

Harness the Generative AI tools for Automated Feedback in EAP Writing Assessment

Mengqi Hu* and Rui Xu

English Language Centre, School of Languages, Xi'an Jiaotong-Liverpool University, China

ARTICLE INFO

Keywords:

*Automated Feedback,
Generative Artificial
Intelligence (GenAI),
Essay Writing,
EAP Assessment*

ABSTRACT

This study explores the impact of five generative AI tools (Grammarly AI Plus, XIPU AI, Kimi, Prepostseo, Magic School AI) on English for Academic Purposes (EAP) writing assessment at a Sino-British EMI university. Thirty-seven first-year students wrote 250-word drafts, received AI-generated feedback, and submitted revised versions. Study 1 employed surveys and interviews, revealing that students valued the AI tools for their speed and grammar accuracy. Study 2 analyzed draft and revision scores in Excel, showing improvements in task response and essay structure. However, gains in grammar and vocabulary were limited, which contrasted with student perceptions. The findings indicate that AI feedback can effectively supplement traditional assessment methods, supporting learner autonomy and enhancing certain aspects of writing. Overall, the study highlights the potential of AI tools to aid EAP writing development, while also emphasizing the need to understand their limitations in improving language accuracy.

1. Introduction

Generative AI (GenAI) is increasingly embedded into English for Academic Purposes (EAP) writing. It offers various potentials, such as enhanced efficiency and effective, immediate and personalized feedback, with numerous studies exploring its applications and implications (Liu et al., 2025; Ngo & Hastie, 2025; Steiss et al., 2024). Some popular applications like Chat GPT and Grammarly can help students identify common errors to improve grammar and vocabulary, thereby supporting formative assessment practices (Liang et al., 2024; Kohnke et al., 2025). However, integrating GenAI into academic writing also raises the risk of plagiarism and over-reliance (Thaweesak, 2025). There are still critical concerns about academic integrity and the need for AI literacy training, among which the accuracy and reliability of such AI-generated feedback still need to be carefully evaluated to ensure that it provides meaningful and actionable feedback (Marzuki et al., 2023; Mabrito, 2025; Steiss et al., 2024).

Teachers' perceptions of GenAI tools in academic writing are generally mixed, with acknowledgement of their instant feedback and support for students, and concerns about AI's hindrance to autonomous writing skills (Nicolas, 2024; Saleh, 2024). While students

* Corresponding author's E-mail address: mengqi.hu@xjtlu.edu.cn

Cite this article as:

Hu, M., & Xu, R. (2025). Harness the Generative AI tools for Automated Feedback in EAP Writing Assessment. *European Journal of Teaching and Education*, 7(4): 29-47. <https://doi.org/10.33422/ejte.v7i4.1659>

© The Author(s). 2025 **Open Access**. This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author(s) and source are credited.



generally demonstrate positive perceptions towards the time-efficient AI-generated feedback (Wang, 2024). For example, according to Cao (2025), students appreciate the reduced writing anxiety and increased confidence associated with AI-generated feedback. However, some students expressed difficulties in understanding the AI-generated feedback and raised concerns about the reliability of AI feedback, particularly in terms of coherence and cohesion (Hou et al., 2024; Sogut, 2024).

Future research should establish comprehensive guidelines and frameworks to ensure teachers' and students' responsible and effective integration of GenAI tools into academic writing practices. This need drives further empirical study projects on GenAI tools' automated feedback on EAP writing assessment. Therefore, it is worthwhile to conduct this study to explore how to harness GenAI tools to support EAP writing assessments, thereby improving the overall learning experience for students and reducing the workload for educators.

2. Overview of Relevant Literature

2.1. The Theoretical Underpinnings of Academic Writing Feedback

2.1.1 Theory of Effective Feedback

Understanding of good feedback in academic writing has its theoretical foundation mostly in Hattie and Timperley's (2007) classic work *The Power of Feedback*. Though its effects might be positive or negative, this thorough framework shows that feedback is among the most effective factors affecting learning and achievement (Hattie & Timperley, 2007). The model's power is in its methodical two-dimensional approach to understanding feedback efficacy: function and level.

Three basic questions must be addressed by effective feedback according to Hattie and Timperley's framework: Where am I going? (feed up), How am I going? (feedback), and where to next? (feedforward). Moreover, the model separates four different degrees of feedback: self-level feedback, process-level feedback, self-regulation feedback, and task-level feedback. While higher levels cover more intricate cognitive and metacognitive processes, the first level, task-related feedback, reveals whether a learning task has been successfully grasped and carried out.

Empirical research has shown how immediately relevant this model is for university settings. Through a case study focusing specifically on applying the Hattie and Timperley feedback model in thesis evaluation in higher education, LipschWijnen and Dirkx (2022) provided evidence of its significance for postgraduate academic writing situations. Their study proves that the model's framework can efficiently examine and improve written feedback techniques in college settings, making it rather pertinent for grasping how AI-created feedback would operate within academic writing evaluation.

2.1.2 Technology Acceptance Model (TAM)

Initially created by Davis (1989), the Technology Acceptance Model (TAM) offers a vital theoretical viewpoint for grasping how users accept technology advancements in educational environments. Perceived usefulness and perceived ease of use are the two main variables the model suggests that determine possible users will accept a computer system. Perceived usefulness refers to the degree to which a person believes that using a particular system would increase their job performance; perceived ease of use refers to the degree to which a person thinks that using a particular system would be free of effort (Davis, 1989).

The concept of perceived usefulness is fundamental in educational environments when looking at student and teacher acceptance of AI-powered writing feedback technologies. According to the model, users' perceptions of how these tools will improve their academic performance and learning results directly affect their willingness to use and continue employing such systems. This theoretical base offers a methodical grasp of how stakeholders welcome GenAI tools in academic writing situations.

2.2. Current Applications and Cases of GenAI in EAP Writing Feedback

2.2.1 GenAI in Language Learning

Especially Large Language Models (LLMs), the development of generative artificial intelligence signals a disruptive technological change in language instruction and learning. GenAI has had worldwide appeal since OpenAI debuted its ChatGPT in November 2022, driven by the generative pretrained transformer3 (GPT3) big language model (OpenAI, 2023). To forecast the following word and improve by human feedback, LLMs such as GPT4—the most recent in the GPT family—are trained on huge internet text. Models resulting from this process, able to generate coherent, grammatically accurate text and carry out several natural language jobs, have significant ramifications for language education (Law, 2024).

