Multidisciplinary Reviews

QUESTIONNAIRES APPLICATION
Published Online: January 16, 2026
https://doi.org/10.31893/multirev.2026380

Evaluating students' perceptions, writing outcomes, and challenges in using AI-based feedback systems for english academic writing

Sri Wahyuni[a] ✉ | **Desi Surlitasari Dewi**[b] | **Rini Susanti Wulandari**[a] | **Thohiriyah**[a] | **Didit Kurniadi**[a] | **Fani Hindun Muawanah**[a] | **Nadya Zulfa Chairunnisa**[a] | **Mega Aulia Ardhi**[a]

[a]Department of English Education, Faculty of Languages and Arts, Universitas Negeri Semarang, Semarang, Indonesia.
[b]Department of English Education, Faculty of Teacher Training and Education, Universitas Riau Kepulauan, Batam, Indonesia.

Abstract The increasing prevalence of artificial intelligence (AI) in higher education has transformed student engagement with academic writing, especially via automated feedback systems. This study seeks to investigate undergraduate students' experiences with AI-driven feedback tools by analyzing three primary dimensions: perceived utility, usability, and system dependability. The study employed a mixed-methods approach and included 156 English Education undergraduates from two Indonesian institutions who often utilized multiple AI-assisted platforms, including Grammarly, ChatGPT, Quillbot, and Microsoft Copilot. Quantitative data were collected using a five-point Likert-scale questionnaire, while qualitative insights were derived from open-ended responses aimed at capturing personal perspectives on benefits and obstacles. The results indicate that students predominantly possess favorable views on AI-assisted feedback, with usefulness attaining the highest rating (M = 4.24, SD = 0.67). Students indicated enhancements in grammatical precision (M = 4.32), vocabulary sophistication (M = 4.21), and increased understanding of syntactic structure and lexical selection. The ease of use received a favorable rating (M = 4.18), mostly attributed to the accessibility and immediacy of automatic response. Nonetheless, reliability received a marginally lower grade (M = 3.91), as students observed sporadic mistakes, inconsistent recommendations, and a restricted comprehension of context or rhetorical meaning. Qualitative responses underscored difficulties including excessive dependence on AI corrections, ambiguous error explanations, and inadequate help for advanced writing elements such as idea development and coherence. The study indicates that AI-driven feedback is a beneficial adjunct to writing training, providing efficiency and linguistic assistance; yet, it cannot entirely supplant human feedback in managing content arrangement and more profound conceptual requirements. An equitable integration of AI technologies with instructor facilitation is crucial to optimize their educational influence on students' academic writing advancement.

Keywords: automated writing evaluation, student technology acceptance, digital learning tools, generative AI systems, EFL composition, revision practices

1. Introduction

The rapid advancement of artificial intelligence (AI) has ushered in a new era of innovation across multiple domains, including education. Within higher education, AI technologies are reshaping pedagogical practices by enabling automation, personalization, and continuous data-driven feedback (Haetami, 2025; Yenduri et al., 2024). This transformation extends beyond administrative convenience, reaching deep into the core of learning and assessment practices. One area that has particularly benefited from AI-driven development is second language (L2) writing instruction, where feedback plays a pivotal role in helping students refine linguistic accuracy, coherence, and academic expression. The emergence of AI-based feedback systems such as ChatGPT, Grammarly, Quillbot, WordTune, Microsoft Copilot, and Write & Improve by Cambridge has redefined how feedback is generated, delivered, and consumed. These systems offer students real-time error detection, language enhancement suggestions, and model-based revisions, transforming the writing process into an interactive and iterative learning experience (El Shazly, 2021; Kim et al., 2025).

Historically, writing instruction in higher education has relied heavily on teacher-provided feedback, a practice rooted in traditional pedagogical models emphasizing direct correction, comments, and revision cycles (Peng & Barrot, 2023; Sparks et al., 2014). While effective, manual feedback is often constrained by time, workload, and inconsistency, particularly in large classes. Furthermore, feedback quality can vary depending on instructors' expertise, availability, and subjective judgment (Yu et al., 2022). In contrast, AI-based feedback systems promise objectivity, speed, and scalability—qualities that have made them increasingly popular among both instructors and students (Hayder & Mahdi, 2025). They allow writers to access immediate, context-sensitive responses and iterative learning opportunities that were previously unattainable in traditional classroom

settings. From grammar and vocabulary to organization and argument structure, AI-based tools provide multi-dimensional input, enabling learners to detect weaknesses and revise accordingly.

Nevertheless, the integration of AI into academic writing instruction is not without controversy. While numerous studies report positive outcomes such as improved linguistic accuracy, enhanced self-efficacy, and greater learner autonomy (Dewi et al., 2025; Jeong, 2018; Su et al., 2023), others raise concerns regarding overreliance, cognitive passivity, and ethical implications (Pratama et al., 2025). Students may become dependent on automated corrections, limiting their engagement with the reasoning behind revisions and thereby reducing opportunities for critical thinking and metalinguistic awareness (Barrot, 2023). Moreover, AI tools differ significantly in their analytical depth and pedagogical alignment. While some systems (e.g., Grammarly, Write & Improve) focus primarily on surface-level accuracy, others (e.g., ChatGPT, Copilot) employ generative algorithms capable of offering higher-order content suggestions. This diversity raises questions about how students interpret and internalize feedback, and whether these systems genuinely support learning or simply facilitate text production (Haddadian, 2024).

Another dimension of the debate centers on students' perceptions and acceptance of AI-based feedback systems. Perception is a key determinant of educational technology adoption and continued use (Jeong, 2018; Lestari et al., 2024). In writing contexts, students' trust in the accuracy, fairness, and pedagogical relevance of AI-generated feedback significantly influences how they use it for self-improvement (Kacena et al., 2024; Nazari et al., 2021; Salvagno et al., 2023). Research shows that positive perceptions can enhance motivation and engagement, while skepticism may hinder uptake and learning effectiveness (Hellín et al., 2023; Putri & Dewi, 2025). Perceived ease of use and usefulness are two major constructs of the Technology Acceptance Model (TAM) that are central to understanding this relationship (Scherer et al., 2018). When students view AI feedback as accessible, understandable, and relevant to their needs, they are more likely to integrate it meaningfully into their writing processes. Conversely, when feedback appears generic, opaque, or linguistically inaccurate, it can lead to frustration and disengagement.

