ASIAN JOURNAL OF EDUCATIONAL SCIENCES
ISSN: 0866-7675
eISSN: 2815-567X
DONG THAP UNIVERSITY JOURNAL OF SCIENCE - DONG THAP UNIVERSITY
https://ajesdthu.vn

DOI: ***https://doi.org/10.52714/dthu.ajes.27.2022***

Comparing Learners' Perceptions of AI-Enabled Writing Feedback and Human Teacher Feedback in IELTS Writing Preparation

Quoc Khanh Nguyen

The University of Da Nang, Vietnam

Email: av32.1b2cva1d047@oncce.udn.vn

Article Info.

Received: June 27, 2026
Revised: August 17, 2026
Accepted: August 28, 2026

Keywords

AI feedback, *automated writing evaluation*, *human teacher feedback*, *IELTS Writing*, *L2 writing feedback*, *trust in AI*

Cite

Nguyen, Q. K. (2026). Comparing Learners' Perceptions of AI-Enabled Writing Feedback and Human Teacher Feedback in IELTS Writing Preparation. *Asian Journal of Educational Sciences*, *1*(2), 62-83. https://doi.org/10.52714/dthu.ajes.27.2022

Abstract

This study investigated International English Language Testing System (IELTS) Writing learners' perceptions of AI-enabled writing feedback and human teacher feedback in relation to perceived writing improvement, IELTS Writing areas, and trust. Using an explanatory sequential mixed-methods design, questionnaire data were collected from 130 learners at a private IELTS teaching center in Ho Chi Minh City, Vietnam, and follow-up interviews were conducted with eight purposively selected participants. The questionnaire used a paired-source structure, enabling learners to evaluate AI feedback and human teacher feedback across parallel IELTS Writing dimensions. Quantitative data were analyzed in RStudio through reliability analysis, descriptive statistics, paired-samples t-tests, and Cohen's dz, while interview data were analyzed in QualCoder using hybrid deductive–inductive thematic analysis. The findings showed no significant difference between AI feedback and human teacher feedback in overall perceived writing improvement. However, AI feedback was rated higher for grammar correction, vocabulary improvement, and revision speed, whereas human teacher feedback was rated higher for coherence and cohesion, task response, idea development, and trust. The integrated findings indicate differentiated, task-sensitive perceptions of the two feedback sources and perceived complementarity in how some learners combined them, although the study did not evaluate objective writing outcomes.

1. INTRODUCTION

Feedback is central to second language (L2) writing development because it helps learners identify the gap between their current performance and expected writing goals. In IELTS Writing preparation, feedback is especially important because learners are expected not only to produce accurate language, but also to respond to the task, organize ideas coherently, use appropriate vocabulary, and demonstrate grammatical control. The official IELTS Writing criteria include task response or task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy (IELTS, 2023). Therefore, effective IELTS Writing feedback needs to address both lower-level language accuracy and higher-order writing quality.

In recent years, AI-based feedback tools such as ChatGPT, Grammarly, QuillBot, and IELTS writing applications have become increasingly available to language learners. These tools can provide immediate correction, vocabulary suggestions, rewriting support, and sometimes band-score estimation. Their speed and accessibility make them attractive to IELTS learners who need frequent writing practice outside the classroom. At the same time, the rapid growth of generative AI in education has created a need to understand how learners use, evaluate, and trust AI-generated feedback. UNESCO's guidance emphasizes human-centered and pedagogically responsible uses of generative AI rather than uncritical adoption (Miao & Holmes, 2023).

In this study, AI-enabled writing feedback (WF) is used as an operational umbrella term for automated correction, evaluation, suggestion, or revision support that participants obtained from AI-enabled tools during IELTS Writing preparation. The category includes heterogeneous tools such as ChatGPT, Grammarly, QuillBot, IELTS writing applications, and other reported tools and is not treated as a homogeneous technological intervention. Human teacher feedback refers to IELTS Writing feedback provided by a human teacher, tutor, or IELTS instructor. This category is likewise source-based and was not standardized by provider, delivery mode, or feedback protocol.

Existing research has shown that automated writing evaluation (AWE) and AI-generated feedback can support L2 writing practice, revision, and learner engagement (Karatay & Karatay, 2024; Wei et al., 2023). However, concerns remain about feedback accuracy, relevance, over-reliance, and the balance between AI and human input (Crosthwaite & Sun, 2026; Yoon et al., 2023). These concerns are particularly relevant to IELTS Writing, where feedback quality depends not only on correcting grammar and vocabulary, but also on interpreting task requirements, evaluating coherence, and judging argument development.

Human teacher feedback remains important because it can be adapted to learners' proficiency, writing history, classroom goals, and test-preparation needs. L2 WF is not merely correction; it is also pedagogical, dialogic, and contextual (Hyland & Hyland, 2006). For IELTS learners, teachers may provide guidance on task response, idea development, paragraph logic, and exam-specific expectations that AI feedback may not consistently address.

Although recent studies have examined automated feedback, teacher feedback, and generative AI in L2 writing, there remains a need for more IELTS-specific research comparing AI-enabled WF and human teacher feedback within the same learner group. Much existing research focuses on general L2 writing, academic writing, or the effectiveness of AI feedback alone. Less attention has been given to how IELTS learners compare AI and teacher feedback across the same writing dimensions and explain their trust in each source.

The present study examines the two feedback-source categories within the same IELTS learner sample and across parallel evaluative dimensions. It compares perceived writing

improvement and specific writing areas and examines trust through quantitative comparison and qualitative explanation. The study therefore focuses on learners' comparative perceptions and reported feedback-use practices rather than on the objective effectiveness of a particular AI system or a standardized teacher-feedback intervention.

The study is guided by the following research questions:

RQ1. How do IELTS learners compare AI-enabled WF and human teacher feedback in terms of perceived writing improvement?

RQ2. In which IELTS Writing areas do learners perceive AI-enabled writing feedback and human teacher feedback as more helpful?

RQ3. To what extent do learners' trust ratings differ between AI-enabled WF and human teacher feedback, and how do learners explain any observed difference?

2. LITERATURE REVIEW

2.1. Feedback in L2 Writing

Feedback plays an important role in writing development because it helps learners understand where they are, where they need to go, and how they can improve. Hattie and Timperley (2007) argue that feedback can strongly influence learning, although its effectiveness depends on its quality and focus. In L2 writing, feedback is especially complex because learners need support with linguistic accuracy, textual organization, meaning development, and reader expectations.

Hyland and Hyland (2006) explain that feedback on L2 writing is not only a matter of error correction. It also involves interaction, motivation, learner identity, and instructional context. This is important for IELTS Writing because grammar correction may help learners reduce sentence-level errors, but it may not help them understand whether their essay fully answers the task or develops a clear argument.

