SANTIAGO E-ISSN 2227-6513

SANTIAGO, 168, 2026
Artículo de Investigación

Human–AI Feedback in EFL Academic Writing: Perceptions and Complementarity in Ecuadorian Education

Retroalimentación humano–IA en la escritura académica en inglés: percepciones y complementariedad en la educación ecuatoriana

Feedback humano–IA na escrita acadêmica em inglês: percepções e complementaridade no educação equatoriana

Edisson Javier Vélez Sánchez[1], https://orcid.org/0009-0002-5156-1556
Jostin Javier Fernández Macías[2], https://orcid.org/0009-0004-5323-799X
Pamela Nicolle Cedeño Loor[1], https://orcid.org/0009-0003-5047-4140
Steeven Josue Valencia Cedeño[3], https://orcid.org/0000-0002-1886-5534
Miguel Macías Loor[4], https://orcid.org/0000-0002-5958-3541

[1]Investigador independiente, Ecuador
[2]Unidad Educativa San Rafael, Ecuador
[3]Unidad Educativa Bilingüe Arco Iris, Ecuador
[4]Facultad de Ciencias de la Educación de la Universidad Técnica de Manabí, Ecuador

Autor para correspondencia: evelez8387@utm.edu.ec

ABSTRACT
Generative artificial intelligence has expanded the sources of feedback available to secondary school students who write in English as a foreign language, but its educational value depends on how automated responses are articulated with teacher judgement rather than on speed alone. This study examined the perceived roles, strengths, and complementarity of teacher and AI-generated feedback in Ecuadorian secondary education. A quantitative, cross-sectional, non-experimental, descriptive-comparative design was used with the original dataset of 60 secondary school students and 5 English-language teachers. Students completed a 48-item questionnaire and teachers a 54-item questionnaire, both using five-point Likert scales. Descriptive statistics, confidence intervals, agreement percentages, paired comparisons, and exploratory non-parametric tests were applied. Students evaluated teacher feedback (M = 4.04) and AI-generated feedback (M = 3.98) favourably, without a statistically significant global difference, t (79) = 1.40, p = .164. AI was mainly associated with objectivity, accessibility, and immediacy, whereas teachers were valued for personalization, understanding learner needs, and explanation quality. Teachers also reported favourable perceptions of institutional readiness (M = 4.07), behavioural intention (M = 4.03), and AI effectiveness (M = 3.99), alongside moderate ethical concern (M = 3.02). Because several item blocks showed weak internal-consistency diagnostics, findings are interpreted primarily at the item level and as exploratory evidence. The results favour a complementary human–AI feedback model while underscoring the need for validated instruments and direct measures of writing performance.
Keywords: Human–AI feedback; academic writing; English as a foreign language; generative artificial intelligence; secondary education.
RESUMEN
La inteligencia artificial generativa ha ampliado las fuentes de retroalimentación disponibles para estudiantes de secundaria que escriben en inglés como lengua extranjera, pero su valor educativo depende de cómo las respuestas automatizadas se articulan con el juicio docente y no únicamente de su rapidez. El estudio examinó los roles percibidos, las fortalezas y la complementariedad de la retroalimentación docente y la generada por IA en la educación secundaria ecuatoriana. Se utilizó un diseño cuantitativo, transversal, no

experimental y descriptivo-comparativo con la base original de 60 estudiantes de secundaria y 5 docentes de inglés. Los estudiantes respondieron un cuestionario de 48 ítems y los docentes uno de 54, ambos con escala Likert de cinco puntos. Se aplicaron estadísticos descriptivos, intervalos de confianza, porcentajes de acuerdo, comparaciones pareadas y pruebas no paramétricas exploratorias. Los estudiantes valoraron favorablemente la retroalimentación docente (M = 4,04) y la generada por IA (M = 3,98), sin diferencia global estadísticamente significativa, t (79) = 1,40, p = 0,164. La IA se asoció principalmente con objetividad, accesibilidad e inmediatez; el docente, con personalización, comprensión de las necesidades del estudiante y calidad de las explicaciones. Los docentes también mostraron percepciones favorables sobre preparación institucional (M = 4,07), intención de uso (M = 4,03) y efectividad de la IA (M = 3,99), junto con preocupación ética moderada (M = 3,02). Debido a que varios bloques presentaron diagnósticos débiles de consistencia interna, los hallazgos se interpretan principalmente a nivel de ítem y como evidencia exploratoria. Los resultados favorecen un modelo complementario humano–IA y señalan la necesidad de instrumentos validados y medidas directas del desempeño escrito.
Palabras clave: Retroalimentación humano–IA; escritura académica; inglés como lengua extranjera; inteligencia artificial generativa; educación secundaria.