Beyond individual perceptions, broader pedagogical and ethical concerns also shape the discussion. The deployment of AI systems in education raises questions about authorship, academic integrity, and intellectual agency (Pratama et al., 2025; Salvagno et al., 2023). As generative AI becomes increasingly capable of producing high-quality text, the boundary between "feedback" and "co-authorship" grows blurred. Some educators worry that students may bypass the learning process by relying on AI to produce rather than refine their writing (Barrett & Pack, 2023; Grassini, 2023). Others, however, argue that these technologies can serve as scaffolding tools, empowering students to develop metacognitive awareness, explore alternative linguistic structures, and practice iterative self-correction (Hayder & Mahdi, 2025; Kundu & Bej, 2025). This pedagogical tension underscores the need for empirical evaluation of how students actually engage with, perceive, and benefit from AI-based feedback systems in real academic contexts.

Given these complexities, there is a growing consensus that research should move beyond purely technological evaluations toward a more holistic understanding of how learners experience AI-based feedback. Such research must account for students' affective responses, cognitive engagement, and perceived autonomy, alongside measurable linguistic gains. In this sense, evaluating perceptions, outcomes, and challenges provides a comprehensive framework for understanding the real educational impact of AI feedback systems.

In line with these considerations, the present study aims to evaluate students' perceptions, writing outcomes, and challenges in using AI-based feedback systems for English academic writing. Specifically, it seeks to (1) assess students' perceptions regarding the usefulness, ease of use, and reliability of AI-based feedback systems; (2) identify perceived improvements in grammar, vocabulary, and writing organization as a result of their use; and (3) explore the challenges and limitations students encounter when interacting with these systems. The study also examines whether students perceive AI-based feedback as complementing or replacing traditional human feedback.

Given the increasing adoption of AI tools in higher education, it becomes essential to ensure their pedagogical alignment and ethical implementation. Despite the proliferation of these technologies, empirical research examining their actual pedagogical impact and psychological implications remains limited. It is hypothesized that students with positive perceptions of AI-based feedback systems will demonstrate greater writing improvement and stronger motivation for self-revision.

By addressing these issues, this study contributes to ongoing discussions about the pedagogical role of AI in higher education, particularly in the domain of English language teaching. The findings are expected to inform educators and policymakers about the affordances and limitations of integrating AI-based feedback into writing instruction, fostering a balanced approach that values both technological innovation and human mentorship. Ultimately, this research seeks to advance the theoretical and practical understanding of AI-mediated feedback as a transformative yet ethically sensitive component of 21st-century language education.

2. Materials and Methods

2.1. Research method

This study employed a quantitative descriptive design with supportive open-ended responses to explore students' perceptions of the use of AI-based feedback systems in enhancing their English academic writing competence. The design was considered appropriate for systematically capturing students' levels of agreement with a range of statements related to the usefulness, ease of use, and behavioral intention to continue using AI feedback tools, while also gaining qualitative insights into their experiences, preferences, and challenges. The research framework was informed by the Technology Acceptance Model (TAM) (Davis, 1989), which posits that users' behavioral intention to adopt technology is influenced primarily by their perceived usefulness (PU) and perceived ease of use (PEOU). In the context of this study, the TAM framework guided the development of questionnaire items exploring how AI-based feedback tools (such as ChatGPT, Grammarly, Quillbot, Microsoft Copilot, and similar platforms) are perceived and utilized in supporting the writing process.

2.2. Participants and context

The participants comprised 156 undergraduate students majoring in English Education from two universities in Indonesia: Universitas Riau Kepulauan (n = 51) and Universitas Negeri Semarang (n = 105). These participants were enrolled in academic writing courses, making them an ideal population for evaluating the role of AI-based feedback systems in developing English writing proficiency. The inclusion criteria required participants to (1) be currently enrolled in an English Education program, (2) have prior experience using at least one AI-based feedback tool in their writing tasks, and (3) voluntarily consent to participate in the research. Participation was anonymous and voluntary, and no identifying information was collected to ensure confidentiality. The selection of two universities from different provinces was intended to enhance the representativeness and generalizability of the findings, as students from both institutions are exposed to different institutional environments, teaching approaches, and levels of technological integration in writing courses.

2.3. Research instruments

Data were collected using a self-developed questionnaire constructed based on the Technology Acceptance Model (TAM) framework, which emphasizes perceived usefulness, perceived ease of use, and behavioral intention. The instrument consisted of 30 closed-ended items and 18 open-ended items. The closed-ended items employed a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). These items measured students' perceptions of the effectiveness of AI-based feedback in improving grammar accuracy, lexical choice, writing efficiency, confidence, and overall satisfaction with AI-assisted writing. They also examined ease of access, clarity of feedback, and future behavioral intention to continue using AI-based tools. The open-ended items aimed to complement the quantitative data by exploring students' reflections on (a) the specific AI feedback systems they use most frequently, (b) the writing stages (drafting, revising, editing, etc.) where they use AI tools, (c) challenges encountered during use, and (d) their perceptions of whether AI feedback complements or substitutes human feedback.

The initial draft of the questionnaire was subjected to expert validation to ensure content relevance, clarity, and alignment with the research objectives. Three experts in applied linguistics, educational technology, and English writing pedagogy reviewed the instrument. Their feedback was incorporated to refine item wording, eliminate redundancy, and ensure the coverage of key TAM dimensions. To establish reliability, a pilot study was conducted with 46 students of similar characteristics who were not included in the main study. The data from the pilot were analyzed using Cronbach's Alpha, yielding an internal consistency coefficient of 0.91, which indicates excellent reliability (George & Mallery, 2019). The open-ended questions were reviewed by the same panel to verify face validity and appropriateness for eliciting meaningful qualitative responses.