The official IELTS criteria include task response or task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy (IELTS, 2023). A feedback source that is strong in grammar correction may not necessarily be equally strong in task response, coherence, or idea development. IELTS Writing feedback should therefore be examined by dimension rather than only at the level of overall usefulness.

2.2. Automated and AI-Generated Feedback in L2 Writing

Automated writing evaluation has become increasingly common in L2 writing classrooms. Earlier AWE tools were often designed to provide scores, error correction, and automated comments. Karatay and Karatay's (2024) research synthesis shows that AWE can affect students' writing practices, teacher feedback, and learner engagement, but also highlights limitations in how students engage with automated systems. This suggests that automated feedback should be evaluated not only by whether it can produce comments, but also by how learners understand, trust, and apply those comments.

Recent research has shifted from traditional AWE toward generative AI feedback. Generative AI tools can rewrite sentences, explain errors, suggest vocabulary, and respond to follow-up prompts. Yan (2023) found that ChatGPT showed potential in an L2 writing practicum while raising concerns about academic integrity and learner dependence. Wei et al. (2023) also found that AWE training could improve L2 writing performance under structured learning conditions.

However, AI-generated feedback is not always reliable. Yoon et al. (2023) found that ChatGPT feedback on coherence and cohesion could be abstract, generic, and sometimes inaccurate. This is important for IELTS Writing because coherence and cohesion involve paragraph logic, idea progression, and the relationship between claims and support. Crosthwaite and Sun's (2026) review also identifies unresolved issues involving feedback relevance, learner uptake, prompt transparency, and the balance between AI and human input. These limitations support comparing what AI and human feedback are perceived to do well in specific learning contexts.

Recent L2 writing research further suggests that the value of feedback depends on how learners interpret, evaluate, and act on it, not simply on whether feedback is available. Feedback literacy and engagement research has therefore emphasized learners' judgment of feedback quality, selective uptake, and revision decisions (Zhang & Mao, 2023; Zhang & Hyland, 2025; Lu et al., 2024). These processes are particularly relevant to AI-enabled feedback because learners may need to verify, reject, modify, or seek clarification of automated suggestions before using them.

2.3. Human Teacher Feedback in L2 Writing

Human teacher feedback has a different pedagogical function from automated feedback. Teachers can interpret learners' writing in relation to classroom instruction, learner history, proficiency level, and assessment goals. In L2 writing, teacher feedback can support planning, organization, argumentation, and revision strategy in addition to language correction (Hyland & Hyland, 2006).

Teacher feedback is particularly important in IELTS Writing because teachers can explain why an essay does not fully answer the question, why an example is underdeveloped, or why a paragraph lacks logical progression. These forms of feedback require evaluative judgment and contextual understanding that AI tools may not provide consistently.

Studies also suggest that the two sources may support different aspects of learning. Zhang and Hyland (2018) found that student engagement with teacher and automated feedback can differ, while Link et al. (2022) showed that AWE can interact with teacher feedback in writing instruction. Automated feedback is therefore better understood as part of a feedback ecology rather than as a replacement for teachers.

Human teacher feedback may also be difficult to process. Learners can experience criticism as discouraging, unclear, or difficult to enact, particularly when comments identify higher-order problems without providing immediately actionable solutions. Research on teacher written feedback similarly shows that learner engagement can include both perceived value and affective or interpretive difficulty (Mao et al., 2024; Pearson, 2022).

2.4. Trust in Automated and Human Feedback

Trust is a key issue in AI-assisted writing because learners must decide whether to accept, reject, or modify feedback suggestions. Trust in automation depends on whether users perceive an automated system as reliable and effective for a given task (Jian et al., 2000). In WF, learners may therefore trust AI for some tasks but not for others.

Ranalli (2021) argues that trust can influence how L2 learners engage with automated feedback. If learners over-trust AI, they may accept inaccurate or unsuitable suggestions; if they under-trust it, they may ignore useful feedback. Appropriate reliance requires learners to judge when automated feedback is sufficiently dependable for a particular task and when further verification or human judgment is needed (Lee & See, 2004).

This issue is highly relevant to IELTS Writing. Learners may trust AI when it identifies grammar errors or suggests vocabulary alternatives, but they may trust teachers more when feedback involves band-score judgment, task response, coherence, or idea development. Trust should therefore be examined as a central construct rather than a secondary issue.

2.5. Research Gap and Study Contribution

Existing research includes studies of automated or generative-AI feedback, teacher feedback, and comparisons between feedback conditions. However, less attention has been given to the intersection examined here: the same IELTS learners evaluating heterogeneous AI-enabled and human teacher feedback across parallel writing dimensions while also providing quantitative trust ratings and qualitative explanations of their judgments. This within-participant, IELTS-specific comparison allows perceived usefulness, domain-level judgments, and trust to be examined together without assuming that either feedback source is uniformly superior.

2.6. Integrated Conceptual Framework

The theoretical framework integrates four perspectives: perceived usefulness, L2 writing feedback theory, IELTS Writing assessment criteria, and trust in automation.

First, Davis's (1989) perceived usefulness construct explains why learners may value feedback that helps them revise faster, correct errors, or improve drafts more efficiently. *Second,* L2 writing feedback theory highlights the pedagogical and contextual functions of human teacher feedback. Feedback is not only corrective but also pedagogical and contextual (Hyland & Hyland, 2006). *Third,* IELTS Writing assessment criteria provide the domain-specific structure of the study. The questionnaire compares AI and teacher feedback across grammar, vocabulary, coherence and cohesion, task response, idea development, and revision speed. *Fourth,* trust-in-automation theory explains how learners decide whether to accept AI feedback (Jian et al., 2000), while L2 writing research shows that trust affects engagement with automated feedback (Ranalli, 2021).

The four components are linked through an analytical sequence rather than treated as independent labels. Within an IELTS Writing task context, learners experience feedback from a particular source and evaluate whether that feedback is understandable, relevant, usable, and sufficiently trustworthy for the task. These evaluations inform two central judgments in the present study, perceived usefulness and trust, which in turn help interpret learners' reported preferences and feedback-use decisions. Feedback literacy provides an interpretive lens for the learner-evaluation stage (Carless & Boud, 2018; Zhang & Mao, 2023), but it was not measured as a mediator or separate quantitative construct. The arrows in this framework represent analytical relationships rather than tested causal pathways.

Together, these relationships form an integrated framework for interpreting task-sensitive evaluations of the two feedback sources. The study does not assume that either AI or teacher feedback is superior, but investigates whether learners assign different roles to each source. This framework directly informs the research questions and provides the basis for interpreting the findings. Figure 1 summarizes the integrated conceptual framework and the analytical relationships among feedback-source experience, the IELTS Writing task context, learner evaluative processing, perceived usefulness, task-sensitive trust, and reported feedback use.

