Proceedings of ICLLCD 2026 Symposium: Literary and Linguistic Perspectives on Cultural Narratives
DOI: 10.54254/2753-7064/2026.NS36691

Teacher Strategies for Reducing AI Generated Content in English Major Student Writing

Yuehan Hou

Department of Foreign Languages, Xinjiang Normal University, Urumqi, China
3887147461@qq.com

Abstract. Generative AI can produce fluent prose before students have worked through the linguistic choices that writing courses are intended to teach. This study examined how one university instructor responded to that problem in a second-year English-major writing course. Forty-two students (18 men and 24 women; mean age = 20.3 years) participated in a 16-week concurrent embedded mixed-methods study. The instructor compared AI-probability estimates from GPTZero, Originality.AI, and Turnitin, returned passage-level feedback, and taught four sessions on making context-sensitive lexical choices without asking AI to rewrite students' work. Questionnaires, interviews with six students, three sets of writing samples, and course scores supplied the data. Across Weeks 5, 10, and 15, the mean estimated share of AI-generated text fell from 51.2% to 35.8% and then to 23.6%. Overall AI-tool use also declined, while mean writing scores rose from 72.4 to 85.9. Students most often associated AI use with difficulty expressing ideas accurately, limited time, and pressure to obtain higher grades. By the end of the semester, more students distinguished grammar assistance from having AI compose a paper. Because the study involved one class and no control group, these changes should be read as associations rather than proof of a causal effect. Even so, the findings illustrate how transparent checking, individualized revision, and lexical instruction can be combined in a classroom response to AI-dependent writing.

Keywords: generative AI, English major writing, teacher strategies, academic integrity, AI-generated content

1. Introduction

For English majors, a written assignment is both a finished product and a record of language practice. Generative AI unsettles that relationship. A student can now obtain a coherent essay within seconds, yet the resulting fluency says little about whether the student selected the words, organized the argument, or worked through uncertainty in English. The issue is therefore not simply whether a text contains machine-produced sentences. It is whether repeated dependence on those sentences displaces the independent expression and critical judgment that an English writing course is expected to develop.

A blanket prohibition offers no complete answer. AI tools can support legitimate activities such as checking grammar or exploring alternative wording, and their availability is unlikely to diminish.

DOI: 10.54254/2753-7064/2026.NS36691

The instructional problem lies in drawing a workable boundary between assistance and substitution. When a tool supplies a paper's ideas and language, students avoid the very decisions through which writing competence grows. Teachers consequently need practices that make this boundary visible while still helping students address the linguistic difficulties that lead them to seek automated writing in the first place.

Existing scholarship has approached AI use in student writing mainly from two directions. One line of work evaluates whether detection systems can distinguish human writing from machine-generated prose [1, 2]. A second documents students' patterns of use and their often uncertain understanding of academic integrity [3, 4]. These studies clarify the scale of the problem and the limits of detection, but they offer less detail about what an instructor can do inside a writing course before inappropriate use becomes an assessment issue. Classroom evidence on teacher-led prevention remains comparatively limited.

The present study follows one instructor's semester-long response. Instead of treating a single detector score as a misconduct decision, she compared estimates from three tools, identified passages for discussion, and asked students to revise those passages themselves. Four vocabulary sessions addressed a recurring reason for AI dependence: students' difficulty finding precise expressions. The study examined changes in the submitted writing together with students' views of these practices. It was guided by the following questions:

RQ1: To what extent did the proportion of AI-generated text in students' writing change during the 16-week intervention?

RQ2: How did students perceive the instructor's use of multiple detection tools and vocabulary-focused guidance?

2. Methods

2.1. Participants and instructional setting

The study took place in one intact second-year English-major class at a comprehensive university in eastern China. The 42 participants included 18 men and 24 women. Their mean age was 20.3 years ($SD = 0.8$), and their CET-4 scores ranged from 425 to 580, reflecting the variation commonly found in an intermediate university class. The course instructor was a woman with more than eight years of English-writing teaching experience and a master's degree in Applied Linguistics from a top-tier Chinese university.