2.2.2 GenAI Tools in Academic Writing Contexts

Current studies highlight the significant influence of GenAI technologies on teaching and learning in academic writing. Platforms such as Grammarly AI, Kimi AI and ChatGPT are reshaping writing instruction by providing tailored feedback, grammatical corrections, and content generation. These tools generally function as collaborative writing tools that are used to analyze drafts, to offer real-time feedback, and to support revision in EAP study.

One of the popular writing assistant tools, Grammarly AI merges traditional grammar checks with AI-driven insights into clarity, tone, and stylistic coherence. By delivering contextual suggestions and explanatory notes, it aids learners in refining their academic writing in real time. Research indicates students view Grammarly as effective for improving academic writing, with high ratings for usefulness and ease of use across various departments (Epe et al., 2024).

Kimi AI, created by Moonshot AI, is among China's most widely adopted AI platforms. It handles text, images, and code as a multimodal system, with key features including document processing, long-context analysis, and cross-format problem-solving. A recent study by Yao et al. (2025) investigated how secondary school English teachers in China use AI-guided chatbots, like Kimi, to provide feedback on student writing. This implies that teachers align teaching pedagogy with AI tools to improve the feedback process.

2.3. Stakeholders Evaluation and Impact of GenAI Feedback

2.3.1 Students' and Teachers' Perceptions

Both students and teachers hold varied perceptions of AI-generated feedback compared to human feedback, with some studies indicating positive evaluations and others highlighting challenges or preferences for human input.

Studies report that students rate AI-generated feedback useful and engaging. Jansen et al. (2024) observed that AI feedback was useful for 59% of texts, and Escalante et al. (2023) noted a near-even preference between AI-generated and human feedback. Mahapatra (2024) documented significant improvements in academic writing skills, while Gozali et al. (2024)

linked AI use to enhanced feedback literacy. Some studies point to challenges such as occasional inconsistencies in response quality and sensitivity to prompt design.

A significant benefit highlighted by teachers is the reduction in routine grading and feedback tasks. Integrating AI can alleviate their feedback workload, especially in large class settings, by reviewing linguistic issues like vocabulary and grammar (Lee & Moore, 2024). Teachers also reported enhanced communication and cognitive/emotional support for students. Findings from Lee and Moore (2024) suggest that AI-generated feedback can be more comprehensive and detailed, particularly in linguistic and solution-oriented aspects, compared to their feedback. Teachers have also reported gaining novel insights from the interaction with AI.

Despite these benefits, teachers express concerns. These include occasional inconsistencies in response quality and a dependence on prompt quality (Kurt & Kurt, 2024). Teachers have expressed reservations about the lengthiness of some AI-generated feedback, which might be challenging for lower-proficiency students to comprehend. AI's lack of familiarity with specific classes and students is a significant concern, potentially leading to inappropriate feedback (Xu & Tan, 2024).

2.3.2 Perceived Usefulness: AI Vs Human Feedback

The perception of AI feedback usefulness is influenced by several contextual factors, such as the educational and cultural contexts (e.g., Saudi vs. Chinese EFL learners), the type of AI tool used (e.g., ChatGPT, Kimi), the nature of the writing task (e.g., narrative essays, argumentative writing), and the user experience, including the ease of use and comprehensibility of the feedback. For example, a study of Saudi EFL learners reported an overall positive attitude toward ChatGPT-generated feedback, with no significant difference in effectiveness from teacher feedback, contrasting with findings from Chinese contexts (Alsofyani & Barzanji, 2024).

Teachers regard AI-generated feedback favorably when it aligns closely with human feedback. Jürgensmeier and Skiera (2024) reported a high correlation ($r = 0.94$) between AI and human feedback in marketing analytics, and Pahi et al. (2024) highlighted improvements in technical accuracy and clarity. Benefits also include reducing routine grading tasks (Lee & Moore, 2024), although concerns about inconsistency and diminished depth in processing persist.

2.4. Purpose of Study

The increasing application of GenAI tools by students in their writings makes it urgent to investigate into their feedback provision in the EAP writing tasks in higher education field. While these technologies contribute to continuous and timely support during the writing process, their pedagogical impact on students' EAP writing remains insufficiently evaluated. Given that EAP writing constitutes a critical yet demanding academic skill, this study aims to investigate students' and teachers' perceptions regarding AI-generated feedback and evaluate the effectiveness of AI-generated feedback in enhancing specific academic writing competencies. Furthermore, it seeks to provide empirical evidence to refine the feedback frameworks to balance innovative affordances with academic integrity.

- **RQ 1:** What are the perceptions of students and teachers regarding AI-generated feedback on EAP writings?
- **RQ 2:** Does the use of AI-generated feedback lead to improvements in students' writing?

3. Research methods

3.1. Method

Study 1 explored the first research question using a quasi-experimental essay writing experience. Study 2 investigated the second research question through an anonymous and standardized assessment of the revised final drafts.

3.2. Participants

Both studies were conducted at a Sino-British University during the summer of 2024. 37 participants who were year 1, year 2 and year 3 Chinese students volunteered to participate in the study.

In study 1, a total of 37 students, composed of 15 males and 22 females, whose ages ranged from 18 to 22 years, and who received 5 written feedback files on their first draft from AI tools. The tools used were Magic School (<https://www.magicschool.ai/>), Kimi (<https://www.kimi.com/>), Grammarly (<https://app.grammarly.com/>), Prepostseo (<https://www.prepostseo.com/>) and XIPU AI (<https://xipuai.xjtlu.edu.cn/v3/index?v=3>). An online survey was conducted on Wenjuanxing (<https://www.wjx.cn/>), followed by a semi-structured focus group interview to gauge students' perceptions regarding the feedback generated by those 5 tools.

To address the second research question, in study 2, two female EAP teachers, both with roughly a decade of experience teaching EAP in the same institution, marked the students' revised final drafts and provided band scores based on the standardised task sheet and rubrics.

3.3. Instruments and Procedure

In study 1, to explore students' perceptions regarding the AI-generated feedback, two investigators created a writing coursework task sheet (see Appendix A) based on the essay prompt 'Benefits and drawbacks of integrating Gen AI in Higher Education' and 5 recommended sources was provided. Participants were required to write a 350-word discursive essay based on the provided prompt and integrate at least two of the provided sources in the university required referencing system. After participants completed the first drafts, a Wenjuanxing questionnaire was developed to gather quantitative data (see Appendix B). Six multiple-choice questions captured various dimensions of feedback preference, including AI feedback features, the most helpful areas, satisfaction with AI feedback improvement and two questions to identify participants' understanding of the key components of essay writing.