2.4. Data collection procedure

Data were collected over a four-week period (March–April 2025) using Google Forms to facilitate wide accessibility and minimize logistical constraints. The link to the questionnaire was distributed through university learning management systems (LMS), student WhatsApp groups, and class representatives. Before completing the questionnaire, all participants were provided with a research information sheet and an informed consent form detailing the study's objectives, procedures, and confidentiality assurances. Participants were informed that their responses would be used solely for academic research and that they could withdraw at any stage without penalty. Participation took approximately 15–20 minutes, and the questionnaire automatically recorded responses for export into SPSS and Excel for analysis.

2.5. Data analysis

Quantitative data from the Likert-scale items were analyzed using SPSS (Version 25) to generate descriptive statistics, including means, standard deviations, frequencies, and percentages. These statistics were used to describe students' perceptions across the dimensions of perceived usefulness, perceived ease of use, and behavioral intention. In addition to descriptive analysis, correlation tests were conducted to examine potential relationships between students' perceptions of usefulness, ease of use, and their intention to continue using AI-based feedback tools. The reliability of the final dataset was

reconfirmed through Cronbach's Alpha analysis using the main study data to ensure internal consistency remained stable. Qualitative data from the open-ended questions were analyzed using thematic content analysis. Responses were read repeatedly to identify recurring themes related to perceived benefits, challenges, and pedagogical implications of AI-based feedback in writing. The themes were then synthesized and illustrated using charts and figures created in Microsoft Excel to visualize major findings and highlight patterns of use across writing stages and tool preferences.

2.6. Ethical consideration

This study strictly adhered to the ethical standards required for educational research involving human participants. Approval was obtained from the Ethics Committees of Universitas Negeri Semarang and Universitas Riau Kepulauan prior to data collection. Participants were informed about the voluntary nature of their participation and the confidentiality of their responses. No personally identifiable information was collected, and all data were stored securely with password protection. Participants provided informed consent before accessing the questionnaire.

3. Results

3.1. Perceptions of usefulness, ease of use, and reliability

Table 1 presents the descriptive statistics of students' perceptions regarding the usefulness, ease of use, and reliability of AI-based feedback systems in academic writing courses. The results indicate that overall student perceptions were highly positive, with mean scores for most items exceeding 4.00 on a five-point Likert scale.

Table 1 Descriptive statistics for items measuring usefulness, ease of use, and reliability (N = 156).

Construct	Item Statement	M	SD
Usefulness	1. AI-based feedback helps me identify my writing errors more efficiently.	4.28	0.64
	2. The system improves the overall quality of my academic writing.	4.21	0.68
	3. Feedback from AI tools helps me revise my grammar and vocabulary effectively.	4.34	0.61
	4. Using AI feedback increases my awareness of writing organization.	4.10	0.72
	5. AI feedback contributes positively to my learning outcomes in academic writing.	4.25	0.67
Ease of Use	1. The AI feedback system is easy to navigate and operate.	4.12	0.70
	2. I can use the AI tool without much technical support.	4.05	0.75
	3. The feedback is presented clearly and is easy to understand.	4.18	0.66
	4. I find it convenient to access the system anytime and anywhere.	4.31	0.63
	5. The AI system provides feedback quickly and efficiently.	4.24	0.69
Reliability	1. The AI-generated feedback is consistent across different writing tasks.	3.96	0.77
	2. I can rely on the AI feedback to be accurate and relevant.	3.89	0.80
	3. The system rarely gives incorrect or confusing suggestions.	3.75	0.83
	4. The AI tool performs reliably without frequent errors or system crashes.	4.02	0.74
	5. I trust the AI system to provide dependable feedback on my writing.	3.91	0.78

3.1.1. Usefulness of AI-based feedback

Findings from the questionnaire show that students perceived the AI-based feedback system as highly useful in supporting their writing improvement. The majority of respondents agreed that AI-generated feedback helps them revise and improve their grammar and vocabulary, with 84.6% selecting "agree" or "strongly agree" (M = 4.34, SD = 0.61). Likewise, 80.1% of students agreed that the system contributes to their overall writing quality (M = 4.21, SD = 0.68). The item "AI feedback helps me identify weaknesses in my writing organization" also received a high mean score (M = 4.10, SD = 0.72), suggesting that students recognized the system's role in improving the logical flow and coherence of their writing. Furthermore, 81.4% of students agreed that the feedback supports their learning outcomes in academic writing (M = 4.25, SD = 0.67). Only a small proportion of students (6.4%) disagreed or strongly disagreed with statements under this category. These results collectively indicate that students perceived AI feedback systems as beneficial in facilitating self-revision, identifying areas of weakness, and enhancing the quality of written output.

3.1.2. Ease of use

Students also expressed positive perceptions regarding the ease of using AI-based feedback systems. Mean scores for this construct ranged from M = 4.05 to M = 4.31. The highest-rated statement was "I can access the AI feedback system anytime and anywhere," which recorded M = 4.31, SD = 0.63, with 86.5% of respondents agreeing or strongly agreeing. This indicates that flexibility and accessibility are among the most appreciated features of AI feedback tools. The item "The AI system is easy to navigate and operate" obtained a mean score of M = 4.12, SD = 0.70, while 80.8% of students agreed that the system provides

clear and easy-to-understand feedback (M = 4.18, SD = 0.66). A similarly high mean score (M = 4.24, SD = 0.69) was reported for the item "The system responds quickly when I request feedback," reflecting satisfaction with the immediate response capability of AI platforms. The lowest mean score under this construct was recorded for "I can use the AI system without much technical support" (M = 4.05, SD = 0.75), though this value still reflects generally positive perceptions. Approximately 15% of students reported occasional difficulty when operating the tools, indicating that while most users found them accessible, a small group might benefit from additional guidance or training.

