedram Allahveysi. 2026. [AI usage in academic writing: Perspectives of stakeholders](#). *AI and Tech in Behavioral and Social Sciences*, 3(4):1–12.
- Minh Ngoc Bui and Eun-young Kim. 2025. [Analysis of influential factors on the writing quality of chatgpt-generated academic essays](#). In *Proceedings of the 2024 7th International Conference on Educational Technology Management, ICETM '24*, page 315–321, New York, NY, USA. Association for Computing Machinery.

- Tuhin Chakrabarty, Philippe Laban, and Chien-Sheng Wu. 2025. [Can ai writing be salvaged? mitigating idiosyncrasies and improving human-ai alignment in the writing process through edits](#). In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA. Association for Computing Machinery.
- Alexander Churchill, Shamitha Pichika, Chengxin Xu, and Ying Liu. 2026. [Gpt models for text annotation: An empirical exploration in public policy research](#). *Policy Studies Journal*, 54(1):e70034. E70034 PSJ-24-RNR-0078.R2.
- Yulu Cui. 2025. [What influences college students using ai for academic writing? - a quantitative analysis based on hisam and tri theory](#). *Computers and Education: Artificial Intelligence*, 8.
- Fiona Draxler, Anna Werner, Florian Lehmann, Matthias Hoppe, Albrecht Schmidt, Daniel Buschek, and Robin Welsch. 2024. [The ai ghostwriter effect: When users do not perceive ownership of ai-generated text but self-declare as authors](#). *ACM Trans. Comput.-Hum. Interact.*, 31(2).
- Wanyu Du, Vipul Raheja, Dhruv Kumar, Zae Myung Kim, Melissa Lopez, and Dongyeop Kang. 2022. [Understanding iterative revision from human-written text](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3573–3590, Dublin, Ireland. Association for Computational Linguistics.
- Google Gemini Team. 2025. [Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities](#). *arXiv preprint arXiv:2507.06261*.
- Michael Gerlich. 2025a. [Ai tools in society: Impacts on cognitive offloading and the future of critical thinking](#). *Societies*, 15(1).
- Michael Gerlich. 2025b. [From offloading to engagement: An experimental study on structured prompting and critical reasoning with generative ai](#). *Data*, 10(11).
- Pierre Guiraud. 1954. [Les caractères statistiques du vocabulaire: essai de méthodologie](#). *Presses universitaires de France*.
- Alicia Guo, Leijie Wang, Jeffrey Heer, and Amy Zhang. 2024. [Preserving writer values in ai writing assistance tools](#). In *Proceedings of the Third Workshop on Intelligent and Interactive Writing Assistants*, In2Writing '24, page 58–61, New York, NY, USA. Association for Computing Machinery.
- Andrew Jelson, Daniel Manesh, Alice Jang, Daniel Dunlap, Young-Ho Kim, and Sang Won Lee. 2025. [An empirical study to understand how students use chatgpt for writing essays](#). *Preprint*, arXiv:2501.10551.
- Nikhita Joshi and Daniel Vogel. 2025. [Writing with ai lowers psychological ownership, but longer prompts can help](#). In *Proceedings of the 7th ACM Conference on Conversational User Interfaces*, CUI '25, page 1–17. ACM.
- Eman Khalid, Hammad Javaid, Yashal Waseem, and Natasha Sohail Barlas. 2025. [AI as a mind partner: Cognitive impact in Pakistan's educational landscape](#). In *Proceedings of the Artificial Intelligence in Measurement and Education Conference (AIME-Con): Works in Progress*, pages 14–19, Wyndham Grand Pittsburgh, Downtown, Pittsburgh, Pennsylvania, United States. National Council on Measurement in Education (NCME).
- Jinhee Kim, Seongryeong Yu, Sang-Soog Lee, and Rita Detrick. 2025. [Students' prompt patterns and its effects in ai-assisted academic writing: Focusing on students' level of ai literacy](#). *Journal of Research on Technology in Education*, pages 1–18.
- Alfred Malengo Kondoro. 2025. [AI writing assistants in tanzanian universities: Adoption trends, challenges, and opportunities](#). In *Proceedings of the Fourth Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2025)*, pages 37–46, Albuquerque, New Mexico, US. Association for Computational Linguistics.
- Nataliya Kosmyrna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes. 2025. [Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task](#). *Preprint*, arXiv:2506.08872.
- Mina Lee, Katy Ilonka Gero, John Joon Young Chung, Simon Buckingham Shum, Vipul Raheja, Hua Shen, Subhashini Venugopalan, Thiemo Wambsganss, David Zhou, Emad A. Alghamdi, Tal August, Avinash Bhat, Madiha Zahrah Choksi, Senjuti Dutta, Jin L.C. Guo, Md Naimul Hoque, Yewon Kim, Simon Knight, Seyed Parsa Neshaei, and 17 others. 2024. [A design space for intelligent and interactive writing assistants](#). In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery.
- Zhuoyan Li, Chen Liang, Jing Peng, and Ming Yin. 2024. [How does the disclosure of AI assistance affect the perceptions of writing?](#) In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4849–4868, Miami, Florida, USA. Association for Computational Linguistics.
- Yijun Liu and Tal August. 2025. [From crafting text to crafting thought: Grounding intelligent writing support to writing center pedagogy](#). In *Proceedings of the Fourth Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2025)*, pages 47–61, Albuquerque, New Mexico, US. Association for Computational Linguistics.