After participants submitted the final drafts, they were grouped into 3 or 4 student groups for a semi-structured focus group interview (see Appendix C). They were interviewed by some semi-structured focus group questions to provide an explanation for their choice and preference towards AI-generated feedback tools. The investigators used Tencent Meeting platform (<https://meeting.tencent.com/>) to record and automatically transcribe the videos into Microsoft Word files.

In study 2, to assess whether AI-generated feedback led to any improvement in students' writing, two investigators marked the 37 first drafts and final drafts on the same criteria of task (30%), organisation (30%), vocabulary (20%), grammar and mechanics (20%). They used Excel to record the band score for each criterion and to calculate the total score.

3.4. Analysis

Data analysis was systematically organised to address both studies. For quantitative data, questionnaire results were analysed in Wenjuanxing platform’s embedded statistical analysis functions, enabling further interpretation to explore relationships between variables like AI feedback satisfaction and revision frequency. As for focus group interviews that were conducted via Tencent Meeting, the interview scripts were transcribed in the Tencent Meeting platform’s built-in functions and then imported into software MAXQDA 24.9 version for thematical analysis, though manual coding was prioritized to maintain contextual nuance. Researchers identified recurring themes through inductive coding, cross-referencing with quantitative findings to triangulate results. Two investigators verified the themes to ensure validity.

First and final draft scores were recorded and calculated by two investigators in Excel files, where each band criterion scores were listed under predefined rubrics for content, structure, and language accuracy (Appendix A). Excel’s data sorting and visualisation tools facilitated comparative analyses across different bands, highlighting trends in score changes. This mixed-methods approach integrated platform-specific functionalities with analytical techniques, adhering to academic standards for data transparency and validity.

4. Results

4.1. RQ 1: What are the perceptions of students and teachers regarding AI-generated feedback on EAP writings?

Table 1. Student perceptions of AI-Generated feedback from questionnaire

Item	n	p	Mean	Deviation
Strongly agree that AI feedback improved writing	15	40.54%	4.05	0.89
Agree that AI feedback improved writing skills	19	51.35%	3.86	0.72
Total positive responses	34	91.89%	3.95	0.76
Find AI feedback helpful for grammar, syntax, and vocabulary	28	75.68%	3.78	0.65
Consider immediacy as the most helpful feature	24	64.86%	3.65	0.59
Consider objectivity as the most helpful feature	22	59.46%	3.55	0.62
Likely or very likely to use AI feedback in the future	32	86.49%	3.87	0.71
Positive Perception of AI Feedback (Grammar Accuracy)	13	65.00%	3.65	0.69
Positive Perception of AI Feedback (Vocabulary Choice)	11	61.11%	3.55	0.61
Feedback Components - Overall Feedback Quality	15	40.54%	3.35	0.56

As illustrated in Table 1 and Figure 3, students’ perceptions regarding AI-generated feedback are generally positive. Specifically, 40.54% (n = 15) of the students strongly agreed, 51.35% (n = 19) agreed that AI feedback significantly improved their writing skills, leading to a total of 91.89% (n = 34) positive responses. This result indicates that most participants claim that they benefited from AI-generated feedback in their writing process. Furthermore 75.68% (n = 28) of the respondents highlighted their preferences of grammar, syntax, and vocabulary suggestions from AI-generated feedback. These items’ mean and deviation values were 3.75 and 0.65 respectively, indicating a strong positive perception in this study.

Further analysis of students’ perceptions on certain feedback components also provided valuable insights on the correlation between their perceived importance and their perceived usefulness. Among the 20 participants (54.05%) who considered grammar accuracy to be a key feedback component, 13 (65.00%) had a positive perception of AI-generated feedback, with a mean of 3.65. Even though students perceive idea development and essay structure as the key factors when writing their first drafts, they still ranked vocabulary choice and

grammar accuracy as the most significant features in written feedback (Figure 1 and Figure 2).

The cross-analysis of AI feedback perceptions by draft difficulty level revealed some interesting patterns. Twenty participants (64.52%) held a positive view of AI feedback, with a mean of 3.65. In contrast, only 6 participants found the writing task to be complicated. This suggests that students who perceive drafting as more challenging may be slightly less receptive to AI feedback.

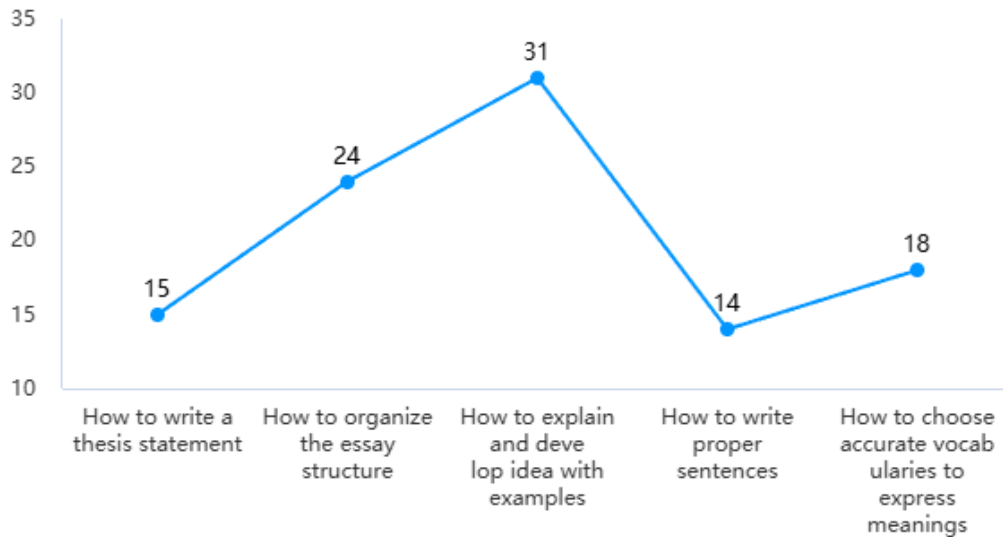


Figure 1. Which three factors do you consider are the most important ones when you are writing the first draft?