3.1.3. Reliability of AI feedback

Perceptions of reliability were moderately positive, with mean values ranging from M = 3.75 to M = 4.02. The item "The AI tool performs reliably without frequent errors or crashes" obtained the highest mean in this category (M = 4.02, SD = 0.74), with 77% of students agreeing or strongly agreeing. Similarly, 72% agreed that they could rely on AI-generated feedback to be accurate and relevant (M = 3.89, SD = 0.80). However, some students expressed caution toward the system's consistency. The statement "The system rarely gives incorrect or confusing suggestions" received the lowest mean score (M = 3.75, SD = 0.83), with 19% of students expressing neutrality and 13% disagreeing. The item "I trust the AI feedback to be dependable" recorded M = 3.91, SD = 0.78, suggesting that while trust was generally high, full confidence was not yet achieved. Overall, these results indicate that while most students regarded AI feedback systems as reliable and functionally stable, a small portion experienced uncertainty about the accuracy or contextual appropriateness of feedback, particularly in more complex or discipline-specific writing tasks.

The overall mean score across the three constructs was M = 4.11, reflecting positive perceptions toward AI-based feedback systems. Among the three dimensions, usefulness recorded the highest overall mean ($\bar{M}$ = 4.24), followed by ease of use ($\bar{M}$ = 4.18) and reliability ($\bar{M}$ = 3.91). These findings suggest that students generally appreciated the pedagogical and functional advantages of using AI tools to support their writing development, particularly in terms of usefulness and accessibility.

3.2. Perceived improvement in writing performance

Table 2 presents the descriptive statistics for students' perceived improvements in their writing performance after using AI-based feedback systems. The analysis focuses on three core areas of writing proficiency—grammar, vocabulary, and organization—each measured through a set of Likert-scale items (1 = strongly disagree to 5 = strongly agree). Overall, students reported moderately to highly positive perceptions regarding their improvement, with mean scores ranging from 3.95 to 4.32 across all items. The findings indicate that the use of AI-based feedback contributed most notably to the enhancement of grammatical accuracy, followed by vocabulary development and organizational coherence.

Table 2 Descriptive statistics for perceived improvements in writing performance (N = 156).

Skill Area	Item	M	SD	% Agree/Strongly Agree
Grammar	I make fewer grammatical errors after using AI feedback	4.32	0.59	86.0%
	I can identify grammatical mistakes more easily	4.27	0.61	83.9%
	I am more aware of sentence structure	4.24	0.66	81.4%
	I understand grammar rules more clearly	4.18	0.68	79.5%
Vocabulary	I use more appropriate vocabulary in my writing	4.21	0.63	82.1%
	I can replace repetitive words with better alternatives	4.16	0.64	80.8%
	I have learned new vocabulary from AI feedback	4.13	0.67	79.5%
Organization	My ideas are more logically arranged	4.02	0.71	77.6%
	My writing is more coherent and cohesive	3.95	0.74	75.0%
	I use transitions more effectively between paragraphs	3.98	0.73	76.3%

3.2.1. Improvements in grammar

Students demonstrated the highest perceived improvement in grammatical accuracy after using AI-based feedback tools. As shown in Table 2, the mean scores for grammar-related items ranged between 4.18 and 4.32, indicating a generally strong consensus that AI systems contributed positively to their grammatical awareness and accuracy. The item "I make fewer grammatical errors after using AI feedback" recorded the highest mean (M = 4.32, SD = 0.59), with 86% of students agreeing or strongly agreeing. This suggests that most participants felt that the AI-generated corrections and explanations directly supported their understanding of correct grammatical forms.

Similarly, 83.9% of students agreed that they could identify grammatical mistakes more easily after engaging with AI feedback (M = 4.27, SD = 0.61). The feedback provided by these systems often includes instant identification of errors in tense, subject–verb agreement, and article use, which may have facilitated immediate recognition and correction. Moreover, "I am more aware of sentence structure" (M = 4.24, SD = 0.66) indicates that beyond surface-level correction, learners also gained

deeper syntactic awareness. This awareness appears to stem from seeing alternative sentence constructions suggested by the AI, which allowed students to internalize grammatical patterns more effectively.

Meanwhile, 79.5% of students reported that they "understand grammar rules more clearly" after using AI feedback (M = 4.18, SD = 0.68). Although slightly lower than other items in this category, the result still reflects strong positive perception, suggesting that AI feedback may serve not only as corrective but also as instructional support. In summary, students perceived AI systems as reliable tools for identifying and understanding grammatical errors, leading to greater confidence and awareness in their writing.

3.2.2. Improvements in vocabulary

Vocabulary development also emerged as a strong area of improvement. The mean scores for vocabulary-related items ranged from 4.13 to 4.21, with most students acknowledging that AI-based feedback systems helped enhance their lexical variety and precision. The item "I use more appropriate vocabulary in my writing" received the highest rating in this group (M = 4.21, SD = 0.63), with 82.1% of participants agreeing or strongly agreeing. This finding indicates that AI feedback not only corrected misused words but also provided contextualized suggestions that expanded students' lexical awareness.

Additionally, "I can replace repetitive words with better alternatives" (M = 4.16, SD = 0.64) shows that learners benefited from synonym suggestions offered by AI tools, enabling them to diversify word choice and avoid redundancy. A similar trend was observed in the item "I have learned new vocabulary from AI feedback" (M = 4.13, SD = 0.67), where 79.5% of students indicated that AI systems served as an incidental source of vocabulary acquisition. These systems often recommend alternative phrases or collocations during text revision, helping students to discover and apply new lexical items.

Collectively, the data suggest that AI-based feedback supported learners' development of vocabulary accuracy and diversity. Although the gains were slightly lower than those in grammar improvement, the consistently high mean values across all vocabulary-related items indicate that students perceived tangible benefits from the system's lexical suggestions.

3.2.3. Improvements in organization

The perceived improvement in writing organization also received positive responses, though with slightly lower mean scores compared to grammar and vocabulary. As shown in Table 2, organizational items recorded means between 3.95 and 4.02. The statement "My ideas are more logically arranged" achieved the highest score in this category (M = 4.02, SD = 0.71), with 77.6% of students agreeing or strongly agreeing. This finding implies that AI-based feedback helped students structure their writing more clearly and logically. The item "My writing is more coherent and cohesive" (M = 3.95, SD = 0.74) garnered agreement from 75% of respondents, suggesting that while AI tools provided useful insights into coherence, their feedback on paragraph transitions and overall flow may have been less explicit than on grammar or vocabulary. Similarly, "I use transitions more effectively between paragraphs" (M = 3.98, SD = 0.73) shows moderate improvement, indicating that students may have applied feedback related to linking words and sentence connectors more selectively.