RESUMO

A inteligência artificial generativa ampliou as fontes de feedback disponíveis para estudantes do ensino médio que escrevem em inglês como língua estrangeira, mas seu valor educacional depende de como as respostas automatizadas se articulam com o julgamento docente, e não apenas de sua rapidez. O estudo examinou os papéis percebidos, as fortalezas e a complementaridade do feedback docente e do feedback gerado por IA no ensino médio equatoriano. Utilizou-se um desenho quantitativo, transversal, não experimental e descritivo-comparativo com a base original de 60 estudantes do ensino médio e 5 professores de inglês. Os estudantes responderam a um questionário de 48 itens e os professores a um de 54, ambos em escala Likert de cinco pontos. Foram aplicadas estatísticas descritivas, intervalos de confiança, percentuais de concordância, comparações pareadas e testes não paramétricos exploratórios. Os estudantes avaliaram positivamente o feedback docente (M = 4,04) e o gerado por IA (M = 3,98), sem diferença global estatisticamente significativa, t (79) = 1,40, p = 0,164. A IA foi associada principalmente à objetividade, acessibilidade e imediaticidade; o professor, à personalização, compreensão das necessidades e qualidade das explicações. Os docentes também relataram percepções favoráveis de preparação institucional (M = 4,07), intenção de uso (M = 4,03) e efetividade da IA (M = 3,99), com preocupação ética moderada (M = 3,02). Como vários blocos apresentaram diagnósticos frágeis de consistência interna, os resultados são interpretados sobretudo no nível dos itens e como evidência exploratória. Os achados favorecem um modelo complementar humano–IA e reforçam a necessidade de instrumentos validados e medidas diretas do desempenho escrito.
Palavras-chave: Feedback humano–IA; escrita acadêmica; inglês como língua estrangeira; inteligência artificial generativa; ensino médio.

Recibido 21/06/2026 Aprobado 3/07/2026

Introduction

Academic writing feedback is most useful when it helps learners make decisions about revision rather than simply informing them that an error exists. In formative assessment, comments become educationally meaningful when students can understand the expected standard, compare it with their current performance, and act on the information provided (Black & Wiliam, 1998; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006). This process also depends on feedback literacy: students need to interpret comments, judge their relevance, manage affective reactions, and decide what to change in subsequent drafts (Carless & Boud, 2018). In English as a Foreign Language (EFL) academic writing, this decision-making process is demanding because revision involves several levels at once. Students must consider grammar and vocabulary while also attending to coherence, argumentation, organization, academic register, and communicative purpose. Teacher feedback can connect these elements to course objectives and individual learning trajectories, yet the time required to provide detailed comments limits how frequently it can be offered. Automated writing evaluation has long attempted to address this constraint, with research showing useful but heterogeneous effects that depend on the technology, duration of use, proficiency level, and opportunities for revision (Fu ***et al***., 2024; Li & Hegelheimer, 2015; Ngo ***et al***., 2024; Stevenson & Phakiti, 2014).

The central issue is therefore not whether feedback is human or automated in the abstract, but what each

source enables students to do. Teacher comments may be direct, indirect, metalinguistic, global, or focused, and their value depends on timing, task purpose, and the learner's capacity to use them. Similarly, automated comments may be technically accurate yet have little formative value if students accept them without reflection. Research on engagement with feedback has consequently shifted attention from the amount of information delivered to the cognitive and behavioural work students perform after receiving it (Carless & Boud, 2018; Zhang & Hyland, 2018).

Generative artificial intelligence has complicated this comparison because contemporary language models can do more than identify local errors. They can explain a suggestion, reformulate a passage, propose alternatives, respond to follow-up questions, and adapt the apparent level or tone of an answer. These capabilities make GenAI attractive for writing support, but they also introduce risks related to fabricated information, bias, privacy, authorship, dependence, and uncertain accountability (Kasneci ***et al***., 2023; Tlili ***et al***., 2023; UNESCO, 2023; Yan ***et al***., 2023). Effective educational use therefore requires AI literacy that combines prompt design, verification, linguistic judgement, and ethical awareness (Kohnke et al., 2023; Ma ***et al***., 2024; Ng ***et al***., 2021; Walter, 2024).

