Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Can GAI improve English writing speed and writing quality? A survey study of college students in China

Bijun HE[a*]
[a]*Joint Education Institute of Zhejiang Normal University and University of Kansas (ZJNUKU), Zhejiang Normal University, Jinhua, 321004, P.R. China*
*hebijun@zjnu.edu.cn

Abstract: This study investigates the role of Generative Artificial Intelligence (GAI) usage on college students' English writing speed and quality. Based on cases from China's higher education system, this study employs a mixed-methods approach with 214 participants. Results demonstrate an increase in writing output (word count) after GAI usage across disciplines, particularly for humanities and social sciences (HSS) students. There is also a decrease in the frequency of grammar errors after GAI usage, especially for non-STEM contexts. In general, GAI usage enhances productivity and accuracy, albeit with disciplinary nuances. The findings show GAI's potential as a pedagogical tool while highlighting the need for tailored integration strategies to address discipline-specific challenges and ethical considerations.

Keywords: Educational Technology; Higher Education; AI for Education; English Writing

1 Introduction

The recent advancement of Generative Artificial Intelligence (GAI) has sparked interest across various fields, including but not limited to technology and education. With digital tools shaping the academic landscape, the potential of GAI to enhance college students' learning and writing has become a topic of growing interests. As a key for students' academic success, writing in a second language, as represented by English, is a Time-consuming and challenging task. With GAI assistance offering the possibility of generating text quickly and improving writing quality, issues arise about their effectiveness in a real-world higher educational setting. Specifically, can GAI improve not only the writing speed of students but also the quality of their writing performance? In China, where education is highly competitive and technology adoption is advancing, this question has gained particular relevance.

As teachers and students seek ways to optimize learning experiences, exploring the impact of GAI on writing presents both practical and theoretical value, leaving it an exciting area of educational research. The primary aim of this study is to re-evaluate the role of GAI in enhancing the writing skills of college students, with a specific reference to the improvements in their English writing speed and quality. Specifically, this study targets at accessing the role of GAI tools, such as ChatGPT and DeepSeek, in assisting students. It will also analyze practices and challenges associated with integrating GAI tools into writing instructions, and provide recommendations for university instructors on how to effectively utilize GAI to benefit students' writing skills.

The implications of this study exist in both educational theory and real-world practice. From a theoretical perspective, the findings will contribute to our understanding of how GAI can be integrated into existing theories of writing instruction and language learning. They may challenge or complement current paradigms, offering new insights into the role of technology in facilitating writing development. In educational practice, the study's recommendations will guide instructors on how to effectively utilize GAI tools, as represented by ChatGPT, to enhance their students' English writing skills. By highlighting workable practices and potential challenges, the findings will help educators adapt their instructional strategies to incorporate GAI in ways that maximize learning outcomes while minimizing the

risks from AI. This study thus has the potential to inform decision makers in terms of writing instruction in the age of artificial intelligence.

2 Literature Review

Recent studies have well highlighted the integration of AI tools such as ChatGPT and DeepSeek into English teaching, including enhancing feedback provision, assessment, instructional design and so on. For instance, Hwang et al. (2024) have explored how learners' prompting behaviors with ChatGPT during writing revisions lack alignment with their objectives. Their findings emphasize the need for instructions on prompt engineering to maximize AI's pedagogical roles. Similarly, Barrot (2023) has shown ChatGPT's potential as a supplementary tool for second language writing, highlighting its capacity to generate instant feedback while cautioning against risks to academic integrity. While language teachers have used AI to automate tasks such as grading, human oversight on AI remains critical. Indeed, by comparing ChatGPT-generated and human feedback on student essays, Steiss et al. (2024) have found that while AI feedback is more consistent, human raters excel in addressing higher-order elements such as content and organization. Additionally, Lee (2024) has demonstrated that structured teacher and peer collaboration with ChatGPT can foster metacognitive awareness and emotional reassurance among English-language learners, proving AI's role as a scaffold rather than a replacement for human guidance. However, ethical concerns persist despite these advancements, including biases in AI-generated content (Santiago de Roock, 2024) and the need for policy frameworks to regulate usage (Playfoot et al., 2024).