These results reflect that while AI-based feedback systems contribute to enhancing organizational quality, their automated nature might limit nuanced feedback on discourse-level coherence. Nonetheless, the majority of students still reported noticeable improvements in structuring ideas and ensuring smoother transitions, indicating that AI systems play a supportive, though partial, role in developing organizational skills. In summary, students perceived AI-based feedback systems as highly beneficial for improving their overall writing performance, particularly in grammar and vocabulary. The highest levels of agreement were observed in grammatical improvement items (M = 4.32, SD = 0.59), followed closely by vocabulary enhancement (M = 4.21, SD = 0.63), and slightly lower for organizational development (M = 4.02, SD = 0.71). These findings collectively demonstrate that while AI feedback tools effectively support local-level revisions (grammar and vocabulary), their role in facilitating global-level revisions (organization and coherence) appears more limited

3.3. Challenges and limitations in using AI-based feedback systems

To address the third research question—exploring the challenges and limitations students encounter when interacting with AI-based feedback systems—qualitative data were collected through open-ended responses in the questionnaire. A total of 156 students provided written reflections on their experiences using various AI-assisted feedback platforms, such as ChatGPT, Grammarly, Quillbot, and Write & Improve. The responses were analyzed using thematic analysis following Braun and Clarke's (2006) six-step framework. This method involved iterative reading, coding, and theme refinement to identify patterns of meaning across participants' accounts. The analysis yielded three major themes: (1) Technical and Functional Constraints, (2) Cognitive and Emotional Challenges, and (3) Pedagogical Limitations of AI Feedback. These themes, together with their subthemes and relative frequencies, are presented in Table 3.

Table 3 Summary of themes on challenges and limitations in using AI-based feedback systems.

Theme	Subtheme	% of Students Mentioning	Example Excerpt
Technical and Functional Constraints	Inaccurate or inconsistent feedback	42%	"Sometimes Grammarly highlights correct sentences as errors, which makes me unsure which one is right."
	Connectivity and access issues	31%	"It takes too long to load or sometimes crashes when I upload long essays."
Cognitive and Emotional Challenges	Over-reliance and reduced self-editing	35%	"I tend to just accept the AI's suggestion without really thinking if it's correct."
	Frustration or confusion with unclear feedback	29%	"The feedback doesn't explain why the change is needed, so I can't learn from it."
Pedagogical Limitations of AI Feedback	Lack of contextual understanding	38%	"The AI corrects grammar but doesn't understand what I really mean."
	Limited guidance on content and organization	27%	"It doesn't give suggestions for how to improve the flow or argument of my essay."

3.3.1. Technical and functional constraints

The first theme concerns the technical and functional issues that hinder students' ability to effectively utilize AI-based feedback systems. Approximately 42% of respondents mentioned that AI feedback was occasionally inaccurate, inconsistent, or overly rigid, leading to confusion rather than improvement. Several students noted that automated corrections were sometimes contextually inappropriate, particularly when the software failed to recognize idiomatic expressions or academic phrasing. One participant commented, "Sometimes the AI marks correct grammar as wrong, and it makes me confused whether to trust it or not." In addition, 31% of students cited connectivity and access problems as obstacles to efficient use. These challenges included system lags, limited access to premium features, or unstable internet connections that disrupted the feedback process. A student from Universitas Negeri Semarang remarked, "When the internet is unstable, the feedback takes forever to appear, and sometimes I lose my work."

While most participants recognized the overall convenience of AI tools, these technical limitations appeared to reduce students' confidence in relying on them consistently. Moreover, students from Universitas Riau Kepulauan—where digital infrastructure is comparatively less stable—reported a slightly higher frequency of connectivity-related difficulties. This suggests that technological accessibility remains a contextual factor influencing students' engagement with AI-based feedback platforms.

3.3.2. Cognitive and emotional challenges

The second theme highlights students' psychological, cognitive, and affective responses to the use of AI feedback tools. About 35% of participants admitted to over-relying on AI-generated corrections, which consequently reduced their active engagement in the revision process. One student expressed, "I rely on AI too much. I just accept its corrections without thinking, so I don't really learn." This reflects a recurring concern that while AI can enhance efficiency, it may inadvertently discourage critical reflection and independent editing—key elements of writing skill development.

A related subtheme pertains to emotional frustration and confusion, reported by 29% of students. Participants described feelings of irritation when the feedback was vague, overly technical, or inconsistent across platforms. A common complaint was the lack of explanatory depth in the feedback provided. For example, one respondent noted, "It tells me to change the sentence but doesn't explain why, so I don't understand what's wrong." Others described feeling overwhelmed by the volume of suggested revisions, particularly in early drafts, which created a sense of cognitive overload rather than clarity.

Moreover, some students perceived that the impersonal nature of AI feedback limited its motivational value. Unlike teacher comments, which often include praise or encouragement, AI-generated responses were viewed as mechanical and emotionally detached. As one participant wrote, "AI just corrects mistakes. It never says anything positive, so sometimes I feel discouraged." These findings suggest that while AI-based systems can offer immediate and detailed feedback, they may lack the human interaction and empathy that sustain learners' confidence and persistence in writing.

3.3.3. Pedagogical limitations of AI feedback

The third theme relates to the pedagogical shortcomings of AI-based feedback systems, which limit their ability to support complex aspects of writing development. Approximately 38% of students observed that AI tools lacked contextual understanding, resulting in feedback that failed to align with their intended meaning. For instance, one student reported, "AI doesn't understand what I want to say. It changes my sentence so it sounds correct grammatically but loses the meaning." This

issue was especially prominent in more advanced writing tasks, such as argumentative or analytical essays, where nuance, tone, and cohesion are essential.

In addition, 27% of respondents highlighted the limited guidance AI provides on higher-level writing dimensions, such as content development, paragraph organization, and logical flow. Students emphasized that while AI systems excel at surface-level corrections—grammar, punctuation, and word choice—they are less effective in helping improve idea organization or argument coherence. As one participant noted, "It only focuses on grammar, not on how to make my writing more logical or convincing."