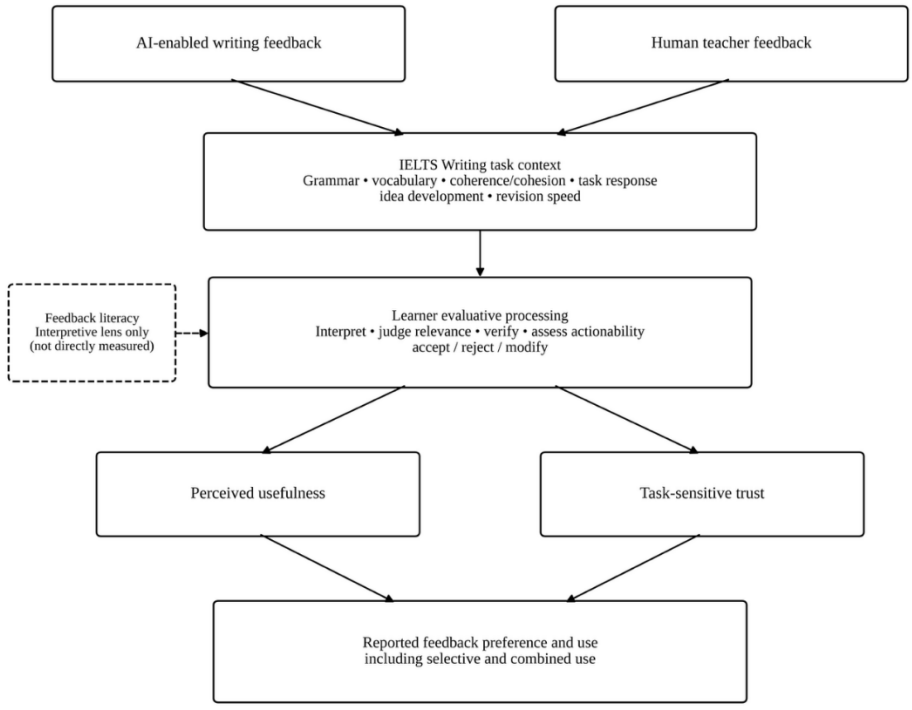

Figure 1. Integrated Conceptual Framework for Learner Evaluation of AI-Enabled and Human Teacher Feedback in IELTS Writing

Note. The figure synthesizes the conceptual relationships specified in Section 2.6. Arrows represent analytical rather than tested causal relationships.

3. MATERIALS AND METHODS

3.1. Research Design

This study employed an explanatory sequential mixed-methods design (Creswell & Plano Clark, 2018; Ivankova et al., 2006). *In the first phase,* quantitative questionnaire data were collected to identify learners' comparative ratings of the two feedback sources. The questionnaire also contained five open-ended items (OE1–OE5) addressing perceived strengths, differential trust, and source preference; these responses provided the connecting mechanism for selecting information-rich participants for the qualitative follow-up. *In the second phase,* semi-structured interviews elaborated the same comparative domains.

RQ1 was addressed through composite comparisons of AI and human teacher feedback perceived improvement. RQ2 was addressed through paired comparisons of grammar, vocabulary, coherence and cohesion, task response, idea development, and revision speed. RQ3 was addressed through comparisons of AI and teacher trust, supported by interviews. Integration occurred again at the interpretation stage, where corresponding quantitative and qualitative findings were compared by analytical domain and summarized in a joint display.

3.2. Research Context and Participants

The study was conducted at a private IELTS teaching center in Ho Chi Minh City, Vietnam. Participants were recruited through convenience sampling. To be eligible, they had to be at least 18 years old, have prepared for IELTS Writing, and have experience with both AI-based and human teacher feedback. Screening questions were used to confirm eligibility.

The final quantitative sample consisted of 130 IELTS Writing learners, including recent high school graduates, university students, and working adults. Eight participants were purposively selected from the questionnaire sample for qualitative follow-up. Selection was informed primarily by responses to OE1–OE5, which addressed the areas in which each feedback source was perceived as most helpful, circumstances underlying differential trust, and overall source preference. Priority was given to information-rich responses that could support fuller explanation of these comparative domains. Selection was not based on prespecified high/low trust scores, AI-use thresholds, or extreme-case cut-offs.

3.3. Research Instrument

The main instrument was a paired-source perception questionnaire that allowed learners to evaluate AI and human teacher feedback across parallel IELTS Writing dimensions. All main perception items used a five-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree.

The questionnaire collected demographic and IELTS preparation information and included eight parallel items for each feedback source. These items focused on overall writing performance, grammar, vocabulary, coherence and cohesion, task response, idea development, revision speed, and understanding of writing weaknesses. Four parallel items measured trust in each source, and seven direct comparative items examined preferences for specific writing areas and combined feedback use.

The questionnaire was researcher-developed for the present IELTS Writing context rather than reproduced from a previously validated instrument. Item content was informed by perceived usefulness, L2 WF research, official IELTS Writing criteria, and trust-in-automation research. Parallel wording was used wherever possible across AI-enabled and human teacher feedback to support within-participant comparison. A construct-to-item specification of the questionnaire is provided in Supplementary Table S1. DEM7 recorded participants' main AI feedback tool rather than exclusive tool use; therefore, tool categories were used descriptively and were not treated as mutually exclusive platform interventions.

3.4. Interview Protocol

Semi-structured interviews were conducted after the questionnaire phase to elaborate participants' comparative experiences with the two feedback sources. The guide contained ten core questions covering five domains: experience with each source, perceived usefulness across writing areas, differential trust, problematic or difficult-to-apply feedback, and combined use and overall source preference.

The ten questions provided a common structure across participants while allowing follow-up probes when clarification or concrete examples were needed. The protocol was aligned with the questionnaire and research questions but was intended to elicit conditions, reasoning, examples, and revision experiences rather than simply repeat the closed-response items. The complete interview guide is provided in Appendix A. The alignment of the interview questions with the questionnaire components and research questions is summarized in Supplementary Table S2.

3.5. Data Collection Procedure

Data collection was conducted in two phases. *First,* eligible learners completed an online questionnaire after reading a consent statement explaining the study purpose, voluntary participation, confidentiality, and academic use of the data. *Second,* responses to OE1–OE5 were reviewed to identify information-rich questionnaire participants for follow-up, and eight participants were purposively selected. Semi-structured interviews were then conducted to

elaborate the same broad writing-dimension, trust, and feedback-use domains examined in the questionnaire.

3.6. Quantitative Data Analysis

Quantitative data were analyzed using R through RStudio (Posit team, 2025; R Core Team, 2025). Composite variables were computed by averaging the relevant items for AI perceived improvement, teacher perceived improvement, AI trust, and teacher trust.