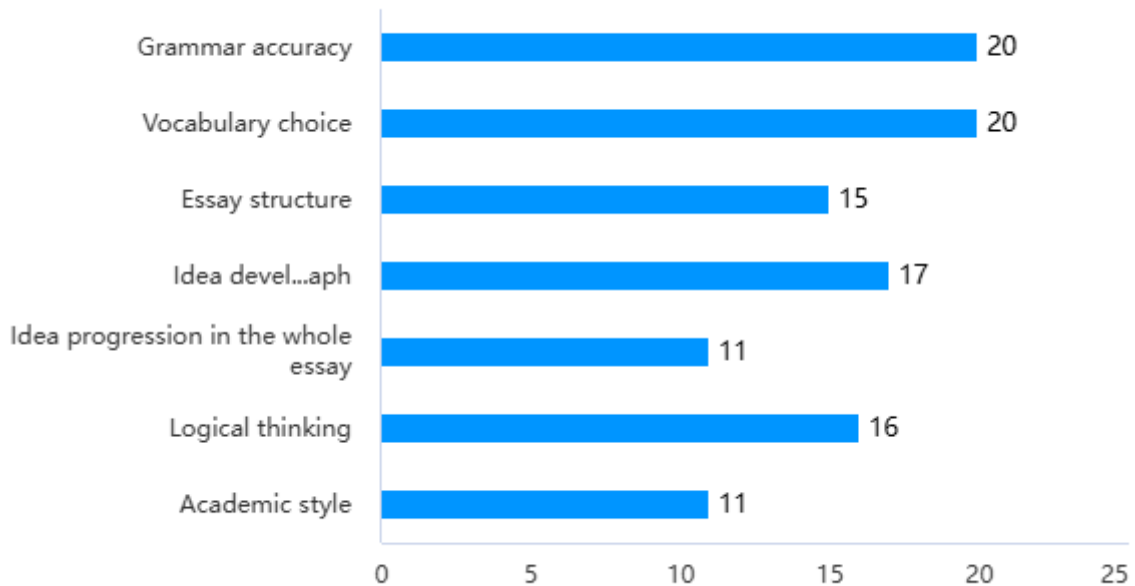


Figure 2. What do you think are the three important components of written feedback?

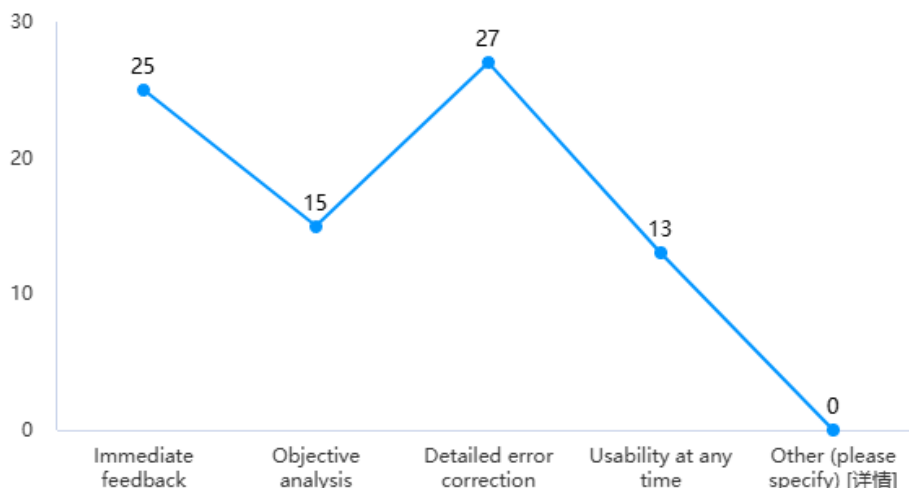


Figure 3. Which features of AI feedback do you find most helpful?

Table 2. Summary of themes retrieved from focus group interviews

Theme	AI's Role	Key Student Expectation
Efficiency	Technical assistant	Accuracy and speed for grammar vocabulary check
Inadequacy for Higher-Order Issues	Limited resource	Guidance on logic, in-depth development and creativity
Tool Preferences	Reliable tool	Context-aware, actionable suggestions
Complementary to Teachers	Secondary support	Human insight for complex needs
Accurate Feedback	Unreliable for context	Error-free, relevant feedback
Emotional Impact	Demotivating tone	Encouraging, personalized support

Table 3. Tool preferences

Tool	Perceived Strengths	Perceived Weaknesses
Grammarly	Detailed grammar/vocabulary fixes; actionable	Overwhelming minor corrections
Magic School	Holistic scoring (e.g., structure, language)	Generic advice; repetitive feedback
XIPU AI	Context-aware academic style suggestions	Inconsistent accuracy
Kimi	Organizational feedback	Vague; lacks implementation guidance

Based on the focus group interview statistics in Table 2, six core themes that have high recurring frequency emerged regarding students' perceptions of AI-generated feedback. This result matches with the questionnaire finding about participants valuing AI for timely and accurate corrections of grammar, vocabulary, and mechanics (e.g., punctuation, and citations). Some participants made positive comments, “Grammarly points out the types of mistakes and tells me how to correct it”, “AI can detect grammar errors and suggest synonyms, but lacks context”, while some made neutral comments, “AI feedback is mechanical—it fixes errors but doesn’t teach principles”. Overall, participants identified that AI-generated feedback is often generic and vague, with a lack of high-order issues. Such feedback also fails to address argument depth, originality, or discipline-specific nuances. For example, some students commented, “AI repeats my ideas but doesn’t guide deeper development”, and “Kimi tells me to ‘improve coherence’ but never explains how”. Further varied perceptions towards AI tools can be found in Table, “Grammarly is precise; Magic School’s scoring helps me prioritize”, Kimi is useless—it diagnoses problems but offers no solutions”. While others compared AI with Human teachers, “AI can’t assess creativity or emotional depth—only teachers can”. Such comparisons between AI and human feedback

indicate that AI feedback does not replace human teachers in logic development and emotional support.

“Teachers understand my voice and provide tailored logic fixes—AI can’t.”

“AI checks grammar; teachers fix my ideas and critical thinking.”

“Teachers encourage me; AI feels transactional.”

“Teachers provide encouragement, build confidence, and reduce anxiety.”

“Teacher feedback includes empathy—they know my struggles.”

“AI’s cold tone makes me doubt my work; teachers reassure me.”

Overall, the results indicate that participants generally hold positive perceptions of AI-generated feedback in EAP writing. These results are similar to what Burner, Lindvig & Wærness (2025) find that students generally appreciate AI-generated feedback’s immediacy and objectivity. However, participants also made negative comments such as AI feedback are usually mechanical and lacking more profound teaching. Unlike teachers—who offer tailored logic guidance, emotional support, and encouragement essential for developing critical thinking, confidence, and personalized feedback—AI often provides generic, vague advice that overlooks higher-order issues like argument depth and originality (Wang & Huang, 2025).

RQ 2: Does the use of AI-generated feedback lead to improvements in students’ writing?