Several students also remarked that AI systems lack adaptability to discipline-specific writing conventions. For example, participants from English education programs found that AI-generated feedback occasionally contradicted the academic norms they had been taught in class. This suggests a mismatch between automated feedback algorithms and pedagogical goals, particularly in contexts emphasizing genre awareness and critical writing.

Overall, the findings reveal a nuanced picture of students' experiences with AI-based feedback systems. While most participants acknowledged the utility of these tools for identifying grammatical errors and improving sentence clarity, they also reported substantial technical, cognitive, and pedagogical challenges. Technical constraints reduced trust and usability; cognitive and emotional factors influenced motivation and engagement; and pedagogical limitations restricted the systems' ability to foster deep, reflective learning.

These results indicate that AI-based feedback systems, although effective in enhancing linguistic accuracy, remain inadequate as standalone pedagogical instruments. The absence of contextual interpretation and the lack of personalized guidance underscore the continuing importance of human-mediated feedback in writing instruction. Consequently, integrating AI tools as complementary aids rather than replacements for teacher feedback appears to be a more sustainable approach for developing students' writing competence.

4. Discussion

The findings of this study provide important insights into students' perceptions and experiences with AI-based feedback systems in academic writing contexts. Overall, students demonstrated high acceptance and positive attitudes toward the integration of AI tools such as Grammarly, Quillbot, and ChatGPT in their writing processes. These perceptions were largely shaped by the perceived usefulness, ease of use, and reliability of the tools—dimensions central to the Technology Acceptance Model (TAM) (Scherer et al., 2018; Jeong, 2018). The results also highlight that students felt notable improvements in their grammatical accuracy, vocabulary range, and organizational skills, underscoring AI's pedagogical potential as an assistive feedback mechanism. Nevertheless, qualitative responses revealed concerns regarding overreliance, contextual misunderstanding, and limited guidance for higher-order writing features. These findings collectively suggest that while AI-based feedback systems effectively enhance surface-level writing performance and learner autonomy, they must be pedagogically scaffolded to ensure critical engagement and sustained learning.

In addressing the first research question, students reported high mean scores across the three constructs—usefulness (M = 4.24), ease of use (M = 4.18), and reliability (M = 3.91). These results indicate that most participants perceived AI tools as beneficial, accessible, and relatively dependable in improving their writing outcomes. Such findings align closely with the core assumptions of the TAM, which posits that individuals are more likely to adopt technology when they find it both easy to use and useful for achieving desired outcomes (Scherer et al., 2018). Consistent with Yenduri et al. (2024) and Haetami (2025), the data underscore that perceived usefulness plays a central role in students' sustained engagement with AI technologies. The high agreement levels (over 80%) for items such as "AI feedback helps me identify my writing errors more efficiently" and "AI feedback contributes positively to my learning outcomes" reflect students' recognition of AI as a time-saving and reliable support system for self-editing and revision. This perception resonates with Dewi et al. (2025), who observed that students tend to associate AI feedback with enhanced productivity and academic confidence.

Ease of use was another strong determinant of positive perceptions. Participants expressed that AI tools were simple to navigate and provided feedback in a clear, user-friendly format. Such accessibility lowers the cognitive load associated with understanding corrective feedback—a factor crucial for second language (L2) writers (Nazari et al., 2021; Kim et al., 2025). In line with findings from Hayder and Mahdi (2025), students valued the immediacy and flexibility of AI systems, which allowed them to engage in multiple revision cycles independently. The convenience of 24-hour access and rapid response time also mirrors the growing preference for digital learning tools that support autonomous and self-paced learning. However, reliability emerged as the least strongly rated dimension, with several students noting occasional inaccuracies or inconsistencies in feedback. This finding echoes Salvagno et al. (2023), who warned that while AI can provide extensive linguistic feedback, its interpretive reliability still depends on algorithmic precision and contextual sensitivity. Hence, while students perceive AI feedback as largely dependable, they remain cautious about its accuracy—an attitude that reflects emerging digital literacy among L2 learners.

The second research question explored students' perceived improvements in specific aspects of writing performance—grammar, vocabulary, and organization—after using AI-based feedback systems. Quantitative data revealed strong perceived gains in grammatical accuracy (M = 4.32, SD = 0.59) and vocabulary enhancement (M = 4.21, SD = 0.63), with slightly lower

improvements in writing organization (M = 4.02, SD = 0.71). The majority of participants (above 80%) agreed that AI feedback helped them identify grammatical errors and use more appropriate vocabulary, confirming earlier claims that automated feedback enhances form-focused learning (Jeong, 2018; Su et al., 2023). This supports Kim et al. (2025), who found that real-time corrective feedback from AI systems promotes immediate self-correction and deeper language awareness. The findings also align with the cognitive-interactionist perspective, which emphasizes that noticing errors through feedback facilitates linguistic development and internalization (Peng & Barrot, 2023). Students' recognition of grammatical patterns and improved word choice further indicates that AI feedback can function as a formative assessment tool—offering individualized, iterative input that complements instructor feedback.

However, perceived gains in organization were relatively modest, suggesting that AI-based systems are less effective in guiding higher-order writing concerns such as coherence, argument development, and rhetorical structure. This limitation reflects a key pedagogical issue identified by Barrot (2023) and Haddadian (2024): while AI excels in surface-level correction, it lacks deep contextual and discourse understanding. Students' open-ended responses further illustrate this gap, noting that the systems "correct grammar but don't understand what I really mean" or fail to "give suggestions for improving argument flow." These insights indicate that while AI-based feedback can significantly assist mechanical accuracy, it remains insufficient as a substitute for human feedback that addresses content quality and rhetorical intent. Therefore, instructors should integrate AI feedback strategically—encouraging students to use it for linguistic refinement while relying on teacher feedback for conceptual and organizational development.