Cronbach's alpha was used to examine the internal consistency of the four scales, with values above .70 considered acceptable (Nunnally & Bernstein, 1994). Descriptive statistics were calculated for the composite variables and comparative-preference items. Paired-samples *t*-tests compared AI and teacher feedback for overall improvement, grammar, vocabulary, coherence and cohesion, task response, idea development, revision speed, and trust. Cohen's *dz* was reported as the paired-comparison effect size (Cohen, 1988). The direct comparative items were analyzed descriptively only. DEM7 was not used as a primary inferential grouping variable because it recorded only the main AI tool rather than exclusive platform exposure, and the tool groups were substantially unequal in size. Tool composition was therefore interpreted descriptively rather than as evidence of platform-specific effects.

3.7. Qualitative Data Analysis

Interview data were analyzed in QualCoder (Curtain, n.d.), a qualitative data analysis tool supporting systematic coding and data organization (Brailas et al., 2023), using hybrid deductive–inductive thematic analysis (Fereday & Muir-Cochrane, 2006). Initial deductive codes were based on the research questions and IELTS Writing dimensions, while inductive codes were added for recurring ideas not captured by the initial list. Cross-case comparison subsequently examined how participants accepted, rejected, or modified feedback; responded to disagreements between sources; verified or filtered AI suggestions; and adapted their AI-use practices.

The transcripts were read in full and coded using meaningful qualitative units, defined as phrases, sentences, or short passages expressing one complete idea (Chenail, 2012). Related codes were grouped into broader themes, and coding memos documented decisions to create, merge, rename, or group codes. QualCoder frequency summaries and coding-density outputs supported interpretation but were not treated as statistical tests. After thematic analysis, quantitative and qualitative findings were organized by common analytical domain and juxtaposed in a joint display (Fetters et al., 2013). The two strands were examined for convergence, elaboration, qualification, or divergence, and an integrated interpretation was developed for each domain.

3.8. Instrument Development, Reliability, and Trustworthiness

The quantitative instrument was developed specifically for the present study. Its item specification linked the paired feedback items to perceived usefulness, L2 WF research, official IELTS Writing domains, and trust-in-automation research. Grammar, vocabulary, coherence and cohesion, and task response were anchored in IELTS assessment domains, whereas idea development, revision speed, understanding of writing weaknesses, and overall perceived improvement were treated as study-specific feedback dimensions. The AI-enabled and human-teacher versions used parallel wording to support direct comparison.

Internal consistency was examined separately for the four composite scales. Cronbach's alpha was .87 for AI-enabled-feedback perceived improvement, .87 for human-teacher-feedback perceived improvement, .85 for AI trust, and .82 for human-teacher trust. These coefficients are interpreted as reliability evidence rather than evidence of comprehensive

construct validity. No documented expert-panel validation, cognitive interviewing, or factor-analytic validation was available; accordingly, the manuscript does not claim that the instrument has been fully psychometrically validated.

For the qualitative phase, trustworthiness was supported through systematic coding, memo writing, and an auditable QualCoder project. These procedures made the analysis more transparent and traceable (Nowell et al., 2017).

3.9. Ethical Considerations

Participation was voluntary and limited to learners aged 18 years or above. Before data collection, participants were informed about the purpose of the research, confidentiality, academic use of the data, and their right to withdraw. No personal names or personally identifiable information were reported, and interview participants were represented using pseudonymous participant codes (P1–P8) for analysis and reporting.

4. RESULTS

4.1. Reliability and Participant Profile

As shown in Table 1, all four scales demonstrated acceptable to good internal consistency. Cronbach's alpha was .87 for AI perceived improvement, .87 for human teacher feedback perceived improvement, .85 for AI trust, and .82 for human teacher feedback trust.

Table 1. Reliability of the main questionnaire scales

Scale	Items	Cronbach's α
AI perceived improvement	8	.87
Human teacher feedback perceived improvement	8	.87
AI trust	4	.85
Human teacher feedback trust	4	.82

Note. α = Cronbach's alpha. Values above .70 indicate acceptable internal consistency.

The final sample included 130 learners with experience of both feedback sources. Most reported a current IELTS Writing band of 5.0–6.5 and had prepared for IELTS Writing for 3–12 months. ChatGPT was the most frequently used AI tool, followed by Grammarly. These categories represent participants' main reported AI tool and should not be interpreted as mutually exclusive platform-use groups. Accordingly, the subsequent results concern aggregated perceptions of heterogeneous AI-enabled feedback rather than effects attributable to any individual platform. Most participants used AI feedback and received teacher feedback often or very often (see Table 2).

Table 2. Relevant participant profile (N = 130)

Variable	Category	n	%
Self-reported IELTS Writing band	Below 5.0	8	6.2
	5.0–5.5	40	30.8
	6.0–6.5	63	48.5
	7.0 or above	15	11.5
	Not sure	4	3.1
Length of IELTS preparation	Less than 3 months	23	17.7

Variable	Category	n	%
	3–6 months	55	42.3
	7–12 months	36	27.7
	More than 12 months	16	12.3
Main AI feedback tool used	ChatGPT	61	46.9
	Grammarly	34	26.2
	QuillBot	14	10.8
	IELTS writing application	17	13.1
	Other	4	3.1
Frequency of AI feedback use	Rarely	9	6.9
	Sometimes	32	24.6
	Often	56	43.1
	Very often	33	25.4
Frequency of human teacher feedback	Rarely	10	7.7
	Sometimes	33	25.4
	Often	74	56.9
	Very often	13	10.0
Current status	Recent high school graduate	31	23.8
	University student	67	51.5
	Working adult	29	22.3
	Other	3	2.3

Note. Only participant characteristics directly related to IELTS Writing preparation and feedback experience are reported. Percentages may not total 100 due to rounding.

4.2. Descriptive Statistics

Participants evaluated both feedback sources positively. Human teacher feedback received a slightly higher mean for perceived improvement than AI-enabled writing feedback, while the difference was greater for trust, with human teacher feedback receiving the higher rating (see Table 3).

Table 3. Descriptive statistics for main composite variables

Variable	M	SD
AI perceived improvement	3.78	0.67
Human teacher feedback perceived improvement	3.88	0.68
AI trust	3.61	0.81
Human teacher feedback trust	4.13	0.68

Note. All variables were measured on a five-point Likert scale, from 1 = strongly disagree to 5 = strongly agree.

4.3. Paired-Samples Comparisons

There was no statistically significant difference in overall perceived improvement, $t(129) = -1.50$, $p = .137$, $dz = -0.13$. However, AI-enabled writing feedback

was rated significantly higher for grammar correction, vocabulary improvement, and revision speed. Human teacher feedback was rated significantly higher for coherence and cohesion, task response, idea development, and trust (see Table 4). The effects were small for grammar and vocabulary, whereas the differences for the remaining dimensions were moderate or approximately medium in size. Effect sizes were interpreted following conventional reporting guidance (Cohen, 1988; Lakens, 2013).