- Atil Mansour, Ofra Amir, and Liat Levontin. 2024. [The potentially negative effects of ai writing assistants on self-disclosure](#). In *Proceedings of the Third Workshop on Intelligent and Interactive Writing Assistants, In2Writing '24*, page 1–3, New York, NY, USA. Association for Computing Machinery.
- Sheshera Mysore, Debarati Das, Hancheng Cao, and Bahareh Sarrafzadeh. 2025. [Prototypical human-AI collaboration behaviors from LLM-assisted writing in the wild](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 16830–16857, Suzhou, China. Association for Computational Linguistics.
- Andy Nguyen, Yvonne Hong, Belle Dang, and Xiaoshan Huang. 2024. [Human-ai collaboration patterns in ai-assisted academic writing](#). *Studies in Higher Education*, 49(5):847–864.
- National Council of Teachers of English. [Essay rubric](#).
- OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, and 262 others. 2024. [Gpt-4 technical report](#). *Preprint*, arXiv:2303.08774.
- Hongchen Pan, Eduardo Araujo Oliveira, and Rafael Ferreira Mello. 2025. [Exploring human-ai collaboration in educational contexts: Insights from writing analytics and authorship attribution](#). In *Proceedings of the 15th International Learning Analytics and Knowledge Conference, LAK '25*, page 903–909, New York, NY, USA. Association for Computing Machinery.
- Nils Reimers and Iryna Gurevych. 2019. [Sentence-bert: Sentence embeddings using siamese bert-networks](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Atiya Rohilla. 2025. [Impact of excessive ai tool usage on the cognitive abilities of undergraduate students: A mixed method study](#). *Applied Science and Sustainable Agriculture Journal*, 4(1):2131–2143.
- Ji-Young Seo. 2024. [Exploring the educational potential of chatgpt: Ai-assisted narrative writing for efl college students](#). *Language Teaching Research Quarterly*, 43:1–21.
- Yixin Shen. 2025. [Ai-driven scaffolding and affective support in esl argumentative writing: A multimodal analytics approach](#). In *Proceedings of the 2025 6th International Conference on Education, Knowledge and Information Management, ICEKIM '25*, page 604–608, New York, NY, USA. Association for Computing Machinery.
- Antonette Shibani, Simon Knight, Kirsty Kitto, Ajanie Karunanayake, and Simon Buckingham Shum. 2024. [Untangling critical interaction with ai in students' written assessment](#). In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '24*, New York, NY, USA. Association for Computing Machinery.
- Chaoran Wang, Zixi Li, and Curtis Bonk. 2024. [Understanding self-directed learning in ai-assisted writing: A mixed methods study of postsecondary learners](#). *Computers and Education: Artificial Intelligence*, 6:100247.
- Mengying Wang and Andreas Spitz. 2025. [Quantifying the risks of LLM- and tool-assisted rephrasing to linguistic diversity](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 22561–22574, Suzhou, China. Association for Computational Linguistics.
- Zhijie Wang. 2020. [Computer-assisted efl writing and evaluations based on artificial intelligence: a case from a college reading and writing course](#). *Library Hi Tech*, 40(1):80–97.
- Ruben Weijers, Gabrielle Fidelis de Castilho, Jean-François Godbout, Reihaneh Rabbany, and Kellin Pelrine. 2024. [Quantifying learning-style adaptation in effectiveness of LLM teaching](#). In *Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024)*, pages 112–118, St. Julians, Malta. Association for Computational Linguistics.
- Qimin Xiao. 2024. [Chatgpt as an artificial intelligence \(ai\) writing assistant for efl learners: An exploratory study of its effects on english writing proficiency](#). In *Proceedings of the 2024 9th International Conference on Information and Education Innovations, ICIEI '24*, page 51–56, New York, NY, USA. Association for Computing Machinery.
- Ziyang Xu. 2025. [Patterns and purposes: A cross-journal analysis of ai tool usage in academic writing](#). *Preprint*, arXiv:2502.00632.
- Tae-An Yoo, Min-Kyu Kim, Ho-Suk Yang, and Ji-Bum Chung. 2025. [The impact of ai-assisted writing on students' learning effectiveness and perceptions: A mixed-methods study](#). *SSRN Electronic Journal*. Available at SSRN: <https://ssrn.com/abstract=5396600>.
- Cheng Zhai, Setio Wibowo, and Li Da Li. 2024. [The effects of over-reliance on ai dialogue systems on students' cognitive abilities: a systematic review](#). *Smart Learning Environments*, 11(28).
- Tingting Zhang. 2025. [Evaluating the impact of chatgpt-assisted writing: A controlled study on undergraduate academic compositions](#). In *Proceedings of the 2025 6th International Conference on Education, Knowledge and Information Management, ICEKIM '25*, page 258–262, New York, NY, USA. Association for Computing Machinery.