The conversational character of GenAI can be especially relevant to EFL writers. A student can request clarification, ask for examples, compare formulations, or repeat a question without the social pressure that may accompany classroom interaction. Such accessibility can support autonomy, yet it can also reduce productive struggle if the system becomes a substitute for analysis. Barrot (2023) notes that ChatGPT may generate plausible but contextually inappropriate writing advice, while Kohnke ***et al***. (2023) and Ma ***et al***. (2024) stress that language learners need strategies for evaluating outputs rather than treating fluency as evidence of correctness. Human oversight remains particularly important when feedback affects argumentation, academic voice, or assessed work (Long & Magerko, 2020; UNESCO, 2023).

This tension suggests that the relevant pedagogical question is how responsibilities should be distributed between teachers, students, and AI. Immediate automated feedback may be useful for repeated low-stakes revision, but students still need to diagnose why a suggestion is appropriate, compare alternatives, and preserve ownership of the text. Recent research indicates that guided use can support revision and feedback engagement, whereas uncritical use can encourage superficial changes or dependence (Teng, 2024; Zhan & Yan, 2025; Zou ***et al***., 2025). Perceptions of usefulness should therefore be examined together with satisfaction, intention to use, and ethical concerns.

Evidence comparing human and AI feedback does not establish a universal winner. Steiss ***et al***. (2024) found stronger human performance in several formative-feedback components, even though AI offered substantial time advantages. Guo and Wang (2024) identified potential for ChatGPT to support EFL teacher feedback while also reporting limitations in contextualization and accuracy. Banihashem ***et al.*** (2024) found that AI-generated comments could be extensive and structured, but students still had to evaluate and transform those comments into revisions. These studies indicate that speed and volume are only part of feedback quality.

Hybrid feedback models provide a more productive way to frame the issue. Research combining ChatGPT and teacher feedback has reported favourable outcomes in EFL writing, suggesting that automated linguistic support and human pedagogical judgement may serve different but compatible functions (Asadi ***et al***., 2025; Lo ***et al***., 2025; Zhang ***et al***., 2025). Venter ***et al***. (2025) similarly argue that effective AI-supported feedback requires explicit decisions about who provides feedback, for what purpose, and at which stage of the learning process. Complementarity is therefore a design principle rather than the simple addition of two comment sources.

Perceptions are important in this design because students and teachers determine whether feedback is trusted, ignored, questioned, or incorporated into revision. Students commonly value GenAI for availability and immediacy but may also perceive its responses as generic or in need of verification (Barrett & Pack, 2023; Chan & Hu, 2023; Teng, 2024). Teachers often recognize possibilities for efficiency and diversification while expressing concerns about quality, academic integrity, dependence, equity, and institutional policy (Chan, 2023; Cotton et al., 2024; Crawford ***et al***., 2023; Farazouli ***et al***., 2023; Moorhouse ***et al***., 2023; Perkins, 2023).

In Ecuadorian secondary education, published evidence on the interaction between teacher feedback and GenAI-supported feedback in EFL academic writing remains limited. National and regional discussions increasingly emphasize faculty digital competence, responsible AI adoption, and institutional support (Escobedo Nicot ***et al***., 2025; Mendoza Vergara & Macias Sera, 2026; Ossa & Willatt, 2023). Examining how students and teachers differentiate the strengths of each source can therefore provide a practical basis for designing feedback sequences that use technology without displacing pedagogical responsibility.

Accordingly, this study reframes the original comparison around complementarity. It examines how Ecuadorian

high school students from educational institutions in Portoviejo, Manabí evaluate teacher and AI-generated feedback, which attributes they associate with each source, how they perceive changes in their own writing and future use, and how teachers assess AI literacy, feedback practices, effectiveness, ethics, institutional readiness, and behavioural intention. It also explores whether response patterns vary by year of high school, gender, AI-use frequency, previous AI training, or teaching experience. Because the dataset contains perceptions rather than pre- and post-writing samples, the study does not claim causal improvement in writing performance.

Methodology

Design

The study used a quantitative cross-sectional design with a descriptive-comparative analytical orientation. Rather than testing an instructional intervention, the analysis examined how participants evaluated two feedback sources and how those evaluations were distributed across relevant background characteristics. No condition was experimentally manipulated, participants were not randomly assigned, and no pretest-posttest writing measure was available. Accordingly, comparisons are interpreted as perceptual evidence and not as estimates of causal effects on writing achievement.

