

Examining the Effects of ChatGPT Integration on AP English Writing Proficiency in Williamson County High School Students: A Qualitative, Survey-Based Study

Kanishk Sama

Received June 05, 2024

Accepted August 27, 2024

Electronic access September 30, 2024

This research delves into the intersection of artificial intelligence (AI) and education, addressing a significant gap in the existing literature by examining the impact of ChatGPT utilization on high school students' English proficiency in AP English Writing. Motivated by the increasing prevalence of AI technologies in educational settings and the need for empirical insights, the study posits a hypothesis suggesting a positive correlation between ChatGPT usage and enhanced performance in AP English assignments, as evidenced by elevated assessment scores. To explore this hypothesis, a survey comprising four questions, each answered with a thesis statement, was administered, with respondents responding to the first two questions unaided and receiving ChatGPT assistance for the subsequent two questions. Grading of responses on a scale of 1 to 10 facilitated trend identification, supported by the construction of tables and graphs for visual representation. Analysis of survey data ($n = 18$) revealed a significant 18% increase in scores when ChatGPT was utilized, indicating substantial agreement beyond chance. Moreover, responses improved by ChatGPT exhibited heightened authenticity, correlating with higher scores, which challenges previous assertions regarding the detrimental effects of ChatGPT usage on performance. These findings underscore the potential efficacy of AI-driven tools in educational contexts and emphasize the importance of students' proficiency in interacting seamlessly with such technologies, suggesting avenues for the development of educational programs aimed at optimizing user engagement.

Keywords: Artificial intelligence (AI), ChatGPT, English proficiency, AP English Writing, Educational technology

Introduction

ChatGPT, developed by OpenAI, launched on November 30, 2022, amid widespread anticipation of its transformative potential. However, just five weeks after its release, numerous school districts across the nation began blocking its use in English essays and other writings. This rapid blockage rate prompted questions from school districts nationwide about why ChatGPT was being restricted more quickly than any other site. Educational boards perceived ChatGPT as a threat to learning, creativity, and perseverance¹. However, many of these conclusions were premature due to insufficient research on the effects of AI chatbots on students' performances. Therefore, the research question remains: "What is the impact of ChatGPT utilization on a high school student's English proficiency in AP English Writing?"

To understand the effects of ChatGPT on written aptitude, it is crucial to grasp its functionality thoroughly. ChatGPT operates as an advanced artificial intelligence model trained on diverse datasets to refine its outputs and interactions with users through deep learning, an artificial neural network method. While the concept of neural networks and artificial intelligence models dates back to the early 1900s, what sets ChatGPT apart is its un-

precedented accessibility as a powerful tool available to a broad audience. This accessibility represents a significant evolution in AI technology, allowing individuals of varying backgrounds and skill levels to leverage sophisticated language processing capabilities for educational purposes². Previously, students could access an instructional program called SCHOLAR, which functioned as an AI tutor and employed natural language processing (NLP) technology to comprehend, manipulate, and interpret human language. NLP machines, utilizing neural networks as described by Krogh, allowed SCHOLAR to generate and analyze responses, providing instant feedback on errors and misconceptions³. Contrary to common perception, users of SCHOLAR benefited significantly from its interactive learning approach, which focused on explaining errors to foster student comprehension and growth. However, due to its recent development and limited testing, SCHOLAR remained inaccessible to many students who could have greatly benefited from its educational support³.

While SCHOLAR was primarily designed for academic activities, the use of artificial intelligence and machine learning for English translation has roots dating back to World War II. In 1946, Weaver and Booth embarked on translating German

codes using computer-based applications, pioneering machine translation (MT) technology. Their approach aimed to leverage differences in vocabulary, word order, and syntax to achieve logical translations. However, this early MT system encountered challenges due to lexical ambiguity inherent in natural language, prompting advancements toward more nuanced theories of language, ultimately contributing to the development of generative grammar⁴. Despite its limitations, this historical endeavor highlights the evolving utilization of AI within the realm of English language processing.