A concurrent embedded mixed-methods design was used to place students' reported experiences alongside changes in their written work. Quantitative evidence came from questionnaires, detector estimates, and writing scores. Interviews and open-ended questionnaire responses provided qualitative evidence about students' reasons for using AI and their interpretation of the instructor's approach.

2.2. Intervention procedures

The course and intervention ran for 16 weeks. The first practice concerned how AI presence was estimated. Each paper was checked with GPTZero, Originality.AI, and Turnitin. The instructor used the median of the three reported percentages rather than relying on any one score, an approach intended to limit the influence of single-tool false positives [5]. No individual detector result was treated as conclusive on its own.

DOI: 10.54254/2753-7064/2026.NS36691

The second practice turned the reports into revision guidance. The instructor marked passages with comparatively high AI probabilities, discussed why those passages appeared machine-like, and required students to restate the content in their own words. Detection was thus used during feedback rather than only after a paper had been graded.

The third practice consisted of four vocabulary sessions in Weeks 4, 7, 10, and 13. Students examined basic or overused wording and considered more precise alternatives in context. The instructor did not ask an AI system to perform the substitution. Instead, she demonstrated how meaning, register, and sentence context should guide a lexical choice so that students could apply the procedure independently in later assignments.

2.3. Data collection and analysis

At the beginning of the semester, students completed an anonymous questionnaire covering frequency and forms of AI use, reasons for using AI in writing, ethical views, and self-assessed writing ability. The questionnaire was administered again at the end of the course. Six students - two from each of the high-, middle-, and lower-proficiency groups - took part in semi-structured interviews before and after the intervention. The interviews explored their motives, their understanding of acceptable assistance, and their responses to the instructor's practices.

Three major assignments, submitted in Weeks 5, 10, and 15, were retained for text analysis. Each was processed by the same three detection tools, and the median percentage was recorded for each student at each time point. Descriptive statistics summarized the questionnaire and writing measures. A one-way repeated-measures ANOVA tested change in estimated AI-generated content across the three assignments. The reported chi-square and paired-samples t tests were used for pre-post comparisons, while interview transcripts and open-ended responses were coded thematically for RQ2.

3. Results and discussion

3.1. RQ1: reduction in AI presence

The measures moved in the same general direction over the semester (Table 1). Reported use of AI tools declined from 78.60% before the intervention to 45.20% after it ($\chi^2 = 8.42$, $p < .01$). The proportion of frequent users fell from 32.10% to 12.80%, whereas the proportion reporting no AI use rose from 21.40% to 41.50%.

Table 1. Changes in AI use, AI-generated content, and writing performance across the intervention

Measure	Time point/group	Value (M ± SD or %)	Statistic	p value
Overall AI-tool use	Pre-intervention	78.60%	$\chi^2 = 8.42$	< .01
	Post-intervention	45.20%		
Frequent AI users	Pre-intervention	32.10%	-	-
	Post-intervention	12.80%		
Never-users of AI	Pre-intervention	21.40%	-	-
	Post-intervention	41.50%		
Estimated AI-generated content rate (%)	Week 5 (Time 1)	51.2 ± 18.4	F(2, 82) = 24.37	< .001

DOI: 10.54254/2753-7064/2026.NS36691

Table 1. (continued)

	Week 10 (Time 2)	35.8 ± 15.2	$\eta^2 = .37$	All pairwise: $p < .01$
	Week 15 (Time 3)	23.6 ± 12.1		
Overall writing score	Pre-intervention	72.4 ± 8.2	t(41) = 8.56	< .001
	Post-intervention	85.9 ± 7.1	18.7% increase	
Vocabulary subscore gain	Pre-post difference	4.2 points	Largest gain across dimensions	-