Context and Participants

The study involved students and teachers from the Ecuadorian educational context. The student group was composed of learners from secondary education, while the teacher group included high school teachers from three educational units located in Portoviejo. The participants provided information related to their perceptions of teacher feedback and artificial intelligence-generated feedback in academic writing. The analysis considered relevant demographic and professional characteristics of the participants while maintaining the original research structure.

Instruments

Two project-specific questionnaires were used. The student instrument included 48 five-point Likert statements distributed across teacher feedback quality (TFQ, 10 items), AI-generated feedback quality (AIFQ, 10), comparative effectiveness (CE, 10), self-reported academic writing performance (AWP, 10), and satisfaction and behavioural intention (SBI, 8). The content addressed clarity, usefulness, personalization, speed, reliability, organization, confidence, accessibility, and intention to continue using AI-supported feedback.

The teacher instrument contained 54 five-point Likert statements covering AI literacy (AIL, 10 items), teacher feedback practices (TFP, 10), perceived effectiveness of AI (PEAI, 10), ethical concerns (EC, 8), institutional readiness (IR, 8), and behavioural intention (BI, 8). Demographic questions recorded gender, teaching experience, and previous AI training. Because the questionnaires were developed for the project rather than adopted as established psychometric scales, the analysis treats their thematic blocks cautiously.

Data Analysis

Data quality checks identified out-of-range Likert values and incomplete records before analysis. Categorical variables were summarized with frequencies and percentages. For questionnaire responses, item means, standard deviations, and the proportion of ratings of 4 or 5 were calculated. Thematic block averages and 95% confidence intervals were retained as descriptive summaries to facilitate comparison of response patterns.

Within the student sample, the overall TFQ and AIFQ summaries were compared with a paired-samples t test and checked with the Wilcoxon signed-rank test; standardized paired effect size was expressed as dz. Year of high school and gender comparisons used Mann-Whitney U tests, whereas AI-use frequency was explored with the Kruskal-Wallis test. Teacher comparisons by gender and previous AI training used Mann-Whitney U tests, and associations with years of experience were examined using Spearman correlations. Holm adjustment was applied to families of teacher comparisons, with $\alpha = .05$.

Cronbach's alpha was calculated as a diagnostic for each thematic block. Several coefficients were low or negative, indicating that the item groupings should not be interpreted as validated unidimensional scales. For that reason, item-level evidence is prioritized throughout the manuscript. Block means are used only to describe broad response tendencies; they are not treated as latent constructs and were not used for factor analysis, regression, or structural modelling.

Ethical Considerations

The dataset available for analysis was anonymized and contained no names, email addresses, or direct personal identifiers. Before journal submission, the manuscript should report the verified institutional ethics approval or exemption, the informed-consent procedure, confidentiality safeguards, and data-custody arrangements that applied to the original data collection.

Results and Discussion

Participant Characteristics

The student distribution was balanced by year of high school: 40 participants were in the second year and 40 in the third. A total of 52.5% identified as men and 47.5% as women. AI use for learning English was frequent: 33.8% reported using it often and 20.0% always; 36.3% responded sometimes and 10.0% rarely. Among teachers, 56.5% were men and 43.5% women. Mean professional experience was 13.35 years (SD = 7.13; range = 5-22), and 56.5% reported prior AI training.

Table 1. Characteristics of Students and Teachers

Group/variable	Category or statistic	n	% or value
High school students	Second year of Bachillerato	40	50,0
High school students	Third year of Bachillerato	40	50,0
High school students	Men	42	52,5
High school students	Women	38	47,5
AI use	Rarely	8	10,0
AI use	Sometimes	29	36,3
AI use	Often	27	33,8
AI use	Always	16	20,0
Teachers	Men	3	60
Teachers	Women	2	40
Teachers	AI training: yes	3	60
Teachers	AI training: no	2	40
Teachers	Experience, mean (SD)	13,35	(7.13); range 5–22

Source: Nstudents = 80; Nteachers = 5. Percentages may differ slightly because of rounding