Alan Turing, an English mathematician and computer scientist, played a pivotal role in shaping the field of artificial intelligence. Turing is renowned for formulating the Turing test, a benchmark for assessing a computer's intelligence by challenging a human interrogator to discern between computer-generated responses and those from a human subject⁵. Passing the Turing test signifies a machine's capability to exhibit human-like intelligence. Additionally, Turing's contributions extended to practical applications of artificial intelligence, such as developing the Bombe machine to decipher the Enigma code during World War II⁶.

Today, accessing different chatbots on the internet has become effortless for anyone. ChatGPT, an open-source chatbot, allows users to prompt it with diverse questions and engages in interactions that mimic human conversation rather than robotic responses. This capability is enabled by ChatGPT's advanced natural language processing abilities, continuously honed through training and refining its responses based on user input. By leveraging deep learning, a data processing technique inspired by human cognition, ChatGPT analyzes patterns in input text and generates responses in natural language. The widespread adoption of ChatGPT in schools and educational institutions underscores the importance of evaluating its impact on rhetorical writing skills, an area that still needs to be explored despite its prevalent use. Understanding how ChatGPT influences rhetorical abilities is crucial for optimizing its educational applications and ensuring its effectiveness as a tool for improving language proficiency.

Literature Review

Steffen Herbold conducted a comprehensive study comparing human-written and ChatGPT-generated argumentative essays. Herbold collected 90 essay prompts and corresponding human-written essays from Essay Forum, a database for essay questions. Using basic prompts, ChatGPT-3 and ChatGPT-4 responded to the same essay questions. Subsequently, 108 experts evaluated the essays based on seven aspects of argumentative writing and rated their proficiency on a scale of 1 to 5. The results revealed that ChatGPT essays surpassed human-written essays significantly, and the data reliability was high with an inter-rater agreement of 0.9⁷. While this study delves into the skillful-

ness of ChatGPT, it leaves a gap regarding the direct impact of ChatGPT on student writers in the context of argumentative essays.

Yongqiang Ma and colleagues conducted a study evaluating ChatGPT's proficiency in writing scientific research papers. Utilizing "SciBERT," a pre-trained model designed for scientific text, Ma discovered that ChatGPT exhibited a superior understanding of syntactic structure and semantics compared to humans⁸. The study involved scoring scientific abstracts based on metrics such as average word length, frequency, and perplexity. Ma's analysis revealed that ChatGPT tended to be more repetitive and generalized, which could pose challenges for scientific papers requiring specificity. Despite these limitations, Ma suggests that ChatGPT has potential in writing applications, but emphasizes the need for further refinement of large language models (LLMs) to optimize their effectiveness at scale.

Francesco Barbieri investigated the impact of using AI for analyzing tweets. Conducting sentiment analysis across tweets in multiple languages, Barbieri aimed to assess the effectiveness of Twitter deepfakes. Humans typically compose tweets with an "uncurated nature," incorporating elements like misspellings, slang, emoji, and multimodality, which pose challenges for AI-generated deepfakes⁹. Despite these challenges, Barbieri's analysis indicated similarities in sentiment between human and AI-generated tweets, highlighting the capabilities of large language models (LLMs) and natural language processing (NLP). However, Barbieri acknowledges potential biases in his findings due to a custom sentiment analysis program prone to errors and misinterpretations⁹.

In a theoretical study involving English as a Foreign Language (EFL) students, Risang Baskara explored the potential benefits of using ChatGPT to enhance vocabulary, grammar, and syntax skills among students¹⁰. Baskara emphasized the value of repeated interactions between students and ChatGPT in cultivating communication skills and refining English proficiency. However, this research does not directly address the impact of ChatGPT on essay writing. Furthermore, the study's focus on EFL students, rather than high school students pursuing a college curriculum, limits its generalizability. Nonetheless, Baskara's findings align with Herbold's research, underscoring ChatGPT's effectiveness in improving semantic rules, syntactic complexity, and linguistic diversity, all of which are essential skills for EFL learners.

Zana Buçinca conducted a study involving 260 participants to investigate how individuals utilize artificial intelligence (AI) and their attitudes toward AI-generated explanations¹¹. The findings suggest that participants do not typically engage in the analytical reading of AI-generated explanations; instead, they develop general criteria for deciding when to follow the advice provided by AI systems. Despite this, participants demonstrated a willingness to utilize AI for various tasks, indicating a tendency towards overreliance on AI assistance. Interestingly, Buçinca's

research suggests that this overreliance may lead to decreased motivation among users, regardless of the depth or quality of the explanations provided by the AI system. This underscores the potential unintended consequences of excessive dependence on AI, which could impact motivation levels adversely.