In study 2, two investigators marked the first and the final drafts, and compiled the scores for each band, which are task, organization, vocabulary, grammar and mechanics. This study employs descriptive statistics and paired-sample t-tests in SPSS 28 to quantitatively analyze writing score improvements between first and final drafts among 34 students. Three samples were removed after 3 participants withdrew from the study. Results demonstrate significant improvements after feedback revision ($t^* = 10.06$, $p^* < 0.0001$), with an average gain of 11.35 points (SD = 6.58). The effect size (Cohen’s $d^* = 1.73$) indicates a large practical significance, confirming the efficacy of AI-generated feedback in academic writing.

Table 4. Comparison analysis of component scores between first draft and final draft

Component	First Draft (Mean)	First Draft (SD)	Final Draft (Mean)	Final Draft (SD)	Improvement (Δ)
Task (30%)	19.41	2.89	22.65	2.62	+3.24
Organization (30%)	19.41	2.89	22.65	2.62	+3.24
Vocabulary (20%)	12.94	1.93	14.88	1.75	+1.94
Grammar (20%)	12.94	1.93	14.88	1.75	+1.94

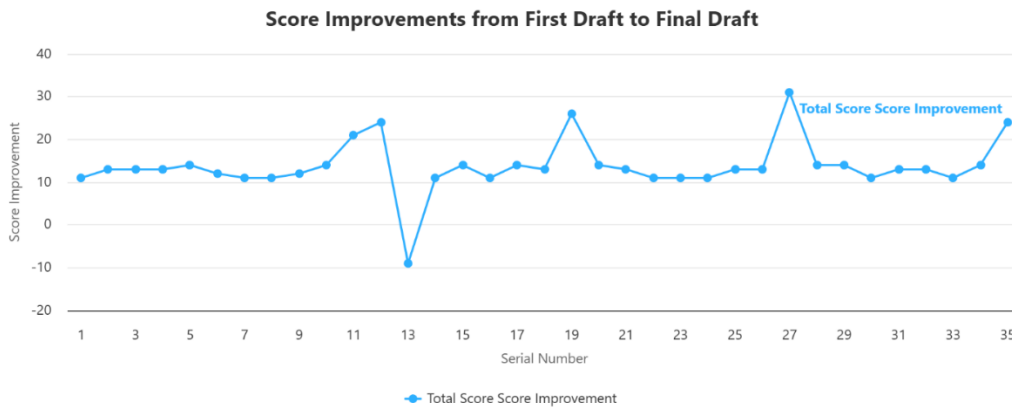


Figure 4. Score improvements from 18 participants

In Figure 4, The x-axis displays participants' serial numbers to display score results anonymously. The y-axis represents the score difference between the first and final drafts. The overall trend demonstrates that improvement in writings, with most improvements clustering around 13 points. While Table 4 reveals an in-depth improvement between higher-order and lower-order writing skills. Task Completion and Organization (macro-skills) demonstrated substantially greater gains (+3.24 points each) compared to Vocabulary and Grammar (micro-skills, +1.94 each). This significant improvement indicates that AI-generated feedback could address organizational errors where students corrected 42% more macro-level errors. Conversely, micro-skills remain challenging, particularly in vocabulary and grammar accuracy, where error rates remained 26% higher post-revision. By comparing standard deviations between macro-skills and micro-skills, it is surprising that AI-generated feedback enhanced participants' task and organization rather than their vocabulary and grammar. Even though there is grammatical improvement in subject-verb agreement, vague word scores stalled near 15/20. This contradictory result shows that although AI-generated feedback could potentially strengthen foundational skills, deeper learning barriers such as student limited vocabulary size and lack of clear expression in precise words may lead to unclear idea development.

5. Discussion and Implications

Study 1 revealed students' perceptions regarding AI-generated feedback for EAP writing. The results indicate that sample students value the feedback immediacy, grammar accuracy and vocabulary choices. These results align with prior research conducted by Hyland (2021), which shows that AI tools like Grammarly enhance learners' accuracy in grammar, syntax, and vocabulary. These tools also reflect students' practical needs for timely, data-driven corrections. However, some negative comments on AI feedback for being "generic" and failing to address argument depth or originality align with concerns about AI system's lack of the contextual understanding to evaluate disciplinary-specific rhetorical strategies (Knoch et al., 2020; Andrea & Halim, 2024). For instance, participants noted tools like Kimi provided vague prompts, which aligns with Wang and Wen (2023)'s finding about AI tools' incompetency in developing logical reasoning.

Some interesting findings about tool preferences highlight the participants' emphasis on grammar accuracy and desire for higher scores through more positive comments on Grammarly's detailed grammar fixes and Magic School's holistic scoring. At the same time, XIPU AI and Kimi faced criticism for inconsistent accuracy or vague suggestions. These findings indicate a gap for an advanced AI tool that could offer context-aware, discipline-specific feedback (Lee & Wolfe, 2022). Another interesting finding is participants' preference for human feedback over AI feedback for emotional support, academic confidence and critical thinking (Hyland, 2021). These findings help educators to reinvent the feedback mechanism in the era of AI to better engage students with human feedback in higher education EAP contexts.

Study 2 compared participants' first draft and final draft scores to identify whether there is improvement. There is a significant improvement in essay quality, as Chen and Parameshachari (2023) have demonstrated that students who received feedback from LLMs showed improved accuracy in their essay-writing tasks and overall essay readability. While GenAI feedback tends to be effective for basic writing skills, it may be less effective for higher-order skills such as logical development and structure (Zhang et al., 2025). There are several possible reasons about the AI-generated feedback's limited impact. Two frequently mentioned factors are lack of personalization and lack of contextual understanding.

According to Yu and Canton (2020), AI-generated feedback frequently lacks depth or personalization compared to provide students with meaningful and individualized guidance. This limitation in tailored support is particularly problematic for developing higher-order skills in EAP writing assessments, which often depend on personalized instruction. As for contextual understanding, GenAI tools often fail to comprehend the wider context of a student's essay, which is a critical element for delivering feedback that fosters critical thinking and creativity (Hou et al., 2024). Although GenAI can offer corrective feedback on language use, it may not fully grasp the creative or critical aspects of a student's writing (Wang, 2024).