Student Perceptions

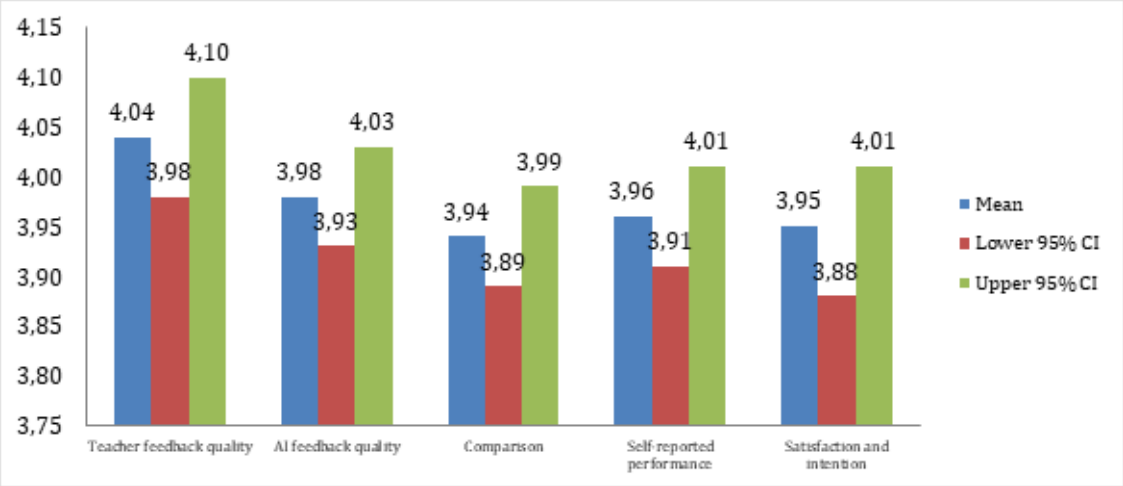
Average ratings were favorable across all blocks. Teacher feedback had the highest descriptive mean (M = 4.04; SD = 0.26), followed by AI-feedback quality (M = 3.98; SD = 0.24), self-reported performance (M = 3.96; SD = 0.23), satisfaction and intention (M = 3.95; SD = 0.28), and the comparative block (M = 3.94; SD = 0.22). The paired difference between TFQ and AIFQ was 0.055 points and was not significant, t (79) = 1.404, p = .164, dz = 0.157. The Wilcoxon test led to the same conclusion (W = 1011, p = .175). This suggests similar overall ratings but does not imply pedagogical equivalence between the sources.

Table 2. Descriptive Summaries of Student Responses

Block	Items	M	SD	95% CI	Diagnostic α
Teacher feedback quality (TFQ)	10	4,04	0,26	3.98–4.10	0,104
AI feedback quality (AIFQ)	10	3,98	0,24	3.93–4.03	-0,124
Comparison (CE)	10	3,94	0,22	3.89–3.99	-0,414
Self-reported performance (AWP)	10	3,96	0,23	3.91–4.01	-0,154
Satisfaction/intention (SBI)	8	3,95	0,28	3.88–4.01	-0,124

Source: Block means are descriptive summaries. The α coefficients do not support interpreting the blocks as unidimensional scales

Figure 1. Student Perceptions by Thematic Block



Source: Authors' elaboration based on the analyzed dataset

Item-level analysis revealed nuances not visible in the global comparison. Within TFQ, the highest ratings concerned motivation to revise, vocabulary improvement, and constructive suggestions (M ranging from 4.09 to 4.13). Within AIFQ, the highest-rated items were improvement in essay quality (M = 4.13), overall benefit (M = 4.10), and error identification (M = 4.09). The lowest means in this block concerned coherence improvement (M = 3.85) and ease of understanding the explanations (M = 3.88), although both remained above the scale midpoint.