Ana Banovac investigated ChatGPT as a writing assistant for argumentative essays written by adults¹². Her study focused on comparing scores before and after using ChatGPT to write essays. Contrary to previous findings, Banovac observed no significant difference in scores between human-written and ChatGPT-assisted essays. Banovac acknowledges that the recency of ChatGPT during the study may have contributed to these comparable scores, as participants were still learning how to use ChatGPT effectively. However, the study's limitation lies in its failure to test high school students, who are more likely to use ChatGPT for academic purposes. Testing high school students at a later stage when they are more familiar with ChatGPT could yield different results.

After reviewing multiple studies on ChatGPT in education, a noticeable gap emerges due to the recent introduction of ChatGPT, resulting in limited research on its effects as an assistant, particularly among high school students¹³. Existing studies often involve toddlers and adults, focusing on neurological changes rather than educational applications relevant to high school students. However, considering that high school students are more likely to use ChatGPT for academic tasks, the implications of this research are pertinent and beneficial to specific populations. Based on this review, I hypothesize that utilizing ChatGPT for AP English essays will positively impact high school students' English abilities and performance, potentially leading to improved scores. I will use an online survey and grade responses to test my hypothesis.

Methods

As mentioned earlier, Ana Banovac conducted a study similar to my research proposal, dividing participants into a control group of nine adults who wrote argumentative essays without ChatGPT assistance and an experimental group of nine adults who used ChatGPT without limitations¹². Each group had 4 hours to complete their essays, which were evaluated by two professors. However, a key limitation of this method is the challenge of directly comparing the effects of ChatGPT by only allowing participants to write with or without its assistance; no study has integrated these two aspects simultaneously. Moreover, the survey described was conducted in person, which is impractical for my research due to potential participant disinterest and resulting impacts on the study's outcome.

The participants of this study were high school students enrolled in AP English Language and AP English Literature because these students would have the greatest use for ChatGPT. The decision to use an online survey methodology through

Google Forms and Google Sheets was guided by several considerations to enhance the comprehensiveness, reliability, and accuracy of the research. Firstly, an online survey format was appealing because of its capacity to reach a wide and diverse audience. This increased reach was deemed crucial for obtaining a representative sample, ensuring that the observations from the study could be extrapolated across a broader spectrum of the student population. Additionally, the online platform provided a convenient medium for participants to engage with the survey, which promoted higher response rates and contributed to the overall validity of the collected data. In the survey, students were tasked with writing four thesis statements to prompts that each had a stimulus (image or quote). For questions with ChatGPT allowed, sample prompts were provided such as "Write me an outline for a thesis statement of a rhetorical essay." Additionally, students were told to write thesis statements in the style they would write for AP English exams. The first two thesis statements were written without the help of ChatGPT. The last two thesis statements were written with the assistance of ChatGPT (this order was used because of other previously researched papers); this meant that improving and writing responses with light involvement from ChatGPT was allowed, but blatant copying was not allowed. In total, the survey response time was estimated to be 20 to 25 minutes. While this may seem short, this focused time period enforced similar AP testing conditions where students were to think on the spot, with and without ChatGPT.

The specific choice of Google Forms and Google Sheets as primary data collection tools was driven by their user-friendly interfaces and seamless integration (see Appendix A for the survey). These widely used platforms facilitated the ease of survey administration and streamlined the organization of responses. The decision to employ these tools was attributed to reduced potential errors in data entry and analysis, given their familiarity and reliability with participants in handling survey-based research, and my expertise with the platforms, which would also reduce errors.

The choice of an online survey methodology using Google Forms and Google Sheets was not only practical but also ethically sound. The choice of an online survey format promoted participant anonymity and confidentiality as no names or email addresses were stored, addressing ethical concerns related to the sensitive nature of writing abilities. Participants engaged with the survey without the fear of judgment, fostering a more open and honest response environment. Additionally, the informed consent process was integrated into the online survey, ensuring that participants were fully aware of the study's objectives and willingly chose to participate.