AI tools, especially generative AI, are reshaping English learning by offering personalized, on-demand support for learning practice. For instance, Boudouaia et al. (2024) have found that students learning English via ChatGPT-4 show significant improvements, attributing this to the tool's ability to provide real-Time feedback and revisions. Similarly, Yan (2023) has observed that learners perceived ChatGPT as a "digital peer" for drafting and editing, though concerns about over-reliance and academic dishonesty have emerged. Established studies have also highlighted AI's role in fostering motivation. Particularly, Guo and Wang (2024) have reported that ChatGPT-assisted feedback increased student engagement in learning, while Liu et al. (2024) have noted its effectiveness in informal language practice, particularly for oral proficiency development. However, challenges remain in the using of AI in learning. Basic et al. (2023) have found no significant improvement when students used ChatGPT independently, suggesting that unsupervised use may not enhance critical thinking. Furthermore, AI-generated texts often lack the depth of human writing in areas such as cohesion and originality (Zhou et al., 2023). Learners also face cognitive overload when navigating AI tools, as shown by Woo et al. (2024), who have highlighted the need for scaffolding to balance efficiency and comprehension. While AI democratizes access to language resources, its efficacy largely depends on learners' metacognitive strategies and institutional support for ethical integration (Yan, 2024; Tram et al., 2024).

Seminal scholarship has also evaluated the application of GAI (generative artificial intelligence), or more specifically ChatGPT, in English writing. For instance, Zhang et al. (2025) reveal that GAI-based logic learning enhanced the quality of logic in English argumentative writing. However, Hu et al. (2025) concludes that GAI does not significantly group writing performance. Additionally, Kim et al. (2024) argues that GAI can assist writing process, performance, and their affective domain, but their qualitative study is only based on 20 participants. Therefore, we still do not have much quantitative evidence to access the real applications and potential benefits of GAI on English writing, regardless of writing quality or writing speed. As such, a study along this line still remains necessary for current academia.

Specific to studies on the use of AI on college students' writing, gaps remain in evaluating both writing speed and quality. So far, qualitative studies have struggled to effectively assess writing speed, as they are largely limited to subjective interpretations of the writing process while making objective measurement challenging. For the measurement of writing quality, while seminal studies such as Noy & Zhang (2023) have focused on

evaluating writing quality through score-based approaches, this research proposes a different approach by assessing writing quality through the frequency and types of grammar mistakes. This shift in methodology allows for a more granular understanding of GAI's role in the accuracy of writing in real-Time, as opposed to a broader, score-based assessment. By addressing the key research question, "can GAI improve the writing speed and/or writing quality of college students", the research targets at offering a nuanced evaluation of GAI's potential to enhance academic writing, contributing insights for educators and practitioners interested in integrating AI into higher education.

3 Materials and Methods

3.1 Survey

The survey design consists of two main parts: the first gathers demographic details, including participants' gender, grade, age, major, school, birthplace, and frequency of AI tool usage, to contextualize their backgrounds and familiarity with technology. The researchers performed coding to quantify these information, to support further correlation analysis and regression analysis.

The second part of our survey involves a comparative English writing test, where participants are divided into two groups. The treated group completes two one-minute writing tasks (one without AI assistance [Group A1] and another with AI support [Group A2]) to measure the role of AI on writing performance under time constraints. Meanwhile, the control group performs two consecutive unassisted one-minute writing tasks (Group B1 and Group B2) to establish a baseline for natural variation or practice effects. This structure aims to isolate the influence of AI assistance on writing outcomes while accounting for individual differences and potential confounding factors.

3.2 Variables

The variables in this study encompass a mix of categorical, ordinal, and continuous measures designed to analyze the role of AI assistance on writing outcomes. Categorical variables, such as Major (STEM vs. humanities/social sciences), University (science/technology-focused vs. general/HSS institutions), and AI (assisted vs. unassisted writing), are represented as dummy variables (0/1) to enable direct comparisons between groups. Province, though originally a 23-category geographic variable, is treated as a set of dummy variables in analysis to isolate regional effects. Ordinal variables like Grade (1–5, reflecting academic year) and Time (first vs. second test) capture ordered distinctions without assuming equal intervals. Continuous variables, such as Word (word count) and Error (grammar error percentage), quantify writing productivity and accuracy. Together, these variables allow researchers to disentangle how AI assistance interacts with academic background, demographic factors, and task sequence to shape writing performance, while controlling for potential confounders like practice effects or regional differences.

In total, this study variables, analyzed in a sample of 214 participants (with 207 valid responses after exclusions), encompass demographic, academic, and task-specific measures to assess AI-assisted writing outcomes. Grade (mean = 2.46, SD = 1.08) reflects participants' academic year (1–5), while Major (34% STEM, coded as 1; 66% HSS, coded as 0) and University (29% science/technology-focused institutions, coded as 1) capture academic backgrounds. Province, a categorical dummy variable (1–23), represents geographic origins, with a mean of 7.84 (SD = 6.01). The Time variable distinguishes the first (1) and second (2) writing tasks (mean = 1.50, SD = 0.50). The binary AI variable (23% AI-assisted tasks, coded as 1) contrasts writing conditions, while Word (mean = 30.29, SD = 46.40) quantifies writing output and Error (mean = 14%, SD = 13%) measures grammar accuracy via Grammarly. These variables collectively evaluate how AI assistance is associated with writing productivity and accuracy, controlling for academic, demographic, and task-related factors (Table 1).