A System Design Details

A.1 Study Procedure

Participants were recruited via email and asked to complete a preliminary survey administered through Google Forms. Upon receipt of the completed survey, participants were scheduled for an individual, in-person study session.

Each session took place in a private room. Participants were first asked to read and sign the informed consent form (Appendix A.1.1) before beginning the study. Following consent, participants received an email containing two links: one to a document outlining study instructions and another to a Google Doc for composing their essay. Participants assigned to the limited-AI and unlimited-AI conditions additionally received login credentials via email to access the AI-assisted writing platform. The full set of condition-specific instructions is provided in Appendix A.1.2.

All participant data were stored in a private Google Drive folder accessible only to members of the research team. Identifying information was removed from all essays and replaced with anonymized user IDs. During data collection through the LLM interface, only anonymized user IDs were recorded to protect participant privacy.

A.1.1 Consent Form

You are invited to participate in a research project investigating the Overdose Effect of AI-Assisted Writing at the University of Minnesota - Twin Cities. We ask that you read this form and ask any questions you may have before agreeing to take part.

Purpose of the Study: This project explores how different levels of reliance on LLM tools (such as ChatGPT or other generative AI systems) affect students' writing ability, creativity, and critical thinking. The goal is to understand how LLM assistance may influence the quality, style, and originality of written work.

Procedures: The participant will engage in a short preliminary survey asking about their background, experience with AI tools, and comfort level using them. We will then have the participant write a short essay (at least 250-300 words) in response to a provided prompt within approximately 30 minutes. Lastly, the participant will be asked to complete a short post-survey reflecting on their writing process and any use of AI during the activity.

Risks and Benefits of Being in the Study: There

are minimal risks to being in the study. While there is no direct benefit to you, this study will help us better understand how AI tools impact writing skills and academic performance. Each user will be compensated with a \$10 gift card for their time.

Data Sharing: The information you provide through this project will be recorded. Data that you provide as part of this research project will be shared with the other students on the project team, and the course staff.