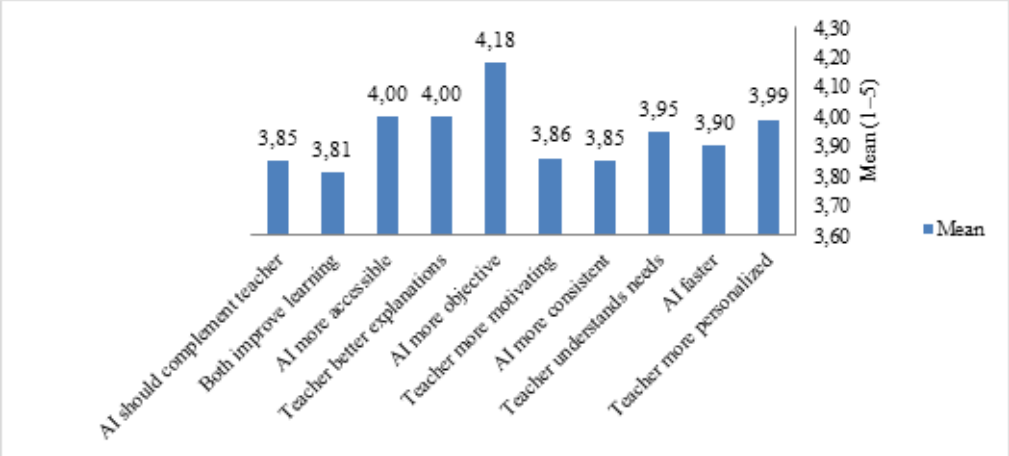
The comparative block revealed a functional distribution of strengths. The statement receiving the strongest agreement was that AI offers more objective feedback (M = 4.18; 83.8% responses of 4 or 5). Students also recognized its accessibility (M = 4.00; 75.0%) and speed (M = 3.90; 70.0%). For teacher feedback, students highlighted better explanations (M = 4.00; 75.0%), greater personalization (M = 3.99; 67.5%), and better understanding of their needs (M = 3.95; 73.8%). A total of 68.8% agreed that combining both sources improve learning, and approximately the same percentage believed that AI should complement rather than replace the teacher.

Table 3. Student Comparison of Feedback Attributes

Summarized statement	M	SD	Agreement %
Teacher feedback is more motivating	3,86	0,76	68,8
Using both sources improve learning	3,81	0,75	68,8
AI feedback is faster	3,90	0,74	70,0
The teacher provides better explanations	4,00	0,89	75,0
AI feedback is more consistent	3,85	0,83	67,5
AI should complement the teacher	3,85	0,78	68,8
Teacher feedback is more personalized	3,99	0,86	67,5
AI feedback is more accessible	4,00	0,84	75,0
The teacher better understands my needs	3,95	0,87	73,8
AI feedback is more objective	4,18	0,76	83,8

Source: Agreement represents responses of 4 or 5 on a five-point scale

Figure 2. Comparative Evaluation of Feedback Sources



Source: Authors' elaboration based on the analyzed dataset

Performance responses were self-reported and did not come from a writing test. Students perceived the greatest improvement in academic style (M = 4.13; 76.2% agreement), writing confidence (M = 4.08; 77.5%), sentence structure (M = 4.04; 78.8%), and paragraph organization (M = 3.99; 73.8%). Grammar improvement had the lowest mean (M = 3.78), although 70.0% agreed or strongly agreed. For satisfaction and intention, the highest item indicated that AI should complement traditional feedback (M = 4.03), followed by satisfaction with AI and intention to continue using it (both M = 4.01). The recommendation that teachers receive AI training reached M = 3.81 and 65.0% agreement.

Table 4. Selected Indicators of Self-Reported Performance, Satisfaction, and Intention

Indicator	M	SD	Agreement %
Sentence-structure improvement	4,04	0,79	78,8
Greater confidence in writing	4,08	0,82	77,5
Grammar improvement	3,78	0,69	70,0
Academic-style improvement	4,13	0,80	76,2
Satisfaction with AI feedback	4,01	0,82	75,0
Vocabulary improvement	3,83	0,79	68,8
Satisfaction with teacher feedback	3,99	0,85	73,8
Paragraph-organization improvement	3,99	0,85	73,8
Intention to continue using AI	4,01	0,88	75,0
Teachers should receive AI training	3,81	0,83	65,0

Source: Performance indicators represent perceptions rather than objective measurements of written texts.

Teacher Perceptions

Teachers reported high ratings for institutional readiness (M = 4.07; SD = 0.22), intention to use (M = 4.03; SD = 0.33), perceived AI effectiveness (M = 3.99; SD = 0.20), feedback practices (M = 3.94; SD = 0.19), and AI literacy (M = 3.92; SD = 0.22). Ethical concerns were close to the midpoint (M = 3.02; SD = 0.29), indicating caution without generalized rejection. The internal-consistency coefficients were inadequate; therefore, these values describe response groupings rather than psychometrically confirmed constructs.