The creation of a custom rubric for scoring responses was a strategy aimed at introducing a systematic and standardized approach to the evaluation of thesis statements (see Appendix A for rubric). A tailored rubric provided a clear set of criteria,

ensuring consistency in scoring across different evaluators and responses. This choice was motivated by the intention to enhance the validity of the research outcomes, contributing to the overall robustness of the study.

In addressing potential bias in the research, careful consideration was given to the design of essay prompts. The formulation of prompts covering a diverse range of subjects was intentional, to capture a comprehensive range of responses and mitigate bias towards specific topics. While reducing the bias, the diverse spectrum of essay prompts also tested the effectiveness of ChatGPT across different contexts, providing a nuanced understanding of its impact on students' writing abilities.

Alternative methods, such as in-person interviews or paper-based surveys, would not be as effective in capturing an extensive range of responses. In-person methods would introduce social desirability bias, where participants would be pressured into answering what is socially accepted rather than their genuine opinion. Paper-based surveys would lack the efficiency and ease of data management provided by digital platforms. The online survey methodology, therefore, stands out as the robust and ethical approach for exploring the subject at hand, ensuring large-scale data collection while prioritizing participant confidentiality and accessibility.

Furthermore, the decision to involve multiple graders in scoring responses was made to augment the objectivity and reliability of the evaluation process. This collaborative approach sought to foster agreement and validation among graders, reducing the potential for individual bias and outliers and contributing to the overall credibility of the study's findings. The data will be analyzed and organized through linear regressions, tables, and charts. By incorporating these methodological choices, the research strived to establish a well-rounded and rigorous framework for assessing the impact of ChatGPT on students' thesis statement writing capabilities.

Results

A total of 18 respondents participated in the survey over four weeks. The participants' demographics were diverse, including individuals who were juniors and seniors in high school. Two-thirds of the participants were males and one-third were females. The inclusion criteria required participants to have a minimum familiarity with academic writing and thesis statement construction consistent with AP English classrooms (AP English Literature and Composition and AP English Language and Composition). No participant's responses were excluded because every response was complete, ensuring a broad representation of people. The response rate was monitored throughout the survey period, with efforts made to encourage participation and maintain data quality. Participants were informed about the purpose of the study and assured of anonymity and confidentiality in

their responses. These measures aimed to mitigate response bias and enhance the reliability of the gathered data.

Qualitative data collection involved analyzing four thesis statements from each participant, resulting in a total of 72 thesis statements. To assess these statements, a customized rubric was developed based on established criteria for strong, effective thesis statements, encompassing dimensions such as clarity, relevance, alignment, persuasiveness, and grammar. The rubric underwent a rigorous research process involving multiple studies to determine the optimal criteria for grading responses. Experts in academic writing and assessment emphasized the importance of comprehensive, clear, and relevant rubrics. Additionally, analysis of AP English rubrics informed the construction of this rubric, given the similarity of prompts to those in AP English, which were familiar to the participants. An example thesis statement in response to Prompt 3 is provided in Figure 1 (see Rubric and Prompt 3 in Appendix A), illustrating how this response was scored 9 out of 10 based on the rubric criteria. The statement received full points for alignment with the prompt, relevance, persuasiveness, and thesis quality but scored lower for clarity and conciseness due to its wordiness.

Results

A total of 18 respondents participated in the survey over four weeks. The participants' demographics were diverse, including individuals who were juniors and seniors in high school. Two-thirds of the participants were males and one-third were females. The inclusion criteria required participants to have a minimum familiarity with academic writing and thesis statement construction consistent with AP English classrooms (AP English Literature and Composition and AP English Language and Composition). No participant's responses were excluded because every response was complete, ensuring a broad representation of people. The response rate was monitored throughout the survey period, with efforts made to encourage participation and maintain data quality. Participants were informed about the purpose of the study and assured of anonymity and confidentiality in their responses. These measures aimed to mitigate response bias and enhance the reliability of the gathered data.