Confidentiality: We will not collect any information that will make it possible to trace your participation back to you. We will not share your participation with anybody outside of the project team. We will keep your participation private to the extent allowable by law.

Voluntary Nature of Research Participation: Participation in this project is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota - Twin Cities.

Contacts and Questions: The students conducting this research project are: Julia Christenson, Shirley Hayati, and Karin de Langis. The instructor supervising this project is Dongyeop Kang. You may ask any questions you have now. If you have questions later, you are encouraged to contact us. You may keep this page for your records after signing and returning the attached sheet.

Statement of Consent: I have read the attached information regarding the project investigation of the Overdose Effect of AI-Assisted Writing at the University of Minnesota - Twin Cities. I have asked questions and have received answers. I consent to participate in this project.

A.1.2 Instruction Forms

Control Group "A Google Doc will have been shared with you for you to write your essay in. Write a response to the prompt below. Your response should be around 300 words (at least 3 paragraphs), however, you can write more if you want to. Take as much time as you need to write your response.

Prompt: Write a response in which you discuss the extent to which you agree or disagree with the recommendation and explain your reasoning for the position you take. *College students should be encouraged to pursue subjects that interest them rather than the courses that seem most likely to lead to jobs.*

Once you are done writing your response fill out

the following survey. Once you are finished with the essay and the survey please let one of the study team members know."

Limited Group "A Google Doc will have been shared with you for you to write your essay in. Write a response to the prompt below. Your response should be around 300 words (at least 3 paragraphs), however, you can write more if you want to. Take as much time as you need to write your response.

When writing your response, **you are allowed to prompt AI 3 times**. You don't have to use all or any of your allotted prompts; however, you cannot use more than 3. You also cannot prompt the AI to write the full response for you. To access AI, go to this link and sign in with the username and password that the study team member gave you.

Prompt: Write a response in which you discuss the extent to which you agree or disagree with the recommendation and explain your reasoning for the position you take. *College students should be encouraged to pursue subjects that interest them rather than the courses that seem most likely to lead to jobs.*

Once you are done writing your response fill out the following survey. Once you are finished with the essay and the survey please let one of the study team members know."

Unlimited Group "A Google Doc will have been shared with you for you to write your essay in. Write a response to the prompt below. Your response should be around 300 words (at least 3 paragraphs), however, you can write more if you want to. Take as much time as you need to write your response.

When writing your response, **you are allowed to prompt AI as many times as you would like** to help you write your response. However, you cannot prompt the AI to write the full response for you. (You cannot copy and paste in the prompt.) To access AI, go to this link and sign in with the username and password that the study team member gave you.

Prompt: Write a response in which you discuss the extent to which you agree or disagree with the recommendation and explain your reasoning for the position you take. *College students should be encouraged to pursue subjects that interest them rather than the courses that seem most likely to lead to jobs.*

Once you are done writing your response fill out the following survey. Once you are finished with

the essay and the survey please let one of the study team members know."

A.2 Study Interface

The full set of instructions for the limited AI model before the user's prompt is below.

```
SYSTEM_PROMPT = "You are a helpful study assistant. Your role is to provide general guidance, suggestions, and help students think through their work. IMPORTANT RULES: - Keep ALL responses under 100 words - be concise and focused - Provide suggestions, outlines, and general guidance only - DO NOT write complete essays, paragraphs, or full answers that students can copy-paste - Help students develop their own ideas through questions and prompts - Offer structural advice (e.g., "Consider organizing your essay with: intro, 3 body paragraphs, conclusion") - Suggest topics to research or think about - Ask clarifying questions to help students think deeper If asked to write a complete essay or full answer, politely decline and instead offer to: 1. Help them brainstorm ideas 2. Suggest an outline structure 3. Provide tips on how to approach the topic 4. Ask questions to help them develop their thoughts"
```

The interactions with the chatbot and user were sent to a csv file using Vercel Blob. Each row of data followed this order: timestamp, user_ID, user_group, prompt_num, prompt, response.

Preliminary Survey

A. Demographics & Background

Below are the questions included in the preliminary survey instrument.