Table 1. *Descriptive Statistics of Major Variables: Full sample*

	N	Minimum	Maximum	Mean	SD
Grade	214.00	1.00	5.00	2.46	1.08
Major	214.00	0.00	1.00	0.34	0.47
University	214.00	0.00	1.00	0.29	0.46
Province	214.00	1.00	23.00	7.84	6.01
Time	214.00	1.00	2.00	1.50	0.50
AI	214.00	0.00	1.00	0.23	0.42
Word	211.00	1.00	393.00	30.29	46.40
Error	207.00	0.00	0.63	0.14	0.13
Valid N (listwise)	207.00				

3.3 Estimation

Our sample is divided into two groups, i.e., the treated group and the control group. Each group is tested twice:

Treatment Group 1: Writing one minute without AI assistance (AI = 0; Time = 1)
Treatment Group 2: Writing one minute with AI assistance (AI = 1; Time = 2)

Control Group 1: Writing one minute without AI assistance (AI = 0; Time = 1)
Control Group 2: Writing one minute without AI assistance (AI = 0; Time = 2)

The estimation strategy is an ordinary Least Squares (OLS) regression analysis which controls three fixed-effects of grade, major, and university:

$$Y_{rt} = \beta_0 + \beta_1 AI_{rt} + \beta_2 I(Time_{rt} \geq 2) + \gamma_1 + \gamma_2 + \gamma_3 + \varepsilon_{rt}$$

Where β_0 is the intercept, β_1 and β_2 are the coefficients of two explanatory variables. ε_{rt} is the error term. γ_1 is the grade fixed effect, γ_2 is the major fixed effect, and γ_3 is the university fixed effect. Here, Y can be either the number of words written or the percentage of grammar errors. We also test this regression model in two sub-groups

4 Results

4.1 Correlation Analysis

The correlation matrix reveals several significant relationships among the variables in the full sample. Grammar error frequency is negatively correlated with AI usage (r = –0.351, p < 0.001), while the number of words written (r = 0.647, p < 0.001) is positively correlated with AI usage, suggesting that AI treatment may reduce grammar errors while advancing writing speed. In addition, Grades have a significant negative correlation with AI usage (r = –0.324, p < 0.001) and the number of words written (r = –0.185, p = 0.007), implying that higher reliance on AI or longer submissions may relate to lower grades. These findings highlight complex interdependencies between academic performance, writing behaviors, and technological assistance.

4.2 Number of Words Written

The regression analysis in Table 2 examines key factors associated with the number of words written in the full sample. The results indicate that AI usage is the strongest and only statistically significant predictor of word count, with a substantial positive coefficient (Estimate = 72.089, p < 0.001), suggesting that AI-assisted writing is linked to significantly

longer submissions. In contrast, all other variables, including Grade level, Major, and University, show no significant association with word count (all $p > 0.05$). The model explains a moderate proportion of the variance in word count (Multiple $R^2 = 0.425$, Adjusted $R^2 = 0.402$), indicating that while AI usage is a major driver, other unmeasured factors may also play a role. The intercept term is statistically insignificant (Estimate = 8.098, $p = 0.395$), reinforcing that AI is the dominant predictor. Overall, these findings suggest that AI tools may encourage more verbose writing, whereas traditional factors like academic level or institutional affiliation is not significantly associated with text length in this sample.

Table 2. *Key Factors Associated with the Number of Words Written and the Percentage of Grammar Errors: Full Sample*

Response Variable	# Words Written (N = 207)	% Grammar Errors (N = 207)
(Intercept)	8.098 (9.491)	0.239*** (0.033)
AI	72.089*** (7.908)	–0.094*** (0.027)
Time	0.840 (6.066)	–0.041. (0.021)
factor(Grade)2	9.239 (6.849)	–0.012 (0.024)
factor(Grade)3	6.192 (7.494)	–0.014 (0.026)
factor(Grade)4	5.441 (8.062)	–0.058* (0.028)
factor(Grade)5	0.646 (26.489)	0.164. (0.091)
factor(University)1	-3.504 (5.709)	–0.001 (0.020)
Adjusted *R*-squared	0.4019	0.1504
'log Lik.'	'log Lik.' -1050.221 (*df* = 10)	log Lik.' 144.2583 (*df* = 10)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