Qualitative data collection involved analyzing four thesis statements from each participant, resulting in a total of 72 thesis statements. To assess these statements, a customized rubric was developed based on established criteria for strong, effective thesis statements, encompassing dimensions such as clarity, relevance, alignment, persuasiveness, and grammar. The rubric underwent a rigorous research process involving multiple studies to determine the optimal criteria for grading responses. Experts in academic writing and assessment emphasized the importance of comprehensive, clear, and relevant rubrics. Additionally, analysis of AP English rubrics informed the construction of this rubric, given the similarity of prompts to those in AP English,

which were familiar to the participants. An example thesis statement in response to Prompt 3 is provided in Figure 1 (see Rubric and Prompt 3 in Appendix A), illustrating how this response was scored 9 out of 10 based on the rubric criteria. The statement received full points for alignment with the prompt, relevance, persuasiveness, and thesis quality but scored lower for clarity and conciseness due to its wordiness.

Response: *The Skittles advertisement appeals to the audience's ethos through the use of vibrant colors and a sense of fun, establishing credibility and trust in the product. It also appeals to pathos by evoking positive emotions and drawing the audience in with its slogan "Touch the Rainbow, Taste the Rainbow" and the visual of the Skittles packet against the background of a rainbow.*

Fig. 1 A sample thesis statement

Example thesis statements and their corresponding scores are provided in Table 1, demonstrating the rubric's application and scoring methodology. The PO calculated was 0.88, and the Pe was calculated to be 0.55. As a result, inter-rater reliability was assessed using Cohen's kappa coefficient, which indicated substantial agreement beyond what would be expected by chance among raters ($\kappa = 0.73$), similar to Banovac's study, and accepted as a reliable agreement.

	Prompt 1	Prompt 2	Prompt 3	Prompt 4
Response #	Score (0 - 10)	Score (0 - 10)	Score (0 - 10)	Score (0 - 10)
1	7	7	8	7
2	6	5	9	10
3	7	8	9	8
4	3	4	6	7
5	6	7	7	8
6	5	6	8	9
7	7	8	6	7
8	4	4	6	6
9	5	7	7	7
10	6	8	7	10
11	7	7	8	8
12	4	4	6	6
13	6	7	8	7
14	6	5	9	9
15	5	8	9	8
16	5	4	6	7
17	7	7	8	7
18	5	5	9	7
Average Score	5.61	6.17	7.56	7.67

Table: 1 Scores for all responses and average scores for each prompt

In the method section, the quantitative analysis compared average scores of thesis statements with and without ChatGPT assistance. Participants were instructed to write two thesis statements with ChatGPT guidance and two without assistance while formulating their thesis statements. Average scores were then calculated for thesis statements related to each of the four prompts, enabling statistical comparisons to demonstrate the impact of ChatGPT on thesis statement quality (see Figure 2). The data presented in Figure 2 supports the hypothesis.

Additionally, the percentage of authentic text in each response facilitated by ChatGPT was determined using the open-source

website GPTZero, a popular and accurate tool for AI detection, and Turnitin. The AI percentages from both tools returned very similar numbers (within 2% of each other). Response scores were plotted against the percentage of authentic text, and a line of best fit was calculated using Google Sheets (see Figure 3).

The results were presented through a combination of tables and figures to enable a comprehensive analysis. Table 1 illustrates example thesis statements along with their scores based on the tailored rubric, showcasing the scoring criteria and inter-rater reliability. Figure 1 displays the average scores of thesis statements with and without ChatGPT assistance, allowing for further statistical comparisons.

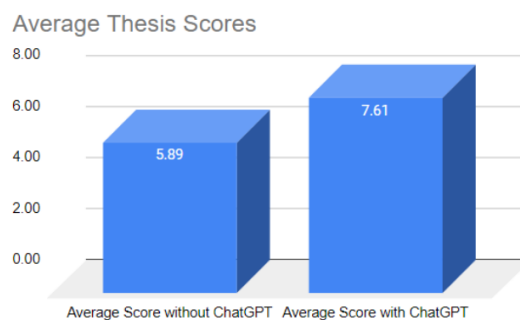


Fig. 2 Average scores for thesis statements with and without ChatGPT

Additionally, scatterplots and best-fit lines (see Figure 3) were used to depict the relationship between scores of responses facilitated by ChatGPT and percent authenticity (100% - AI%), providing insights into the authenticity-performance trade-off when using AI assistance in thesis statement formulation.