Table 4. Paired-samples comparisons between AI-enabled feedback and human teacher feedback

Dimension	AI M	Human M	Mean difference	t	df	p	Cohen's dz
Overall perceived improvement	3.78	3.88	-0.10	-1.50	129	.137	-0.13
Grammar correction	4.07	3.78	0.29	2.88	129	.005	0.25
Vocabulary improvement	4.01	3.80	0.21	2.00	129	.048	0.18
Coherence and cohesion	3.48	4.01	-0.52	-4.81	129	< .001	-0.42
Task response	3.34	4.02	-0.68	-6.50	129	< .001	-0.57
Idea development	3.47	4.12	-0.65	-5.77	129	< .001	-0.51
Revision speed	4.22	3.60	0.62	6.04	129	< .001	0.53
Trust	3.61	4.13	-0.52	-6.98	129	< .001	-0.61

Note. Mean difference was calculated as AI feedback minus human teacher feedback. Positive values indicate higher ratings for AI feedback; negative values indicate higher ratings for human teacher feedback. Cohen's dz was used for paired-samples effect size.

4.4. Direct Comparative Preferences

The direct comparative items supported the paired-samples results. Learners preferred AI feedback for grammar correction, vocabulary improvement, and fast revision, but preferred human teacher feedback for coherence and cohesion, task response, and idea development. Combining both sources received the highest mean score (see Table 5).

Table 5. Direct comparative preference items

Item	Comparative preference	M	SD
CP1	AI feedback was preferred for grammar correction	4.05	0.83
CP2	AI feedback was preferred for vocabulary improvement	3.87	0.88
CP3	Human teacher feedback was preferred for coherence and cohesion	4.13	0.81
CP4	Human teacher feedback was preferred for task response	4.14	0.89
CP5	Human teacher feedback was preferred for idea development	4.05	0.80
CP6	AI feedback was preferred for fast revision	4.28	0.78
CP7	Combining AI and human teacher feedback was preferred overall	4.32	0.75

Note. Direct comparative preference items were analyzed descriptively and were not included in the reliability analysis because they measured specific comparative judgments

rather than one unified scale. The respondent-facing wording "AI feedback" is retained in the comparative items as originally administered.

4.5. Qualitative Findings

The thematic analysis generated five major themes and twelve sub-themes, with 104 coded references, following the identification and grouping of repeated patterns across the interview data (Braun & Clarke, 2006). The five themes were AI-enabled feedback as micro-level revision support, limitations of AI-enabled feedback, human teacher feedback as higher-order writing development, task-sensitive trust, and combined feedback use (see Figure 2).

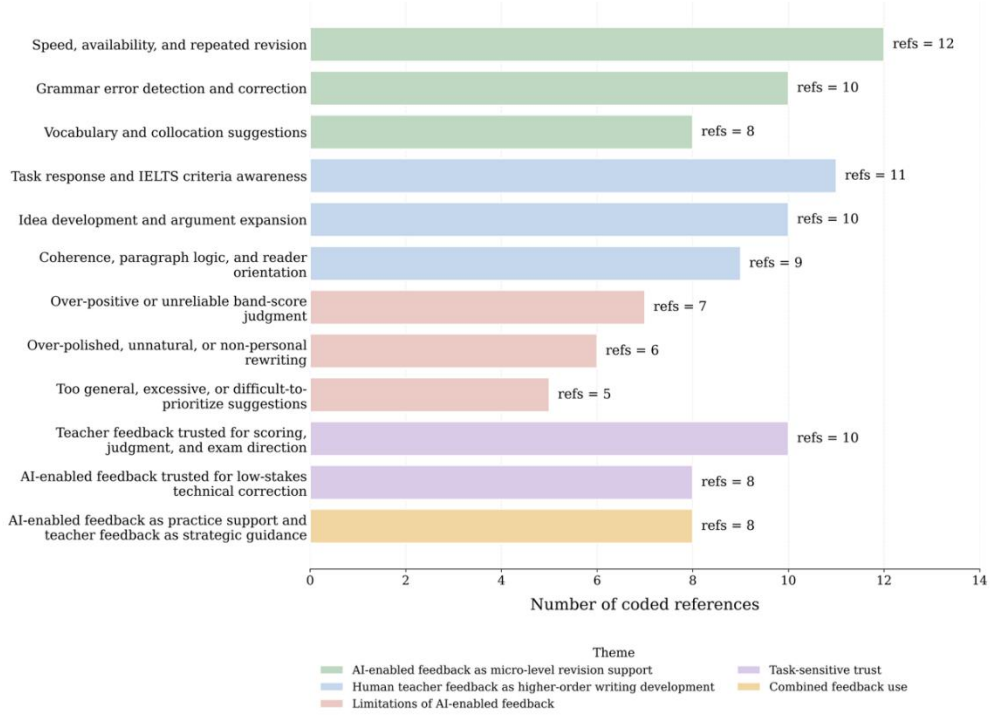

Figure 2. Coding Density of Qualitative Subthemes

Note. Bars show coded references by subtheme.

Beyond confirming the dimension-level survey patterns, the interviews showed how learners evaluated and filtered feedback. P2 initially copied AI-generated rewrites but later changed strategy: "Now I ask AI to list mistakes and explain them." P3 similarly described filtering changes that altered the writer's style, while P8 reported using vocabulary suggestions only when the words were understood. These accounts indicate that feedback uptake was selective rather than automatic.

Conflicts between AI-enabled and teacher feedback also changed learners' trust and subsequent use. P4 described receiving an AI band estimate of 7.0 for an essay that the teacher judged to be approximately Band 6.0 because of weak task response and reported, "Since then, I do not ask AI for band score seriously." P7 similarly stopped copying AI sentences directly after a polished AI-generated conclusion was judged by the teacher to sound memorized and unnatural. These experiences narrowed the purposes for which learners considered AI feedback dependable rather than leading them to reject it entirely.

The interviews also qualified the generally positive interpretation of human teacher feedback. P2 described some teacher comments as emotionally difficult but ultimately useful;

P4 characterized teacher feedback as "harder to accept emotionally, but … useful," and P6 described it as discouraging but realistic. Higher trust in teachers therefore did not necessarily correspond to greater emotional comfort or immediate actionability.

Participants also described feedback usefulness as sensitive to the way AI was queried and to the stage of revision. P7 trusted AI more when asking "simple and specific questions," while P5 noted that broad comments became more useful when followed by more targeted questions. Several participants described staged use. P3 used AI for an initial language-level check, sought teacher feedback on higher-order problems, and then returned to AI for further practice; P8 reported a similar AI–teacher–AI sequence. These accounts suggest perceived complementarity in feedback use, not evidence that the combined sequence objectively produces superior writing outcomes. Supplementary Table S3 provides a cross-case summary of feedback evaluation, source conflict, and adaptation patterns across P1–P8.