4.3 Number of Grammar Errors

The regression results in Table 2 also analyze factors associated with the percentage of grammar errors in the full sample. The intercept is statistically significant (Estimate = 0.239, $p < 0.001$), representing the baseline error rate when all predictors are zero. AI usage shows a strong negative association with grammar errors (Estimate = –0.094, $p = 0.001$), suggesting that AI-assisted writing significantly reduces error rates. Among Grade levels, only Grade 4 demonstrates a significant reduction in errors compared to the reference category (Estimate = –0.058, $p = 0.038$), while Grade 5 shows a marginally positive but non-significant trend toward higher errors (Estimate = 0.164, $p = 0.072$). Major and University affiliations play no significant role on error rates (both $p > 0.05$). Overall, the findings suggest that AI tools improve grammatical accuracy, while additional time and higher grade levels (particularly Grade 4) may also contribute to reduced errors.

5 Conclusions and Future Research Directions

The study reveals that GAI tools boost writing productivity of higher education students, with AI-assisted tasks yielding longer writings (e.g., +72.089 words, $p < 0.001$ in the full sample) and fewer grammar errors (e.g., −0.094 error percentage, $p = 0.001$). However, such roles of GAI vary by discipline. Indeed, HSS students benefit more from AI in error reduction, while STEM students perform differently after GAI usage.

However, there are also limitations within our findings. The first obvious limitation of this study lies in sample representativeness. Most participants of Group 1 come from Zhejiang Province of China, while participants of Group 2 come from different provinces across China. This would partly have an influence on the representativeness of different provinces. Another limitation comes from the students' speeds and ways of typing in English. Indeed, while Chinese students may not be quite familiar with typing in English, this study has not taken into consideration of typing speed and customs of typing.

Such limitations, especially regional sampling bias (overrepresentation of Zhejiang Province) and unmeasured typing proficiency, call for broader and more diverse cohorts. Future studies should explore longitudinal role of AI on writing development, incorporate typing-speed controls, and examine cross-cultural comparisons beyond China. Additionally, qualitative insights into student and instructor perceptions could clarify ethical and pedagogical concerns that quantitative metrics alone cannot capture, such as over-reliance or authenticity.

Reference

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, 100745.

Basic, Z., Banovac, A., Kruzic, I., & Jerkovic, I. (2023). ChatGPT-3.5 as writing assistance in students' essays. *Humanities & Social Sciences Communications, 10*(1), 750.

Boudouaia, A., Mouas, S., & Kouider, B. (2024). A study on ChatGPT-4 as an innovative approach to enhancing English as a foreign language writing learning. *Journal of Educational Computing Research, 62*(6), 1509–1537.

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*(7), 8435–8463.

Hu, W., Tian, J., & Li, Y. (2025). Enhancing student engagement in online collaborative writing through a generative AI-based conversational agent. Internet and Higher Education, 65, 100979.

Hwang, M., Jeens, R., & Lee, H.-K. (2024). Exploring learner prompting behavior and its effect on ChatGPT-assisted English writing revision. *Asia-Pacific Education Researcher*. Advance online publication.

Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on Generative AI-assisted academic writing. Education and Information Technologies. Advance online publication.

Lee, Y.-J. (2024). Can my writing be polished further? When ChatGPT meets human touch. *ELT Journal, 78*(4), 401–413.

Liu, G., & Ma, C. (2024). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching, 18*(2), 125–138.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187-192.

Playfoot, D., Quigley, M., & Thomas, A. G. (2024). Hey ChatGPT, give me a title for a paper about degree apathy and student use of AI for assignment writing. *Internet and Higher Education, 62*, 100950.

Santiago de Roock, R. (2024). To become an object among objects: Generative artificial intelligence, writing, and linguistic white supremacy. *Reading Research Quarterly, 59*(4), 590–608.

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction, 91*, 101894.

Tram, N. H. M., Nguyen, T. T., & Tran, C. D. (2024). ChatGPT as a tool for self-learning English among EFL learners: A multi-methods study. *System, 127*, 103528.

Woo, D. J., Wang, D., Guo, K., & Susanto, H. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. *Education and Information Technologies*. Advance online publication.

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies, 28*(11), 13943–13967.

Yan, D. (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology, 28*(1), 1–19.

Zhang, R., Zou, D., & Cheng, G. (2025). ChatGPT affordance for logic learning strategies and its usefulness for developing knowledge and quality of logic in English argumentative writing. *System, 128*, 103561.

Zhou, T., Cao, S., Zhou, S., Zhang, Y., & He, A. (2023). Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. *System, 118*, 103141.