Universities are increasingly recognizing the importance of integrating GenAI literacy into their graduate-attribute frameworks to prepare students for a technology-driven future. The European Commission (2022) recommends that higher education institutions develop AI-specific learning outcomes that balance technical proficiency with ethical reasoning. Drawing on this, EAP curricula needs to include AI literacy and ethical decision-making skills to ensure that graduates are equipped with both technical skills and critical thinking abilities necessary for the complexities of the modern job market (Siau & Wang, 2024). Successful integration of GenAI also requires faculty development and support. Studies on GenAI literacy support highlight the need for enhanced digital literacy among educators and creation of supportive frameworks to facilitate technological innovation while respecting the unique contributions of human educators (Haroud & Sagri, 2025). The adoption of GenAI in higher education varies globally, with different regions emphasizing various aspects of AI integration and GenAI literacy. For example, Asian universities focus on generated content and maintaining academic integrity, while addressing gaps in policy-making and resource allocation (Bearman et al., 2024). Some challenges such as ethical concerns over data privacy and academic integrity and the need to balance traditional educational methods with AI-driven innovations require strategic curricular transformation, comprehensive policies and ongoing support for both students and educators (Haroud & Sagri, 2025; Valdez & García-Peñalvo, 2025).

6. Conclusion

To effectively incorporate AI-generated feedback into EAP writing assessment, EAP teachers need to engage in professional training about GenAI systems and empirical studies to understand GenAI tools' capabilities and limitations. This could enable educators to integrate those AI tools with pedagogical caution and expertise when applying them to assist students' writing process. Additionally, adopting a hybrid feedback model that combines AI feedback and human feedback can address both basic and higher-order writing skills. This could be time-efficient for both teachers and students while providing timely feedback. Teachers should also emphasize the higher-order skills such as critical thinking and creativity by showcasing such flaws in AI-generated feedback as negative exemplars, thereby discouraging students' over-reliance on GenAI tools.

EAP teachers should revise and refine the academic integrity policy in EAP writing by guiding students on the ethical use of AI tools, stressing the importance of originality and authenticity in their work. Furthermore, project-based learning and peer assessment could be embedded into the EAP assessment to shift from evaluating an end-product to evaluating the whole learning process. Teachers should also encourage students to build their digital literacy by promoting transparent communication about AI tool usage and facilitating a more inclusive learning atmosphere. Ultimately, EAP teachers can help students understand the

strength of AI-generated feedback and why teachers' personalized guidance still matters for individual learning.

In conclusion, while several explorations into stakeholders' perceptions of GenAI feedback potentials exist, this empirical study compares perceptions with results by analyzing students' perceptions of AI-generated feedback and students' improvements between first drafts and final drafts in academic essay writing. As GenAI continues to be developed and applied in the sphere of EAP learning and teaching, it becomes imperative for students and teachers to utilize its strength and reconceptualize what critical thinking, higher-order skills entail in the era of artificial intelligence. Future research should investigate the long-term impact of AI feedback on writing development, explore optimal combinations of AI and human feedback, and examine cultural variations in AI feedback acceptance across different educational contexts.

Funding

This study was funded by SURF (2024-0217) at Xi'an Jiaotong-Liverpool University.

References

- Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research*, 63(2), 431–463. <https://doi.org/10.1177/07356331241307297>
- Andrea, A., & Halim, S. (2024). The Impact of AI-Powered Software on Second Language (L2) Writing: A Systematic Literature Review. *Research and Innovation in Applied Linguistics-Electronic Journal*, 2(2), 138–138. <https://doi.org/10.31963/rial.v2i2.4801>
- Bearman, M., Ryan, J., Ajjawi, R., & Tai, J. (2024). Artificial intelligence assessment in higher education: Rapid review. *Journal of Higher Education Policy and Management*, 46(5), 523–538. <https://doi.org/10.1080/17516234.2024.2379070>
- Burner, T., Lindvig, Y., & Wærness, J. (2025). “We should not be like a dinosaur”—Using AI technologies to provide formative feedback to students. *Education Sciences*, 15(1), 58. <https://doi.org/10.3390/educsci15010058>
- Cao, T. (2025). Undergraduate English majors' views on ChatGPT in academic writing: Perceived vocabulary and grammar improvement. *FWU Journal of Social Sciences*, 19(1), 1–11. <https://doi.org/10.51709/19951272/Spring2025/1>
- Chen, J., & Parameshachari, B.D. (2023). *Integration of Big Data and Artificial Intelligence in Constructing Learners' Individualized Feedback System*. 01–06. <https://doi.org/10.1109/icdsns58469.2023.10245381>
- Epe, D. J. F., Cañamo, J. T., Palioto, K. M. B., & Orong, M. Y. (2024). Evaluating the Grammarly AI tool for improving academic writing: Soliciting students' perceptions of usefulness and ease of use. *International Journal of Scientific Research and Engineering Development*, 7(6), 1–10. <https://ijsred.com/volume7/issue6/IJSRED-V7I6P100.pdf>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, 47. <https://doi.org/10.1186/s41239-023-00425-2>

- European Commission (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators. *Publications Office of the European Union*. <https://doi.org/10.2766/153756>
- Gozali, I., Wijaya, A., Lie, A., Cahyono, B., & Suryati, N. (2024). Leveraging the potential of ChatGPT as an automated writing evaluation (AWE) tool: Students' feedback literacy development and AWE tools integration framework. *The JALT CALL Journal*, 20, 1–22. <https://doi.org/10.29140/jaltcall.v20n1.1200>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Haroud, S., & Saqri, N. (2025). Generative AI in Higher Education: Teachers' and Students' Perspectives on Support, Replacement, and Digital Literacy. *Education Sciences*, 15(4), 396. <https://doi.org/10.3390/educsci15040396>
- Hou, X., He, S., & Rongxiu, C. (2024). Learner use of AI-generated feedback for written corrective feedback in L2 writing: Usefulness, user proficiency and attitude. *Proceedings of the ACM on Human-Computer Interaction*, 70–76. <https://doi.org/10.1145/3678726.3678767>
- Jansen, T., Höft, L., Bahr, J. L., Fleckenstein, J., Möller, J., Köller, O., & Meyer, J. (2024). Comparing generative AI and expert feedback to students' writing: Insights from student teachers. *Psychologie in Erziehung und Unterricht*, 2024(2), 80–92. <https://doi.org/10.2378/peu2024.art08d>
- Khasawneh, M. (2023). Integrating AI-based virtual conversation partners in enhancing speaking skills in foreign languages: Insights from university students. *Journal of Southwest Jiaotong University*, 58(5). <https://doi.org/10.35741/issn.0258-2724.58.5.43>
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers & Education: Artificial Intelligence*, 3, 100074. <https://doi.org/10.1016/j.caeai.2022.100074>
- Kohnke, L., Zou, D., & Su, F. (2025). Exploring the potential of GenAI for personalised English teaching: Learners' experiences and perceptions. *Computers and Education: Artificial Intelligence*, 8, 100371. <https://doi.org/10.1016/j.caeai.2025.100371>
- Lee, S. S., & Moore, R. L. (2024). Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review. *Online Learning Journal*, 28(3), 82–104. <https://doi.org/10.24059/olj.v28i3.4593>
- Liang, J., Huang, F., & Teo, T. (2024). Understanding Chinese university EFL learners' perceptions of AI in English writing. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1), 1–16. <https://doi.org/10.4018/IJCALLT.358918>
- Lipsch-Wijnen, I., & Dirkx, K. (2022). A case study of the use of the Hattie and Timperley feedback model on written feedback in thesis examination in higher education. *Cogent Education*, 9(1), 2082089. <https://doi.org/10.1080/2331186X.2022.2082089>
- Liu, X. J., Wang, J., & Zou, B. (2025). Evaluating an AI speaking assessment tool: Score accuracy, perceived validity, and oral peer feedback as feedback enhancement. *Journal of English for Academic Purposes*, 75, 101505. <https://doi.org/10.1016/j.jeap.2025.101505>