For AI-enabled writing feedback, the most frequent sub-theme was speed, availability, and repeated revision, followed by grammar correction and vocabulary and collocation suggestions. Learners valued AI because it allowed immediate and repeated revision. One participant explained, "I can write, check, revise, and write again in one evening."

For human teacher feedback, the most frequent sub-themes were task response and IELTS criteria awareness, idea development and argument expansion, and coherence and paragraph logic. These themes showed that learners valued teacher feedback for aspects requiring interpretation and judgment. One participant stated, "AI corrected sentences, but teacher corrected the paragraph logic."

Participants also identified several limitations of AI feedback, including over-positive or unreliable band-score judgments, overly polished or impersonal rewriting, and suggestions that were too general or difficult to prioritize. Complementary use emerged when learners used AI for repeated practice and technical revision while relying on teachers for strategic and evaluative guidance.

Learners trusted AI mainly for low-stakes technical correction, such as grammar, vocabulary, and sentence-level revision. They trusted teachers more for scoring, task response, coherence, idea development, and exam direction. Overall, the qualitative findings supported the quantitative distinction between AI-assisted micro-level revision and teacher-supported higher-order writing development.

4.6. Mixed-Methods Integration

The quantitative and qualitative findings were integrated by common analytical domain. Table 6 juxtaposes the principal quantitative pattern, corresponding qualitative explanation, and integrated interpretation. The final column represents mixed-methods interpretation rather than additional statistical analysis.

Table 6. Mixed-methods joint display of quantitative and qualitative findings by analytical domain

Domain	Quantitative pattern	Qualitative explanation	Integrated interpretation
Overall perceived improvement	No significant difference, $p = .137$	Sources valued for different purposes	Similar overall ratings masked differentiated perceived functions

Domain	Quantitative pattern	Qualitative explanation	Integrated interpretation
Grammar/vocabulary	AI-enabled ratings higher	Rapid, readily verifiable language correction	Convergence on language-level perceived usefulness
Coherence/task/ideas	Teacher ratings higher	Context, logic, task interpretation	Convergence on higher-order/contextual perceived usefulness
Revision speed	AI-enabled rating higher	Immediate availability and repeated practice	Advantage concerned perceived immediacy, not objective effectiveness
Trust	Teacher 4.13 vs AI 3.61	Selective AI trust; teacher for judgment; teacher feedback could still be uncomfortable	Trust was task-sensitive and distinct from emotional comfort
Combined use	CP7 M = 4.32	Staged AI–teacher–AI use	Evidence of perceived complementarity, not demonstrated performance superiority

5. DISCUSSION

5.1. Perceived Writing Improvement from AI and Human Teacher Feedback

The findings suggest that learners did not perceive either feedback source as clearly superior in overall writing improvement. Instead, both were regarded as useful, but their usefulness operated in different ways. Within the integrated conceptual framework, perceived usefulness is interpreted here as a learner judgment about whether a feedback source was helpful for a particular writing purpose, not as evidence of objective writing improvement or as a test of the full Technology Acceptance Model. The interview findings further indicate that usefulness depended on learner evaluation: participants considered whether suggestions were understandable, appropriate to their level or voice, and sufficiently actionable before using them.

Learners valued AI-enabled WF because it made revision faster, more accessible, and easier to repeat outside the classroom, consistent with the concept of perceived usefulness (Davis, 1989). However, perceived usefulness did not mean replacement of teacher feedback. Learners still relied on teachers for explanation, judgment, and IELTS-oriented guidance. This pattern is consistent with viewing AI-enabled feedback as one component of a broader feedback ecology rather than as a direct substitute for teacher feedback.

5.2. Writing Areas Where AI and Human Teacher Feedback Were Perceived Differently

AI-enabled WF was perceived as more useful for grammar correction, vocabulary improvement, and fast revision. These are micro-level aspects that can be checked and revised quickly. In contrast, human teacher feedback was perceived as more useful for coherence and

cohesion, task response, and idea development, which require interpretation of the task, argument development, paragraph logic, and awareness of reader expectations.

This distinction reflects the multidimensional nature of IELTS Writing assessment (IELTS, 2023) and is consistent with L2 writing research that views teacher feedback as pedagogical and contextual rather than purely corrective (Hyland & Hyland, 2006). Learners associated AI-enabled feedback more strongly with language-level revision, whereas they associated teacher feedback more strongly with identifying higher-order and contextual writing problems. The interview data suggest one mechanism for this distinction: learners were more willing to act on suggestions that they could independently understand and verify, whereas task response, argument development, and reader-oriented coherence often required broader contextual judgment. Feedback literacy is used here as an interpretive lens for this evaluation and uptake process; it was not directly measured as a quantitative construct.

5.3. Task-Sensitive Trust in AI-Enabled and Human Teacher Feedback

Learners trusted human teacher feedback more overall, but this should not be interpreted as simple distrust of AI. Rather, they showed task-sensitive and selective trust: they trusted AI for low-stakes technical correction but trusted teachers for high-stakes evaluative judgment. Trust also changed through experience. Conflicting AI and teacher judgments, unsuitable rewriting, and overly positive AI evaluation led some interviewees to reduce reliance on AI for particular functions while continuing to use it for narrower, readily verifiable tasks. Because the study did not independently evaluate the accuracy of either feedback source, these patterns should not be interpreted as evidence that trust was objectively calibrated.

This pattern is consistent with trust-in-automation research, which suggests that trust depends on the perceived reliability of a system for a particular task (Jian et al., 2000). Learners trusted AI when checking grammar, vocabulary, and sentence-level alternatives because these tasks were relatively easy to verify. They trusted teachers more for IELTS scoring, task response, coherence, and idea development because these required expert interpretation.

The interview findings also showed that AI feedback could be too general, overly positive, or too polished. This is consistent with research showing that ChatGPT feedback on coherence and cohesion can be abstract, generic, or inaccurate (Yoon et al., 2023). Learners' lower trust in AI therefore reflected awareness of its limitations rather than rejection of technology.

5.4. Integrated Interpretation of Feedback-Source Use

Across the three research questions, the integrated evidence indicates differentiated and task-sensitive evaluations of the two feedback sources. Participants described AI-enabled feedback mainly as a practice and revision tool, helping them notice language errors, generate vocabulary options, and revise quickly. They described human teacher feedback as a pedagogical and evaluative guide, particularly for task requirements, paragraph logic, idea development, and IELTS expectations. The expanded qualitative analysis clarifies the evaluative process underlying these judgments: learners interpreted feedback, compared it with their own knowledge or teacher guidance, and then accepted, rejected, modified, or reused suggestions. Combined feedback use may therefore be described as perceived complementarity, but the present study does not establish that combining the two sources produces superior IELTS Writing outcomes.