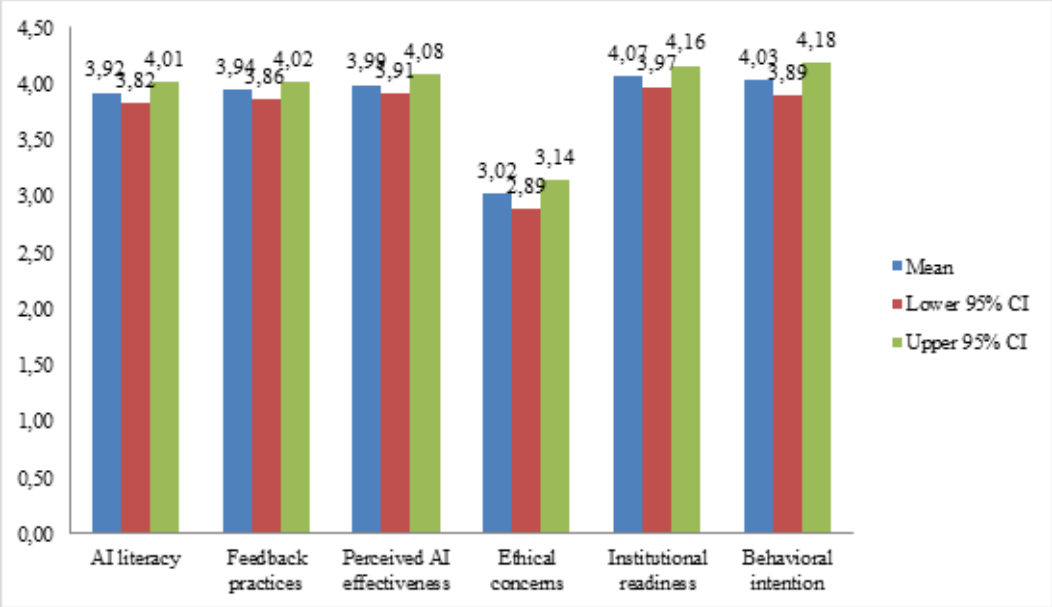
Table 5. Descriptive Summaries of Teacher Responses

Block	Items	M	SD	95% CI	Diagnostic α
AI literacy (AIL)	10	3,92	0,22	3.82–4.01	-0,117
Feedback practices (TFP)	10	3,94	0,19	3.86–4.02	-0,250
Perceived AI effectiveness (PEAI)	10	3,99	0,20	3.91–4.08	-0,542
Ethical concerns (EC)	8	3,02	0,29	2.89–3.14	0,332
Institutional readiness (IR)	8	4,07	0,22	3.97–4.16	-0,351

Behavioral intention (BI)	8	4,03	0,33	3.89–4.18	0,436

Source: The α coefficients prevent interpreting the blocks as validated unidimensional scales.

Figure 3. Teacher Perceptions by Thematic Block



Source: Authors' elaboration based on the analyzed dataset.

Group Comparisons

No statistically significant differences were observed between second- and third-year high school students across the five thematic summaries (p = .253-.907), by gender (p = .118-.965), or by AI-use frequency (p = .419-.907). Among teachers, the comparison by gender initially showed a difference in feedback practices (p = .032), but it was no longer significant after Holm adjustment (adjusted p = .192). No robust differences were found by prior AI training, and years of experience were not significantly correlated with the analyzed blocks. These results require caution, particularly because of the small teacher group and the limited internal consistency of the groupings.

Table 6. Summary of Exploratory Comparisons

Group	Factor	Test	Result
High school students	Level	Mann–Whitney U	No differences; p = .253–.907
High school students	Gender	Mann–Whitney U	No differences; p = .118–.965
High school students	AI-use frequency	Kruskal–Wallis	No differences; p = .419–.907
Teachers	Gender	U + Holm	TFP p = .032; adjusted p = .192
Teachers	AI training	U + Holm	No adjusted differences
Teachers	Experience	Spearman + Holm	No adjusted correlations

Source: The tests are exploratory and do not replace psychometric validation of the questionnaires

The findings are better understood as evidence of differentiated complementarity than as a competition between two feedback sources. Students evaluated both teacher feedback and AI-generated feedback positively, and the small difference between their global descriptive means was not statistically significant. Item-level responses, however, showed that participants did not assign the same role to each source. AI was linked most clearly with objectivity, accessibility, and speed; teachers were associated with personalization, understanding learner needs, and stronger explanations. Similar global ratings therefore conceal different perceived functions (Guo & Wang, 2024; Steiss ***et al***., 2024; Venter ***et al***., 2025).