- Mabrito, M. (2025). Collaborating with generative AI in the English classroom. *The International Journal of Technology Knowledge and Society*, 21(2), 1–23. <https://doi.org/10.18848/1832-3669/cgp/v21i02/1-23>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(1), 2. <https://doi.org/10.1186/s40561-024-00295-9>
- Marzuki, W., Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, 10(2), 2236469. <https://doi.org/10.1080/2331186x.2023.2236469>
- Nicolas, S. (2024). Artificial intelligence in writing courses: Attitudes of university instructors in Lebanon. *Issues in Educational Research*, 34(4), 1469–1487. <http://www.iier.org.au/iier34/nicolas.pdf>
- Ngo, T. N., & Hastie, D. (2025). Artificial intelligence for academic purposes (AIAP): Integrating AI literacy into an EAP module. *English for Specific Purposes*, 77, 20–38. <https://doi.org/10.1016/j.esp.2024.09.001>
- OpenAI. (2023). *GPT-4 system card*. <https://cdn.openai.com/papers/gpt-4-system-card.pdf>
- Saleh, M. (2024). Artificial intelligence as a provider of feedback on EFL student compositions. *World Journal of English Language*, 15(2), 161–171. <https://doi.org/10.5430/wjel.v15n2p161>
- Siau, K., & Wang, W. (2024). Artificial intelligence and graduate employability: What should we teach Generation AI? *Journal of Applied Learning and Teaching*, 7(2), 1–14. <https://doi.org/10.37074/jalt.2024.7.2.27>
- Sogut, S. (2024). Generative artificial intelligence in EFL writing: A pedagogical stance of pre-service teachers and teacher trainers. *Focus on ELT Journal*, 6(1), 58–73. <https://focusonelt.com/index.php/foe/article/view/134/77>
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Waschauer, M., & Olsen, C. B. (2024). Comparing the quality of human and ChatGPT feedback on students' writing. *Learning and Instruction*, 91, 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- Wang, D. (2024). Teacher- Versus AI-Generated (Poe Application) Corrective Feedback and Language Learners' Writing Anxiety, Complexity, Fluency, and Accuracy. *The International Review of Research in Open and Distributed Learning*, 25(3), 37–56. <https://doi.org/10.19173/irrodl.v25i3.7646>
- Wang, D., & Huang, X. (2025). Transforming education through artificial intelligence and immersive technologies: Enhancing learning experiences. *Interactive Learning Environments*. Advance online publication. <https://doi.org/10.1080/10494820.2025.2465451>
- Xu, W., & Tan, X. (2024). Beyond words: L2 writing teachers' visual conceptualizations of ChatGPT in teaching and learning. *Journal of Second Language Writing*, 64, 101110. <https://doi.org/10.1016/j.jslw.2024.101110>
- Yao, Y., Zhu, X., Xiao, L., & Lu, Q. (2025). Secondary school English teachers' application of artificial intelligence-guided chatbot in feedback provision: An activity theory

- perspective. *Journal of Second Language Writing*, 67, 101179. <https://doi.org/10.1016/j.jslw.2025.101179>
- Yu, E., & Canton, S. (2020). Student-inspired optimal design of online learning for generation Z. *Journal of Educators Online*, 17(1), https://www.thejeo.com/archive/archive/2020_71/yupdf
- Valdez, T., & García-Peñalvo, O. (2025). Generative AI tools in Salvadoran higher education: Balancing equity, ethics, and knowledge management in the Global South. *Education Sciences*, 15(2), 214. <https://doi.org/10.3390/educsci15020214>
- Zhang, Z., Aubrey, S., Huang, X., & Chiu, K.F. T. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: a comparative study. *Innovation in Language Learning and Teaching*, 1–19. <https://doi.org/10.1080/17501229.2025.2503890>

Appendix A Writing Coursework Task Sheet

Essay Prompt
Task: Write a discursive essay based on the essay prompt below.
Benefits and drawbacks of integrating Generative AI in Higher Education
Core Requirements
Word count: 300-500 words (excluding the title and reference list). Discuss both the benefits and drawbacks of integrating Generative AI in Higher Education. Use at least 2 sources from the Recommended Reading List. Avoid using ideas from outside source materials. You should refer to XJTLU Harvard Referencing Guide via https://libguides.lib.xjtlu.edu.cn/ld.php?content_id=5887149 and follow the citation and reference requirements.
Recommended Reading List
Source 1. Benefits, Challenges, and Methods of Artificial Intelligence (AI) Chatbots in Education: A Systematic Literature Review https://doi.org/10.46328/ijte.600 Source 2. AI Will Transform Teaching and Learning. Let's Get it Right. https://hai.stanford.edu/news/ai-will-transform-teaching-and-learning-lets-get-it-right Source 3. TED Talk How AI could save not destroy the education https://www.ted.com/talks/sal_khan_how_ai_could_save_not_destroy_education Source 4. Student perspectives on the use of generative artificial intelligence technologies in higher education https://doi.org/10.1007/s40979-024-00149-4 Source 5. The dark side of artificial intelligence in higher education https://doi.org/10.1080/02642069.2023.2258799
Learning Outcomes
Employ a range of level-appropriate reading strategies. Produce written responses with coherence, complexity, and accuracy. Demonstrate awareness and understanding of linguistic forms.
Academic Writing Suggestions
Thesis Statement Clearly state your position on the topic in the introduction. Provide a roadmap of the main points you will discuss in the essay.
Paragraph Structure

Each paragraph should focus on a single main idea.
Start with a topic sentence that introduces the main point of the paragraph.
Support your ideas with evidence and examples.
Use transitions to guide the reader through your argument.