The highest rating for combined feedback indicates that learners did not perceive the two sources as direct competitors. This finding contributes to AI-assisted L2 writing research by moving beyond the question of whether AI feedback is good or bad. The more relevant

issue is how AI and teacher feedback can be combined responsibly within a broader feedback ecology.

5.5. Pedagogical Implications

IELTS Writing instruction should not frame AI-enabled WF as either a threat or a replacement for teachers. Instructors may consider allowing learners to use AI-enabled feedback before class to identify grammar errors, examine vocabulary alternatives, and prepare a cleaner draft. Teacher feedback may then place greater emphasis on task response, coherence, argument development, and IELTS assessment expectations.

After teacher feedback, learners may use AI for further revision and practice, but should be taught to verify suggestions rather than accept them automatically. In such a staged approach, AI may function as a revision assistant, while teacher feedback may provide strategic and evaluative guidance. Such staged use may also encourage feedback-literacy practices by requiring learners to decide when AI suggestions are useful, when they should be questioned, and when further teacher guidance is needed. The staged AI–teacher–AI sequence reported by several interviewees should be regarded as a learner-reported practice worth testing, rather than as an established optimal instructional model.

5.6. Limitations and Future Research

This study examined learners' perceptions rather than directly measured changes in IELTS Writing performance. It also relied partly on self-reported proficiency and was conducted at one private IELTS center. The qualitative phase involved eight interview participants, which provided explanatory depth but limited broader generalization. Several additional limitations qualify the interpretation of the findings. First, the researcher-developed questionnaire showed acceptable internal consistency but was not subjected to documented independent content-validation or factor-analytic validation. Second, DEM7 recorded only participants' main AI tool, and the heterogeneous tools represented in the sample cannot be treated as mutually exclusive platform interventions. Third, the qualitative follow-up was purposively informed by information-rich open-ended responses rather than by a prespecified extreme-case or high/low-score matrix. Finally, the conceptual framework was used to interpret perceived usefulness, trust, evaluative processing, and reported feedback use; these relationships were not tested as causal or mediational pathways.

Future research could compare AI-only, teacher-only, and combined-feedback conditions using independently rated pre-test and post-test writing. Longitudinal and multi-site studies could also examine whether learners' trust and feedback use change across proficiency levels and over time.

6. CONCLUSION

This study investigated IELTS Writing learners' perceptions of AI-enabled WF and human teacher feedback in relation to perceived improvement, specific writing areas, and trust. The findings showed no significant difference in overall perceived writing improvement. However, AI-enabled writing feedback was rated higher for grammar correction, vocabulary improvement, and revision speed, whereas human teacher feedback was rated higher for coherence and cohesion, task response, idea development, and trust.

The findings therefore support a conclusion of differentiated and task-sensitive learner perceptions rather than the objective superiority of either feedback source. Learners associated AI-enabled WF particularly with rapid language-level revision, while reporting greater reliance on human teacher feedback for several higher-order and evaluative judgments. Some participants described the two sources as complementary within their own revision practices,

but this perceived complementarity should not be interpreted as evidence that combined feedback produces better IELTS Writing performance.

Learners also reported task-sensitive trust, using AI-enabled feedback selectively for readily verifiable technical assistance while relying more strongly on teachers for scoring, task interpretation, coherence, and strategic IELTS guidance. The contribution of the study is therefore to clarify how learners evaluated usefulness and trust across the two feedback-source categories and how those evaluations informed their reported feedback-use decisions.

DECLARATIONS

Funding: This research received no external funding.

Institutional Review Board Statement: No formal Institutional Review Board or ethics committee approval was obtained for this minimal-risk educational study. All participants were aged 18 years or above and took part voluntarily after being informed of the study purpose, confidentiality, and their right to withdraw. No personally identifiable information was reported, and the data were handled confidentially, with interview participants represented using pseudonymous codes for academic research purposes.

Transparency: The author confirms that the manuscript presents an accurate and transparent account of the research. All essential aspects of the study design, data collection, data analysis, findings, limitations, and interpretation have been reported.

Competing Interests: The author declares that there are no competing interests.

Author's Contributions: The author was responsible for the conceptualization of the study, research design, methodology, questionnaire development, data collection, data analysis, interpretation of findings, and manuscript preparation. The author has read and agreed to the published version of the manuscript.

Disclosure of AI Use: AI tools were used to support language editing, proofreading, and refinement of academic expression during manuscript preparation. The author reviewed, verified, and approved all AI-assisted content and remains fully responsible for the accuracy, integrity, and originality of the manuscript.

REFERENCES

Brailas, A., Tragou, E., & Papachristopoulos, K. (2023). Introduction to qualitative data analysis and coding with QualCoder. *American Journal of Qualitative Research, 7*(3), 19-31. https://doi.org/10.29333/ajqr/13230

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Chenail, R. J. (2012). Conducting qualitative data analysis: Reading line-by-line, but analyzing by meaningful qualitative units. *The Qualitative Report, 17*(1), 266-269. https://doi.org/10.46743/2160-3715/2012.1817

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Crosthwaite, P., & Sun, S. (2026). Generative AI and L2 written feedback studies: A scoping review. *RELC Journal, 57*(1), 207-219. https://doi.org/10.1177/00336882251386530

Curtain, C. (n.d.). *QualCoder* [Computer software]. GitHub. https://github.com/ccbogel/QualCoder

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

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 5*(1), 80-92. https://doi.org/10.1177/160940690600500107

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—Principles and practices. *Health Services Research, 48*(6 Pt 2), 2134-2156. https://doi.org/10.1111/1475-6773.12117

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching, 39*(2), 83-101. https://doi.org/10.1017/S0261444806003399

IELTS. (2023, May 3). *IELTS Writing band descriptors and key assessment criteria*. https://ielts.org/news-and-insights/ielts-writing-band-descriptors-and-key-assessment-criteria

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods, 18*(1), 3-20. https://doi.org/10.1177/1525822X05282260

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*(1), 53-71. https://doi.org/10.1207/S15327566IJCE0401_04

Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System, 123*, Article 103332. https://doi.org/10.1016/j.system.2024.103332

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning, 35*(4), 605-634. https://doi.org/10.1080/09588221.2020.1743323

Lu, Q., Zhu, X., Zhu, S., & Yao, Y. (2024). Effects of writing feedback literacies on feedback engagement and writing performance: A cross-linguistic perspective. *Assessing Writing, 62*, Article 100889. https://doi.org/10.1016/j.asw.2024.100889