Detector-based estimates from the three assignments showed a progressive decrease. The mean estimated AI-generated share was 51.2% (SD = 18.4) in Week 5, 35.8% (SD = 15.2) in Week 10, and 23.6% (SD = 12.1) in Week 15. The repeated-measures ANOVA was significant, $F(2, 82) = 24.37$, $p < .001$, $\eta^2 = .37$, and all pairwise comparisons were reported at $p < .01$. Mean writing scores also rose from 72.4 (SD = 8.2) to 85.9 (SD = 7.1), $t(41) = 8.56$, $p < .001$, an increase of 18.7%. Vocabulary contributed the largest dimensional gain, with a pre-post increase of 4.2 points. These results describe change within the class; they do not by themselves isolate the intervention from other influences during the semester.

3.2. RQ2: student perceptions and effects on practical judgment and critical thinking

Questionnaire and interview responses help explain why students used AI and how they reacted to the course. Before the intervention, the most common motives were difficulty expressing an intended meaning accurately, pressure to finish work quickly, and the desire for a higher grade. These motives suggest that AI use was connected not only with convenience but also with students' confidence in their own English.

Students did not immediately welcome the use of three detectors. One interviewee reported being annoyed at first but later interpreted the comparison as an effort to be fair rather than punitive. Another said that seeing three reports made the passages judged to be machine-generated more visible and reduced room for dispute. Although the tools could not establish authorship with certainty, showing more than one estimate appeared to make the feedback process easier for students to understand.

Responses to the vocabulary sessions were more consistently positive. A lower-proficiency student explained that limited expressive resources had previously led her to AI; after the sessions, she began keeping a vocabulary notebook and relied less on automated wording. A higher-proficiency student valued the instructor's explanation of why a word suited one context better than another, rather than a list of supposedly advanced substitutes. The emphasis on contextual choice gave students a procedure they could reuse when revising independently.

A change was also recorded in ethical judgment. At the start of the course, 26.2% of students regarded generating an entire paper with AI as academic misconduct. By the end, the proportion was 64.3% ($\chi^2 = 11.24$, $p < .001$). Students increasingly separated limited assistance from the outsourcing of ideas and composition. One student summarized that distinction directly: "I now understand that asking AI to polish my grammar is different from asking AI to write my ideas." The shift indicates a more differentiated understanding of academic integrity, even though views were not uniform across the class.

DOI: 10.54254/2753-7064/2026.NS36691

3.3. Comparison with previous research and pedagogical contributions

The baseline pattern resembles earlier reports of frequent AI use and uncertain ethical boundaries among university writers [3, 4]. The present data add a classroom sequence to that descriptive literature. During one semester, lower detector estimates occurred alongside reduced self-reported use, stronger writing scores, and clearer distinctions between assistance and substitution. The design does not permit a firm causal claim, but the convergence of these measures makes the instructional process worth closer examination.

First, comparing tools gave the instructor more information than a single report. That choice responds to work based on one detector [1] and to wider evidence that AI-detection systems can produce both false positives and false negatives [6, 7]. The median still cannot serve as ground truth, and the findings should not be read as validation of the three products. Its practical value here was narrower: an extreme result from one system did not determine the feedback a student received.

Second, the instructor used flagged passages as material for formative revision. This differs from an administrative-surveillance model in which detection is used only to identify a possible violation [8]. It also fits calls for assessment that makes students' thinking and revision processes visible rather than judging only the submitted product [9]. Research reviews often emphasize the difficulty of detecting AI-written work [2]; the present case suggests that the same reports can be handled as prompts for a conversation about writing, provided their uncertainty is acknowledged.

Third, the vocabulary sessions addressed a reason for dependence rather than only its visible outcome. Wen [10] argues that language education in the AI era must give greater attention to judgment and thinking. In this course, that aim took a concrete form: students compared lexical options, considered context, and made the final selection themselves. The approach is consistent with studies linking AI-assisted writing to students' low confidence in their expressive resources [11, 12]. It may help explain why the vocabulary score showed the largest gain and why students described less need for automated phrasing.