The qualitative data addressing the third research question revealed three major themes: technical and functional constraints, cognitive and emotional challenges, and pedagogical limitations. The most frequently mentioned issue was inconsistent or inaccurate feedback (42%), followed by connectivity problems (31%). Such technical issues have been widely reported in AI-assisted learning environments, where algorithmic bias and limited training data can lead to misleading corrections (Grassini, 2023; Haetami, 2025). From a user-experience perspective, technical disruptions can lower engagement and trust in digital systems (Putri & Dewi, 2025). Cognitive and emotional challenges also emerged as significant barriers. About 35% of students admitted overreliance on AI tools, often accepting suggestions uncritically. This finding supports Pratama et al. (2025) and Barrot (2023), who caution that excessive dependence on automated feedback may reduce learners' self-editing capacity and metacognitive engagement. Some participants expressed frustration over ambiguous feedback, which failed to explain underlying rules—a problem that limits feedback literacy and learning transfer (Yu et al., 2022). These patterns suggest a need for instructional scaffolding that helps students interpret and evaluate AI feedback rather than merely apply it.

Pedagogical limitations of AI feedback formed the third thematic cluster. Around one-third of respondents highlighted the tools' inability to provide contextualized or content-level suggestions. This aligns with previous research indicating that AI-generated feedback is often limited to sentence-level corrections and lacks awareness of argument logic or academic conventions (Peng & Barrot, 2023; Kacena et al., 2024). In this regard, AI functions best as a supplementary feedback source rather than a comprehensive writing tutor. Yet, the study's findings also suggest a pedagogical opportunity: students' awareness of AI's limitations may foster critical thinking and self-regulation if instructors explicitly guide them in reflecting on the appropriateness of automated corrections. As Hayder and Mahdi (2025) argue, AI-based feedback should be embedded within a reflective learning framework, where students are encouraged to question, verify, and justify their revisions.

Taken together, these findings contribute to ongoing discussions about the role of AI in higher education writing pedagogy. Consistent with Yenduri et al. (2024) and Kim et al. (2025), the present study confirms that AI-based feedback can enhance linguistic accuracy, learner autonomy, and revision awareness when used appropriately. However, without sufficient critical mediation, such tools risk promoting superficial learning and dependency. The Technology Acceptance Model provides a useful interpretive lens, illustrating that perceived usefulness and ease of use strongly influence students' willingness to engage with AI systems. Yet, perceived reliability—especially regarding contextual accuracy—remains an area requiring improvement. Pedagogically, integrating AI feedback into writing instruction should not replace human feedback but rather complement it through a blended feedback model that combines the efficiency of automation with the nuance of human interpretation (Dewi et al., 2025; Barrett & Pack, 2023).

Taken together, the findings suggest that AI-based feedback holds substantial promise for improving EFL writing instruction by enhancing learners' accuracy, confidence, and engagement. Nonetheless, its effectiveness depends heavily on how students perceive, interpret, and apply the feedback they receive. Instructors and curriculum designers should therefore aim to integrate AI feedback within a human–AI partnership model, where automated feedback provides formative support, and teacher mediation ensures depth, context, and ethical awareness. Future research could further examine how long-term use of AI feedback affects learners' critical thinking, writing self-efficacy, and sustained writing improvement.

5. Final Considerations

This study set out to examine students' perceptions and experiences with AI-based feedback systems in EFL writing contexts, focusing on perceived usefulness, ease of use, reliability, perceived improvements in writing performance, and the challenges encountered during their use. The findings provide strong evidence that AI-based feedback tools are not only positively perceived by students but also play a meaningful role in enhancing writing accuracy, vocabulary development, and

organizational coherence. In line with the Technology Acceptance Model (TAM) (Scherer et al., 2018), students' high ratings of usefulness and ease of use significantly contributed to their acceptance of and satisfaction with these tools, indicating that intuitive design and perceived pedagogical relevance are central to effective technology integration in writing instruction. Furthermore, the results confirmed that AI feedback systems have a measurable impact on learners' linguistic competence. Participants reported that automated feedback facilitated faster recognition and correction of errors, enabling iterative writing and revision practices that strengthened their grammatical accuracy and lexical variety. Importantly, students valued the immediacy and accessibility of AI-generated feedback, which allowed them to engage with writing beyond classroom boundaries and at their own pace—fostering self-regulated learning. Nevertheless, this study also illuminated several pedagogical and ethical challenges. A subset of participants expressed concerns about the contextual accuracy of AI feedback, its tendency toward surface-level corrections, and the risk of overdependence on automated suggestions. Ethical questions concerning originality, authorship, and academic integrity also surfaced, emphasizing the need for explicit digital literacy and responsible AI use in academic settings. Hence, while AI-based feedback can complement teacher feedback, it should not replace human judgment, critical interpretation, or personalized pedagogical support.

The significance of this study lies in its comprehensive exploration of the intersection between technology acceptance, writing pedagogy, and learner autonomy in the context of AI integration. The findings contribute to the growing discourse on how intelligent feedback systems can transform traditional writing instruction by offering scalable, data-driven, and individualized learning experiences. More importantly, this research highlights the need for a human–AI partnership model, in which teachers act as mediators who guide students to use AI feedback critically and reflectively. Such an approach can balance efficiency with depth, ensuring that technology enhances—rather than diminishes—cognitive engagement and learning ownership. However, this study is not without limitations. The sample was confined a small population of EFL tertiary students, which may restrict the generalizability of the findings. The study also relied on self-reported perceptions and thematic reflections, which may be influenced by individual biases or limited familiarity with AI tools. Future studies should therefore consider longitudinal or experimental designs that trace actual performance improvement over time, incorporate diverse student populations, and examine the long-term effects of AI feedback on writing quality, metacognitive awareness, and learner agency. Additionally, comparative studies exploring different AI systems or hybrid feedback models could offer valuable insights into which technological configurations most effectively promote learning outcomes.

In conclusion, AI-based feedback systems represent a transformative innovation in L2 writing pedagogy, offering efficiency, immediacy, and personalization that traditional feedback mechanisms often lack. When integrated thoughtfully and ethically, these tools have the potential to empower learners, foster autonomy, and elevate the overall quality of writing instruction. Yet, their effectiveness ultimately depends on informed pedagogical use, continuous critical engagement, and the preservation of the human element in teaching and learning. This study thus underscores the importance of designing AI-enhanced writing environments that combine technological precision with human insight—an equilibrium essential for the sustainable advancement of digital learning in the age of artificial intelligence.