Mao, Z., Lee, I., & Li, S. (2024). Written corrective feedback in second language writing: A synthesis of naturalistic classroom studies. *Language Teaching, 57*(4), 449-477. https://doi.org/10.1017/S0261444823000393

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods, 16*(1), 1-13. https://doi.org/10.1177/1609406917733847

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Pearson, W. S. (2022). Student engagement with teacher written feedback on rehearsal essays undertaken in preparation for IELTS. *SAGE Open, 12*(1). https://doi.org/10.1177/21582440221079842

Posit team. (2025). *RStudio: Integrated development environment for R* [Computer software]. Posit Software, PBC. https://posit.co/

R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing, 52*, Article 100816. https://doi.org/10.1016/j.jslw.2021.100816

Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. *Frontiers in Psychology, 14*, Article 1249991. https://doi.org/10.3389/fpsyg.2023.1249991

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies, 28*(11), 13943-13967. https://doi.org/10.1007/s10639-023-11742-4

Yoon, S.-Y., Miszoglad, E., & Pierce, L. R. (2023). *Evaluation of ChatGPT feedback on ELL writers' coherence and cohesion* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2310.06505

Zhang, T., & Mao, Z. (2023). Exploring the development of student feedback literacy in the second language writing classroom. *Assessing Writing, 55*, Article 100697. https://doi.org/10.1016/j.asw.2023.100697

Zhang, Z. V., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing, 36*, 90-102. https://doi.org/10.1016/j.asw.2018.02.004

Zhang, Z., & Hyland, K. (2025). The role of digital literacy in student engagement with automated writing evaluation (AWE) feedback on second language writing. *Computer Assisted Language Learning, 38*(5-6), 1060-1085. https://doi.org/10.1080/09588221.2023.2256815

Appendix A

Semi-Structured Interview Guide

1. Can you describe your experience of using AI feedback for IELTS Writing preparation?
2. Can you describe your experience of receiving human teacher feedback?
3. In which writing areas does AI feedback help you most, such as grammar, vocabulary, coherence, task response, or idea development? Why?
4. In which writing areas does human teacher feedback help you most? Why?
5. When do you trust AI feedback more than teacher feedback?
6. When do you trust teacher feedback more than AI feedback?
7. Have you ever received AI feedback that you felt was inaccurate, unclear, or too general? What happened?
8. Have you ever received teacher feedback that was difficult to apply? What happened?
9. How do you usually combine AI feedback and teacher feedback when revising an essay?
10. If you had to choose one feedback source for IELTS Writing preparation, which one would you choose? Why?

Note. The wording "AI feedback" is retained because it reproduces the interview guide administered to participants. In the revised manuscript, the broader analytical category is referred to as AI-enabled writing feedback.

Supplementary Table S1

Table S1. Questionnaire Construct-to-Item Specification

Questionnaire component	Items	Intended content	Conceptual/ domain basis	Interpretation
AI-enabled feedback perceived improvement	AI1-AI8	Overall perceived improvement, grammar, vocabulary, coherence/cohesion, task response, idea development, revision speed, understanding weaknesses	Davis (1989); IELTS Writing criteria; L2 writing feedback literature	Context-specific perceptions of AI-enabled feedback usefulness
Human teacher feedback perceived improvement	HT1-HT8	Same eight domains	IELTS Writing criteria; L2 writing feedback literature; parallel-source design	Context-specific perceptions of human teacher feedback usefulness

Questionnaire component	Items	Intended content	Conceptual/ domain basis	Interpretation
AI-enabled feedback trust	AIT1-AIT4	Trust in problem identification, confidence applying feedback, perceived reliability, clarity	Jian et al. (2000), contextually rewritten	Context-specific trust composite; not the complete original Jian et al. scale
Human teacher feedback trust	HTT1-HTT4	Same trust-related domains	Parallel-source construction; L2 feedback literature	Human-teacher trust composite
Direct comparative preferences	CP1–CP7	Grammar, vocabulary, coherence, task response, ideas, revision speed, combined use	Study-specific direct comparison	Individual comparative judgments; not treated as one psychometric scale
Open-ended follow-up items	OE1-OE5	Source strengths, differential trust, overall preference	Study research questions	Used to inform purposive selection for the qualitative follow-up

Supplementary Table S2

Table S2. Alignment of Interview Questions With the Quantitative Phase and Research Questions

Interview question	Main purpose	Questionnaire connection	RQ
Q1	Experience with AI-enabled feedback	AI1–AI8; tool/use background	RQ1–RQ2
Q2	Experience with teacher feedback	HT1–HT8; teacher-use background	RQ1–RQ2
Q3	Areas where AI helps most	AI2–AI7; OE1; CP1/CP2/CP6	RQ2
Q4	Areas where teacher helps most	HT2–HT7; OE2; CP3–CP5	RQ2
Q5	Conditions for greater AI trust	AIT1–AIT4; OE3	RQ3
Q6	Conditions for greater teacher trust	HTT1–HTT4; OE4	RQ3
Q7	Problematic AI feedback	AI perceptions and trust	RQ2–RQ3
Q8	Difficult-to-apply teacher feedback	Teacher perceptions and trust	RQ2–RQ3
Q9	Combined use during revision	CP7	RQ1–RQ3
Q10	Overall source choice and rationale	OE5 / comparative preference	RQ1–RQ3

Supplementary Table S3

Table S3. Cross-Case Explanatory Patterns in the Interview Data

Participant	Selective uptake / checking	Conflict or qualification	Adaptation in feedback use
P1	Accepts straightforward grammar correction but checks substantial rewriting	AI band estimate higher than teacher; some AI vocabulary judged unnatural	Uses AI for examples when teacher identifies a higher-order problem
P2	Checks synonym meaning and does not accept all AI suggestions	AI encouraged complexity while teacher emphasized clarity	Shifted from copying rewrites to requesting mistake identification and explanation
P3	Treats AI as a source of options and filters unnecessary changes	AI may smooth language without resolving logical problems	Uses AI before and after teacher feedback
P4	Selective about AI paraphrasing	AI Band 7.0 versus teacher approximately Band 6.0	Stopped taking AI band estimates seriously
P5	Rejects vocabulary that does not fit personal voice	AI comments can be too general; teacher is more direct	Uses follow-up questions and AI for repeated practice
P6	Rejects rewriting that no longer reflects own level	AI may agree too easily; teacher offers reader-oriented judgment	Uses AI for explanation/examples while retaining teacher judgment
P7	Trusts AI more for narrow, specific questions	AI-polished conclusion judged memorized/unnatural by teacher	Became more cautious and stopped copying AI sentences directly
P8	Uses only vocabulary that is understood	AI may provide too many suggestions; teacher prioritizes problems	Uses an AI → teacher → AI revision sequence