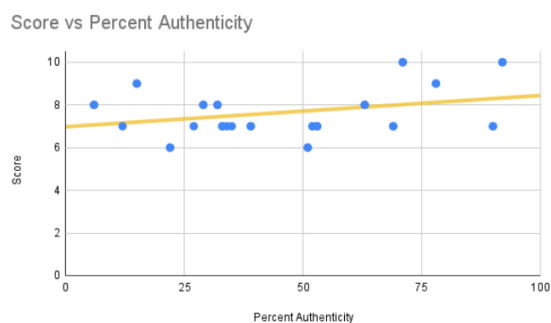


Fig. 3 Relationship between authenticity of text and score received

Observation & Discussion

The analysis of scores for each prompt reveals intriguing insights into the impact of ChatGPT assistance. Prompt 1, focusing on Coca-Cola advertising, had an average score of 5.61 out of 10, with a relatively low standard deviation of 1.19. Prompt 2,

centered on John F. Kennedy's quotes, showed more variability with a standard deviation of 1.54 compared to Prompt 1.

Notably, ChatGPT-assisted statements received relatively higher scores in subsequent prompts. Prompt 3, which explored Skittles advertising, obtained an average score of 7.56 out of 10, with a standard deviation of 1.20. For Prompt 4, discussing Martin Luther King Jr.'s inspirational quotes, the average score was 7.67, with a similar standard deviation of 1.19. These results highlight potential differences in thesis statement quality across different prompts and emphasize the impact of ChatGPT on enhancing scores and response quality.

In Figure 2, the data illustrates a notable divergence in average scores between thesis statements produced with and without ChatGPT assistance, showing a difference of +1.72 points on average. This variance suggests that utilizing ChatGPT positively influenced the quality of thesis statements. The enhancement in quality may be attributed to several factors, including improved content, clarity, structure, grammar, and syntax. ChatGPT's ability to swiftly generate multiple ideas in response to prompts provides students with valuable starting points for developing strong thesis statements, a process that students may find more time-consuming without such assistance.

In Figure 3, the distribution of scores reveals that lower authenticity rates (below 50%) are associated with many scores of 7 and 8. Conversely, responses with authenticity percentages above 50% tended to receive higher scores of 9 or 10. The linear regression analysis, highlighted in yellow in Figure 3, indicates a general trend where higher authenticity rates correlate with higher scores, even with ChatGPT's assistance. This observation suggests that ChatGPT is more effective as an assistant rather than the sole executor of writing tasks. As an assistant, ChatGPT can guide, generate ideas, and improve content quality, but users must strike a balance to maintain authenticity and originality for optimal results. This finding implies that further interaction with ChatGPT may be necessary to refine ideas and produce higher-quality writing while preserving authenticity. Lower authenticity rates may indicate that some students were unsure how to utilize ChatGPT's capabilities to enhance their writing fully.

Additionally, the observed +1.72 score difference signifies augmented learning and skill development among students. Throughout the survey, students not only receive guidance and support from ChatGPT but also actively refine their writing skills based on the feedback and suggestions provided. Furthermore, the correlation between higher authenticity rates and higher scores (Figure 3) reflects a learning process where users recognize the importance of authenticity and strive to be original and genuine in their work. This interpretation highlights ChatGPT's dual role as a writing tool and a platform for learning, skill enhancement, and a deeper understanding of writing principles. However, this score increase could have been caused by familiarity with the prompts or the random variation within

the sample. While the data might not be fully indicative, this interpretation partially underscores the educational value of using ChatGPT as an assistant, promoting continuous learning, skill refinement, improved content quality, and effective communication in rhetorical writing.

In Banovac's findings, she concluded that using GPT as a writing tool did not improve essay quality since the control group (without ChatGPT) outperformed the experimental group (with ChatGPT) in most parameters¹². However, my data contradicts her findings and aligns with many other sources online that suggest ChatGPT as an effective assistant. Banovac attributes her findings to participants' uncertainty in combining their style with generated text, whereas in my research, I have found the opposite to be true, possibly due to the demographic surveyed. Students tend to have an easier time familiarizing themselves with technology and effectively employing it compared to adults. Additionally, Banovac observed a similar score comparison to my data but in the opposite direction. Overall, my research agrees with certain aspects of Banovac's findings but also presents some contrasting results.