Taken together, the practices can be understood as prevention through empowerment rather than detection through surveillance. The contrast is relevant to discussions of both opportunity and risk in language assessment [13]. It also accords with arguments that academic-integrity education should build learners' knowledge and skills instead of relying only on punishment [14, 15]. In the present class, transparency about detection, opportunities to revise, and explicit language instruction worked as connected elements rather than as separate policies.

4. Conclusion

This study documented how one English-writing instructor combined three detector reports, passage-level revision, and four vocabulary sessions over a 16-week course. The estimated AI-generated share of student papers declined across the three assignments from 51.2% to 35.8% and then to 23.6%. At the same time, overall AI use was reported less often, the mean writing score increased by 18.7%, and more students identified full-paper generation as misconduct. The interviews suggest that transparent feedback and attention to students' expressive difficulties were important to how the practices were received.

For teaching, the findings favor a response that separates appropriate assistance from replacement of the writer and gives students a practical route toward more independent work. Multiple detector outputs may support discussion when their limitations are made explicit; vocabulary instruction can address the uncertainty that prompts some students to outsource phrasing; and assessment can attend to drafting and revision rather than only the final text. These implications concern instructional use

DOI: 10.54254/2753-7064/2026.NS36691

and should not be taken as evidence that detector percentages provide definitive judgments of authorship.

Several limitations narrow the conclusions. All participants came from one class at one university, one instructor delivered the intervention, and no comparison group was available. The observation period was limited to a single semester, so the durability of the changes is unknown. Studies involving multiple institutions, different instructors, comparison conditions, and follow-up data are needed to test whether the same pattern appears in other settings and persists over time.

References

[1] Flitcroft, M. A., Sheriff, S. A., Wolfrath, N., et al. (2024). Performance of artificial intelligence content detectors using human and artificial intelligence-generated scientific writing. *Annals of Surgical Oncology, 31*(10), 6387–6393.

[2] Han, J., Yang, Y., & Liu, G. (2025). Are teachers assessing work written by students or by AI? A rapid literature review of research on detecting content generated by generative AI. *European Journal of Education, 60*(1), e70240.

[3] Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239.

[4] Zhao, X., & Li, M. (2024). Chinese university students' perceptions and use of generative AI in academic writing: A mixed-methods study. *Journal of Educational Technology & Society, 27*(4), 112–128.

[5] Wu, R. T., & Dang, R. R. (2023). ChatGPT in head and neck scientific writing: A precautionary anecdote. *American Journal of Otolaryngology Head and Neck Medicine and Surgery, 44*(6), 103980.

[6] Saha, S., & Feizi, S. (2025). AI detection tools: Systematic testing and reliability concerns. *University of Maryland Technical Report*.

[7] Weber-Wulff, D., et al. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*(1), 1-17.

[8] Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era. *Journal of University Teaching & Learning Practice, 20*(2), 1-24.

[9] Kickbusch, S., et al. (2025). The next era of assessment: Key touchpoints in AI-free, AI-assisted, and AI-integrated assessment designs. *Digital Education Council*.

[10] Wen, Q. F. (2024). Foreign language teaching in the age of artificial intelligence: Challenges and opportunities. *Foreign Language World*(1), 2-8.

[11] Malik, J., et al. (2023). Student perceptions of AI in academic essay writing. *Journal of Educational Technology, 15*(4), 223-245.

[12] Xiao, Y., & Zhi, Q. (2023). ChatGPT as a learning partner: Student perspectives on AI-assisted writing. *Computer Assisted Language Learning, 36*(8), 1-25.

[13] Roe, J., Renandya, W. A., & Handoyo, P. (2023). The role of AI in language assessment: Opportunities and threats. *Language Assessment Quarterly, 20*(3), 245-262.

[14] Pecorari, D., & Petrić, B. (2014). Plagiarism in second-language writing. *Language Teaching, 47*(3), 269-302.

[15] Zhang, X. (2024). Solution-oriented versus problem-oriented approaches to academic integrity. *Journal of Academic Ethics, 22*(1), 45-62.