Acknowledgment

The authors would like to express their sincere gratitude to the anonymous student participants from Universitas Negeri Semarang and Universitas Riau Kepulauan for their valuable contributions and participation in this study.

Ethical Considerations

This study was conducted in accordance with the ethical standards of research involving human participants. Ethical approval was obtained from the Research Ethics Committee of LPPM Universitas Negeri Semarang, under approval number B/12161/UN37/1/2/HK.07.10/2025. Participation in this study was entirely voluntary, and informed consent was obtained from all respondents prior to data collection. The participants' anonymity and confidentiality were strictly maintained throughout the research process.

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This research was supported by DPA LPPM Universitas Negeri Semarang under the 2025 Research Grant, contract number 14.14.3/UN37/PPK.11/2025.

References

Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education, 20*(1). https://doi.org/10.1186/s41239-023-00427-0

Barrot, J. S. (2023). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning, 36*(4), 584–607. https://doi.org/10.1080/09588221.2021.1936071

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Dewi, D. S., Hartono, R., Wahyuni, S., & Lestari, S. (2025). Fostering learner autonomy in EFL settings through multiliteracy pedagogy: A case in tertiary education. *Australian Journal of Teacher Education, 50*(2), 1–19. https://doi.org/10.14221/1835-517X.6369

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems, 38*(3), e12667. https://doi.org/10.1111/exsy.12667

George, D., & Mallery, M. (2019). *SPSS for Windows step by step: A simple guide and reference* (15th ed.). Routledge.

Grassini, S. (2023). Shaping the future of education with artificial intelligence: Opportunities, challenges, and implications. *Education Sciences, 13*(7), 692. https://doi.org/10.3390/educsci13070692

Haddadian, G. (2024). Comparing the effects of teacher feedback, automated feedback, and integrative feedback on EFL learners' writing accuracy and writing apprehension. *Computer Assisted Language Learning Electronic Journal (CALL-EJ), 25*(3), 124–147. https://callej.org/index.php/journal/article/view/436

Haetami. (2025). AI-driven educational transformation in Indonesia: From learning personalization to institutional management. *Al-Ishlah: Jurnal Pendidikan, 17*(2), 1819–1832. https://doi.org/10.35445/alishlah.v17i2.7448

Hayder, M., & Mahdi, M. A. (2025). The effect of artificial intelligence in the educational process of the English language. *American Journal of Research in Humanities and Social Sciences, 33*, 21–33. https://americanjournal.org/index.php/ajrhss/article/view/2728

Hellín, C. J., Calles-Esteban, F., Valledor, A., Gómez, J., Otón-Tortosa, S., & Tayebi, A. (2023). Enhancing student motivation and engagement through a gamified learning environment. *Sustainability, 15*(19), 14119. https://doi.org/10.3390/su151914119

Jeong, K. O. (2018). Developing EFL learners' communicative competence through multimedia-assisted language learning. *Journal of Theoretical and Applied Information Technology, 96*(5), 1367–1376.

Kacena, M. A., Plotkin, L. I., & Fehrenbacher, J. C. (2024). The use of artificial intelligence in writing scientific review articles. *Current Osteoporosis Reports, 22*(1), 115–121. https://doi.org/10.1007/s11914-023-00852-0

Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies, 30*(1), 1265–1300. https://doi.org/10.1007/s10639-024-12878-7

Kundu, A., & Bej, T. (2025). Transforming EFL teaching with AI: A systematic review of empirical studies. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-025-00470-0

Lestari, I., Dewi, D. S., & Shalehoddin, S. (2024). Gamifying vocabulary learning: The effects on students' acquisition. *ELT Echo, 9*(2), 130–142. https://doi.org/10.24235/eltecho.v9i2.19077

Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence-powered digital writing assistant in higher education: A randomized controlled trial. *Heliyon, 7*(5), e07014. https://doi.org/10.1016/j.heliyon.2021.e07014

Peng, Y., & Barrot, J. S. (2023). Post-writing form-focused instruction in process-genre-oriented writing classrooms: Effects on second language learners' writing accuracy. *Porta Linguarum, 39*, 249–264. https://doi.org/10.30827/portalin.vi39.24276

Pratama, H., Putri, S. P., & Bahri, S. (2025). Potential risks of ChatGPT-assisted essay writing on knowledge retention among EFL learners in Indonesia. *Language and Education Journal, 10*(1), 148–168. https://doi.org/10.52237/1j0f2165

Putri, A., & Dewi, D. S. (2025). Gamifying vocabulary learning: Students' perceptions of Kahoot! in a junior high school. *ELT Echo, 9*(1), 71–86. https://doi.org/10.36597/jellt.v9i1.19327

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing? *Critical Care, 27*(1), 1–5. https://doi.org/10.1186/s13054-023-04380-2

Scherer, R., Tondeur, J., Siddiq, F., & Baran, E. (2018). The importance of attitudes toward technology for pre-service teachers' technological, pedagogical, and content knowledge: Comparing structural equation modeling approaches. *Computers in Human Behavior, 80*, 67–80. https://doi.org/10.1016/j.chb.2017.11.003

Sparks, J. R., Song, Y., Brantley, W., & Liu, O. L. (2014). Assessing written communication in higher education: Review and recommendations for next-generation assessment. *ETS Research Report Series, 2014*(2), 1–52. https://doi.org/10.1002/ets2.12035

Su, L., Noordin, N., & Jeyaraj, J. J. (2023). Implementation of strategy instruction in teaching English as a foreign language: A systematic review. *International Journal of Learning, Teaching and Educational Research, 22*(7), 156–172. https://doi.org/10.26803/ijlter.22.7.9

Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., Raj, G. D., Jhaveri, R. H., Prabadevi, B., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2024). GPT (generative pre-trained transformer): A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access, 12*, 54608–54649. https://doi.org/10.1109/ACCESS.2024.3389497

Yu, S., Di Zhang, E., & Liu, C. (2022). Assessing L2 student writing feedback literacy: A scale development and validation study. *Assessing Writing, 53*, 100643. https://doi.org/10.1016/j.asw.2022.100643