Q.1 Name (first last):

Q.2 Email:

Q.3 Age:

Q.4 Ethnicity:

Q.5 Year:

- (a) Freshman
- (b) Sophomore
- (c) Junior
- (d) Senior
- (e) Grad student
- (f) Other:

Q.6 Gender:

Q.7 Major/minor/field of study:

Q.8 Category of field of study:

Q.9 Are you a native English speaker?

- (a) Yes
- (b) No

Q.10 Which generative AI tools have you used before? (Select all that apply)

- (a) ChatGPT
- (b) Bard
- (c) Claude
- (d) Co-pilot
- (e) Built-in tools (Grammarly, Word, etc.)
- (f) Other:

Q.11 How often do you currently use AI tools for writing tasks?

(1=Never) 1 2 3 4 5 (5=Always)

B. Self-Assessments

Rate your agreement with the following statements (1 = Strongly disagree, 5 = Strongly agree).

Q.12 I consider myself a strong writer. ()

Q.13 I think AI will improve my writing ability. ()

Q.14 I worry that using AI will make me less capable as a writer. ()

Q.15 I am comfortable using AI tools for drafting, revising, or editing. ()

Q.16 I understand how to prompt AI (I know how to give clear prompts). ()

Q.17 Of your last 10 course-writing tasks, how many used AI in any way? (0-10):

C. Behavioral Baseline & Motivations

Q.18 Why do you use or avoid AI tools for writing? (1-2 sentence open text)

Q.19 Any comments, questions, concerns?

Post-Interaction Survey

A. Usage Log (Only asked to AI groups)

Below are the questions included in the post-interaction survey instrument.

Q.1 Approximately what percent of your final text was directly generated or suggested by AI?

- (a) 0%
- (b) 1–10%
- (c) 11–25%
- (d) 26–50%
- (e) 51–75%
- (f) 76–99%
- (g) 100%

Q.2 Did you copy-paste any AI output into your final submission?

- (a) Yes
- (b) No

Q.3 Did you edit the AI output?

- (a) No edits
- (b) Minor edits (wording)
- (c) Moderate edits (structure)
- (d) Major edits (rewrote most)

B. Process & Experience

Rate your agreement with the following statements (1 = Strongly disagree, 5 = Strongly agree).

Q.4 The AI suggestions helped me generate ideas. ()

Q.5 The AI suggestions saved me time. ()

Q.6 The AI suggestions improved the clarity of my writing. ()

Q.7 The AI suggestions reduced my creative expression. ()

Q.8 I felt I relied too heavily on AI to complete the task. ()

Q.9 I had to do substantial critical thinking to decide what to accept or reject from the AI. . . ()

Q.10 I felt the final product still reflected my own voice. ()

Q.11 I would submit the final essay as my own in a regular class.

- (a) Yes
- (b) No
- (c) Not sure

Q.12 I would disclose to an instructor that I used AI for this piece.

- (a) Yes
- (b) No
- (c) Maybe

C. Cognitive Load / Effort & Time

Q.13 How difficult was it to complete this task? (Very easy) 1 2 3 4 5 (Very difficult)

Q.14 How much of the 30 minutes did you spend editing vs. composing vs. reading AI outputs?

Editing: %
Composing: %
Reading AI outputs: %

D. Reflection on Skills

After this session, indicate how you feel your ability changed in each area (increase / no change / decrease).

Q.15 Organize ideas:

Q.16 Use precise vocabulary:

Q.17 Write creatively:

Q.18 Perform critical analysis:

E. Originality & Authenticity

Q.19 How original do you feel your final essay is? (Not at all original) 1 2 3 4 5 (Very original)

Q.20 On a scale of 0–100, how authentic does the essay feel to you?

Q.21 Would you be comfortable if this exact essay were used as a graded assignment?

- (a) Yes
- (b) No
- (c) Maybe

F. Ethics & Academic Integrity

Q.22 Do you think using AI in coursework should be allowed?

- (a) Yes
- (b) No
- (c) Depends — please explain:

Q.23 If allowed, should students be required to disclose AI usage?