Citations and References

Use XJTLU Harvard Referencing Guide for accurate citations.
Include in-text citations whenever you use information from your sources.
Create a reference list at the end of your essay for all cited sources.

Academic Vocabulary

Use discipline-specific terminology related to Generative AI and higher education.
Avoid informal language and colloquial expressions.

Critical Analysis

Evaluate the strengths and weaknesses of integrating Generative AI in higher education.
Provide a balanced argument supported by evidence.

Coherence and Cohesion:

Ensure logical flow between sentences and paragraphs.
Use transitional words and phrases to connect ideas.

Conclusion

Summarize your key points in the conclusion.
Restate your thesis and provide a final thought on the topic.

Formatting Requirements

File Format: MS Word DOC or DOCX
Filename: SURF_Your full name_your student ID number (it will be coded later to be anonymously)

Line Spacing: 1.5
Word Font: Tahoma
Font Size: 11 or 12 points
Include page numbers

Marking Criteria

Task (30%)

Understanding of the Prompt (10%):
Exceeds Expectations: Demonstrates a deep understanding of the complexities of integrating GenAI in Higher Education and presents nuanced arguments.
Depth of Analysis (10%):
Meets Expectations: Provides a thorough analysis of both benefits and drawbacks with clear examples and well-developed arguments.
Critical Thinking (10%):
Below Expectations: Lacks critical evaluation and fails to consider the broader implications of GenAI integration in higher education.

Organization (30%)

Structure and Coherence (10%):
Exceeds Expectations: Presents ideas in a well-structured and coherent manner with seamless transitions between paragraphs.
Introduction and Thesis Statement (10%):
Meets Expectations: Includes a clear thesis statement that previews the main points of the essay.
Body Paragraph Development (10%):

Below Expectations: Body paragraphs lack depth and fail to provide sufficient analysis and supporting evidence.

Vocabulary (20%)

Academic Vocabulary (10%):

Exceeds Expectations: Demonstrates a sophisticated use of academic vocabulary specific to GenAI and higher education.

Variety of Vocabulary (5%):

Meets Expectations: Shows a good range of vocabulary to express ideas effectively.

Word Choice (5%):

Below Expectations: Word choice is sometimes imprecise or inappropriate for the academic context.

Grammar and Mechanics (20%)

Sentence Structure (10%):

Exceeds Expectations: Constructs varied and complex sentences to convey ideas with clarity and sophistication.

Grammar Accuracy (5%):

Meets Expectations: Demonstrates consistent and accurate use of grammar throughout the essay.

Punctuation and Spelling (5%):

Below Expectations: Contains frequent errors in punctuation and spelling that hinder readability.

Appendix B Wenjuanxing Questionnaire

Demographic questions

1. Please indicate your current year of study

- Year 1
- Year 2
- Year 3
- If specify_____

2. Please indicate your gender

- Male
- Female
- I would rather not say

Essay writing process

1. From 1 (not difficult at all) to 10 (extremely difficult), how would you rate the difficulty of writing the first draft?

2. Which three factors do you consider are the most important ones when you are writing the first draft?

- a. How to write a thesis statement
- b. How to organize the essay structure
- c. How to explain and develop idea with examples
- d. How to write proper sentences
- e. How to choose accurate vocabularies to express meanings

3. What do you think are the three important components of a written feedback?
 - a. Grammar accuracy
 - b. Vocabulary choice
 - c. Essay structure
 - d. Idea development in a body paragraph
 - e. Idea progression in the whole essay
 - f. Logical thinking
 - g. Academic style
4. After finalizing your final draft, if you have a chance to have teacher written feedback, what aspects do you think you need further feedback?

Evaluating the AI tools' generated feedback

1. On a scale of 1 to 5 (1 being "strongly disagree" and 5 being "strongly agree"), how much do you agree with the following statement: "The AI's feedback significantly improved my writing skills."
2. Have you previously used Generative AI to receive feedback on your EAP essays?
 - a) Yes, frequently
 - b) Yes, a few times
 - c) No, but I'm open to trying
 - d) No, and I'm not interested
3. What Generative AI tool have you used before?
4. When comparing AI feedback with tutor feedback, which do you find more helpful?
 - a) AI feedback
 - b) tutor feedback
 - c) Both are equally helpful
 - d) Neither is particularly helpful
5. In which of the following areas do you feel AI feedback has been particularly beneficial for your writing?
 - a) Grammar and syntax
 - b) Vocabulary enhancement
 - c) Sentence structure
 - d) Coherence and organization
 - e) Introduction and conclusion
 - f) All of the above
 - g) None of the above
6. What has been your experience with AI feedback in terms of understanding and evaluating the theme and structure of your essays?
7. Are there any specific improvements or changes you would like to see in the future regarding AI feedback on your writing?
8. On a scale of 1 to 5, where 1 is 'definitely no', 3 is 'maybe', and 5 is 'definitely yes', how likely would you be to continue using AI feedback for your future writing if given the opportunity?
9. How confident are you in AI feedback as compared to feedback from human teachers?
10. Which feature of using AI feedback do you find most valuable?
 - a) Immediate feedback

- b) Objective analysis
- c) Detailed error correction
- d) Usability at any time
- e) Other (please specify)

Appendix C

Semi-structured focus group interview

1. Do you perceive any changes in grammatical accuracy between your first and final drafts? Why?
2. Do you perceive any changes in the vocabulary choice between your first and final drafts? Why?
3. Do you perceive any changes in your paragraph organization between your first and final drafts? Why?
4. Do you perceive any changes in your idea development (e.g. your logic, your choice of example) between your first and final drafts? Why?
5. What are important details that you appreciate in the feedback?
6. What are irrelevant details that you do not want to read in the feedback?
7. Overall, which AI tool's generated feedback do you think is most helpful? Why?
8. Overall, which AI tools' generated feedback do you think is least helpful? Why?
9. After receiving the AI generated feedback, do you think teacher feedback is still needed? If yes, specify which aspects you think need teacher to provide feedback with.

Ethics Approval Statement

This study was conducted in accordance with the ethical standards set forth by the Xi'an Jiaotong-Liverpool University Ethics Committee.

Declaration of Interest Statement

All authors declare no conflict of interest.