Limitations

One primary limitation I encountered was a relatively low number of responses, which was acceptable because several other research studies I encountered had similar participant numbers. Typically, a larger sample size provides more patterns and enables deeper research insights. The limited number of responses constrained the breadth of my conclusions. Difficulty in survey participation at school was another challenge, mainly due to ChatGPT being blocked on school Chromebooks. Consequently, students who participated likely did so independently at home, potentially biasing the sample by excluding individuals who met the criteria but lacked the resources or technology to respond to the survey. Additionally, this limits the generalizability of my findings because there were few participants; the generalizations can only be applied to the sample population and possibly my high school, but nothing beyond that.

Another limitation of the study is the potential inaccuracies in authenticity ratings. ChatGPT detectors can yield false negatives and false positives, which might have skewed the results. Additionally, external factors beyond my control could have influenced the data. Participants may have experienced fatigue or significant distractions during the survey, affecting the accuracy of their responses. Despite employing a tailored rubric and multiple graders, scoring remained subjective. Although I aimed for impartiality in grading, there's a possibility of bias towards certain types of responses. Moreover, students varied in their proficiency with using ChatGPT; some were adept at leveraging its capabilities, while others were still learning. In future studies, it would be beneficial to assess participants' comfort levels with using ChatGPT to inform the analysis.

Conclusions and Future Directions

This study investigated the impact of using ChatGPT on AP English thesis prompts among high school students in Williamson County, Texas. Through a Google Form survey featuring four prompts, the research involving high school students (n = 18) revealed a statistically significant correlation between ChatGPT usage and increased scores. Contrary to some studies suggesting that writing without ChatGPT is more beneficial, my data supports the effectiveness of using ChatGPT, as evidenced by a significant score increase. The study also highlighted the importance of students' ability to interact smoothly with ChatGPT. Figure 2 demonstrated that students who integrated ChatGPT's feedback with their skills achieved higher authenticity rates and scores, emphasizing the need for prompt engineering by students.

However, further research is warranted to validate these findings, given the study's limited scope confined to one high school, with predominantly male participants among other factors. Future studies should consider these aspects to ensure greater accuracy in results. Moreover, this research has implications for ChatGPT's integration into the school system. Rather than advocating for ChatGPT's prohibition, schools should leverage ChatGPT to enhance written English quality. This data suggests that schools and education boards could benefit from investing in ChatGPT training courses. These courses would empower students to interact effectively with LLMs, producing responses that foster skill development, as opposed to relying solely on ChatGPT-generated content.

In conclusion, this study advocates for a proactive approach toward ChatGPT implementation in education, promoting its use as a tool for skill enhancement rather than mere content generation. Further research should explore the efficacy of such programs while addressing the limitations observed in this study, creating a detailed understanding of ChatGPT's role in academic settings.

References

- 1 J. Shen-Berro, *Chalkbeat*.
- 2 A. Krogh, *Nature Biotechnology*, **26**, 195–197.
- 3 C. Guan, J. Mou and Z. Jiang, *International Journal of Innovation Studies*, **4**, 134–147.
- 4 E. Liddy, *Encyclopedia of Library and Information Science*, NY: Marcel Decker, Inc, 2nd edn.
- 5 M. Flasiński, *Introduction to Artificial Intelligence*, 3–13.
- 6 M. Haenlein and A. Kaplan, *California Management Review*, **61**, 5–14.
- 7 S. Herbold, A. Hautli-Janisz, U. Heuer, Z. Kikteva and A. Trautsch, *Scientific Reports*, **13**, 18617.
- 8 Y. Ma, J. Liu and F. Yi, *AI vs. Human – Differentiation Analysis of Scientific Content Generation*, <https://doi.org/10.48550/arxiv.2301.10416>, ArXiv (Cornell University).
- 9 F. Barbieri, L. Espinosa-Anke and J. Camacho-Collados, *XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond*, <https://doi.org/10.48550/arxiv.2104.12250>, ArXiv (Cornell University).
- 10 R. Baskara, *International Journal of Education and Learning*, **5**, 44–55.
- 11 Z. Buçinca, M. Malaya and K. Gajos, *Proceedings of the ACM on Human-Computer Interaction*, **5**, 1–21.
- 12 A. Banovac, I. Kružić, I. Jerković and Bašić, *Humanities and Social Sciences Communications*, **10**, year.