- (a) Yes
- (b) No
- (c) Depends

G. Open Reflection

Q.24 Describe (in 1–3 sentences) how you used AI in this task (what prompts you gave, what you accepted/rejected).

Q.25 What, if anything, did AI prevent you from doing or thinking about?

Q.26 Did anything about the AI outputs surprise you?

- (a) Yes — please explain:
- (b) No

Q.27 Any additional comments about how AI influenced your thinking or writing on this task?

H. Optional Detection / Guessing Task

Q.28 We sometimes run automatic classifiers on texts. Do you think your submitted essay was more likely:

- (a) Human-written
- (b) AI-assisted
- (c) Mostly AI-generated

Essay Grading Rubric

The following rubric was used to evaluate the quality of student essays across all experimental conditions. Each category is scored on a scale of 1–5 points.

Scoring: Each category is scored from 1–5 points. Total possible score for the first six categories: 30 points.

Category	5	4	3	2	1
Focus & Details	Engaging and full development of a clear thesis as appropriate to the assignment purpose.	Competent and well-developed thesis; thesis represents a sound and adequate understanding of the assigned topic.	Mostly intelligible ideas; thesis is weak, unclear, too broad, or only indirectly supported.	Most simplistic and unfocused ideas; little or no sense of purpose.	Ideas are extremely simplistic, showing signs of confusion and misunderstanding of the prompt. The thesis is missing or not discernible.
Ideas, Support, & Development (Evidence)	Consistent evidence with originality and depth of ideas. Ideas work together, and the main points are sufficiently supported with evidence.	Ideas are supported sufficiently; support is sound, valid, and logical.	Main points and ideas are only indirectly supported. Support is loosely relevant to the main points.	Insufficient or non-specific support.	Lack of support for main points.
Organization	Organization is sequential and appropriate. Paragraphs are well developed and appropriately divided.	Competent organization, without sophistication. May lack effective transitions.	Limited attempts to organize around a thesis; paragraphs are mostly stand-alones with weak transitions.	Organization, while attempted, was unsuccessful. Paragraphs were simple, disconnected, and formulaic.	Little to no organization; disjointed or confusing structure.
Voice	The author's purpose is very clear. Their extensive knowledge and/or experience with the topic is/are evident.	The author's purpose is mostly clear. Some evidence of attention to the audience. The author's knowledge and/or experience with the topic is/are evident.	Little or inconsistent sense of the audience related to the assignment purpose. Tone and point of view are not refined or consistent.	Shows almost no awareness of a particular audience; reveals no grasp of appropriate tone and/or point of view for given assignment.	Lacks awareness of a particular audience for the assignment; tone and point of view are somewhat inappropriate or very inconsistent.
Vocabulary and Word Choice	Exceptional vocabulary range, accuracy, and correct and effective word usage.	Good vocabulary range and accuracy of usage.	Ordinary vocabulary range, mostly accurate.	Errors of diction and usage, while evident, do not interfere with readability.	Extremely limited vocabulary; choices lack grasp of diction.
Sentence Structure, Grammar, Mechanics, & Spelling	All sentences are well constructed and have varied structure and length. The author makes no errors in grammar, mechanics, and/or spelling.	Most sentences are well constructed and have varied structure and length. Contains only occasional punctuation, spelling, and/or capitalization errors.	Sentences show a formulaic or tedious sentence pattern. Some errors in grammar, mechanics, and/or spelling, but they do not interfere with understanding.	Most sentences are well constructed, but they have a similar structure and/or length. The author makes several errors in grammar, mechanics, and/or spelling that interfere with understanding.	Sentences sound distractingly repetitive or are difficult to understand. The author makes numerous errors in grammar, mechanics, and/or spelling that interfere with understanding.
AI Detection/ Human-written	I am confident that this text was completely human-written. This essay feels completely authentic and original.	I am confident that this text was written mostly by a human. This essay feels authentic, but may have had some AI assistance.	I believe this essay was done by a human with AI assistance. Some portions of text seem to be AI-generated.	I believe this essay was mostly written with the help of AI. A large portion of the text sounds repetitive and/or unoriginal.	I believe this essay was completely written by AI.