Appendix A

AP Research Thesis Statement Experiment

In this form, you will respond with thesis statements to 4 prompts about rhetorical analysis questions. Some of the prompts will require the help of AI, and some won't. In no scenario will you blatantly copy and paste the sentence of ChatGPT. Please only use it to help you. There is also no need to spend more than 5 minutes on each question because they only require 1 - 2 sentences.

[Redacted]

* Indicates required question

Your participation in this project is voluntary. All of your responses will be anonymous. You may decide not to participate, choose not to answer any question, or stop participating at any time **for any reason** *

I agree

Are you enrolled in an AP English class in Williamson County? *

Yes
 No

What grade level are you in? *

Freshman (9th)
 Sophomore (10th)
 Junior (11th)
 Senior (12th)

What gender are you? *

Male
 Female
 Non-binary
 Prefer not to say
 Other: _____

Next Page 1 of 3 Clear form

First Page of Survey: Demographic Data

Thesis Statements Without ChatGPT

Please spend some time (5 minutes per question at maximum) on each of these prompts and write a well crafted thesis statement to each one without the use of any AI. It should be 1 - 2 sentences in length. Don't use ChatGPT, or any other chatbot at all for these prompts.

Examine this famous advertisement from Coca-Cola. How does the advertisement appeal to ethos, pathos, and logos to the intended target audience?



Your answer

Analyze the persuasive techniques used in the quote below from JFK's inaugural address. How does he employ rhetorical devices to influence the audience's opinion, evoke emotional responses, and impact the overall message/effectiveness of the speech?



Your answer

[Back](#)

[Next](#)

Page 2 of 3

[Clear form](#)

Second Page of Survey: Demographic Data

Thesis Statements With The AID of ChatGPT

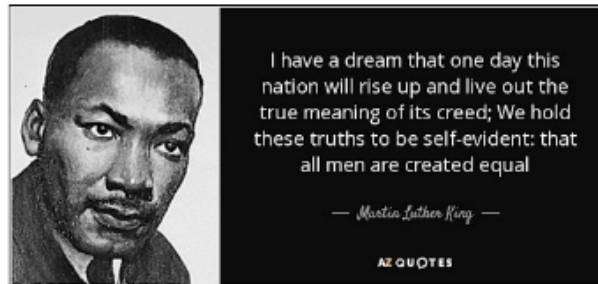
Please spend some time (5 minutes per question at maximum) on each of these prompts and write a well crafted thesis statement to each one with the HELP of ChatGPT (use it to help you write the statements). It should be 1 - 2 sentences in length.

Examine this famous advertisement from Skittles. How does this specific advertisement appeal to ethos, pathos, and logos to the intended target audience? *



Your answer

Analyze the persuasive techniques used in the quote below from Martin Luther King's "I Have A Dream" speech. How does he employ rhetorical devices to influence the audience's opinion, evoke emotional responses, and impact the overall message/effectiveness of the speech? *



Your answer

Back

Submit

Page 3 of 3

Clear form

Third Page of Survey: Demographic Data

Criteria	0 Points	1 Point	2 Points
Alignment with Prompt	The thesis statement does not address the specific prompt.	The thesis partially aligns with the prompt but lacks a clear connection.	The thesis statement directly and fully addresses the provided prompt.
Clarity and Conciseness	The thesis is unclear and overly wordy.	The thesis is somewhat clear but could be more concise.	The thesis statement is exceptionally clear and concise, effectively conveying the intended message.
Relevance to Topic	The thesis statement is not relevant to the topic.	The thesis has some relevance but includes unrelated elements.	The thesis is entirely relevant to the topic and directly addresses the subject matter.
Persuasiveness	The thesis lacks persuasive elements and is unconvincing.	The thesis is somewhat persuasive but could be more compelling.	The thesis statement is highly persuasive, effectively arguing a point of view.
Declarative Statement	The thesis is a question or does not make a clear statement.	The thesis statement is somewhat declarative but may still resemble a question.	The thesis is a strong declarative statement, presenting a clear argument.

Rubric used to grade responses