The high rating for AI objectivity deserves careful interpretation. A generative system can respond repeatedly without fatigue and can appear consistent when applying the same prompt. Yet procedural consistency is not equivalent to correctness, contextual sensitivity, or fairness. Barrot (2023) documents limitations in ChatGPT-supported second-language writing, while Yan ***et al***. (2023) identify reliability, bias, and accountability as persistent educational concerns. Perceived objectivity should therefore be accompanied by verification rather than treated as a reason for automatic acceptance.

Teacher feedback was valued for qualities that depend on situated pedagogical knowledge. Teachers know

the learning objectives, previous student difficulties, expected proficiency, and communicative purpose of an assignment. This allows them to decide which problem deserves attention and whether the best response is correction, questioning, modelling, or guided reflection. Steiss ***et al***. (2024) similarly found advantages for human feedback in accuracy and prioritization, and Zhang and Hyland (2018) showed that learners' engagement is shaped by the feedback source and the relationship established around it.

The student responses consequently support a hybrid rather than substitutive model. Approximately seven in ten participants agreed that combining both sources could improve learning and that AI should complement the teacher. This pattern is consistent with studies in which ChatGPT-supported feedback produced stronger results when integrated with teacher mediation (Asadi ***et al***., 2025; Zhang ***et al***., 2025). Lo ***et al***. (2025) also reported benefits of hybrid arrangements for motivation and feedback quality. The practical implication is not to maximize the amount of feedback, but to assign different functions to different sources.

Complementarity requires coordination. If AI and a teacher offer conflicting recommendations, students need criteria for deciding which suggestion to accept, modify, or reject. If both sources agree, the convergence may increase confidence but should still be examined against the task rubric and the writer's intention. Venter et al. (2025) emphasize the importance of matching feedback source to purpose and timing. In a structured sequence, AI can support frequent low-stakes linguistic revision, while teachers can concentrate on argumentation, evidence, disciplinary conventions, voice, and recurring learning needs.

This perspective also clarifies the non-significant difference between TFQ and AIFQ. Similar averages should not be interpreted as evidence that AI and teachers are pedagogically equivalent. The global evaluation of AI may be strengthened by speed and permanent availability, while teacher feedback may be valued for depth and contextualization. Research comparing AI, instructors, and peers shows that conclusions change depending on whether the outcome is accuracy, grading, tone, formative usefulness, or actionability (Lu et al., 2024; Steiss ***et al***., 2024; Usher, 2025).

For EFL writing instruction, a staged feedback cycle appears more defensible than unrestricted use. Students might first use AI to identify possible grammar, vocabulary, clarity, or phrasing issues and then compare those suggestions with explicit criteria. Teacher feedback can subsequently address higher-order concerns and patterns that require knowledge of the learner and course. Reviews of automated writing evaluation show that feedback is more effective when embedded in sustained revision activity rather than delivered as isolated correction (Fu ***et al.***, 2024; Ngo et al., 2024). Feedback literacy is therefore a necessary condition for hybrid practice (Carless & Boud, 2018; Zhan & Yan, 2025; Zou ***et al***., 2025).

Students also perceived improvement in academic style, confidence, sentence structure, and organization. These results are compatible with research reporting positive experiences with automated or GenAI-supported revision in EFL contexts (Ajabshir & Ebadi, 2023; Guo ***et al***., 2024; Mekheimer, 2025). Teng (2024) likewise describes ChatGPT as a readily available writing companion from the learner perspective. Nevertheless, the present dataset contains perceptions rather than assessed texts. It cannot establish that writing performance objectively improved or attribute improvement to either feedback source.

This distinction matters because self-reported progress and measured writing development are different outcomes. Lu et al. (2024) found that ChatGPT can complement teacher assessment, but its usefulness varies by criterion and depends on oversight. Usher (2025) likewise found differences among AI, instructor, and peer assessment. Fleckenstein ***et al***. (2023) reported positive but context-dependent effects of automated feedback. A stronger next phase would therefore combine perception data with writing samples, analytic rubrics, independent raters, and pre-post comparisons.

Teacher responses provide a second layer to the complementarity argument. Participants reported favourable perceptions of AI effectiveness, institutional readiness, and future use, suggesting openness to incorporation of GenAI into educational practice. Openness, however, is not equivalent to professional competence. AI literacy includes understanding system limitations, designing appropriate prompts, checking outputs, protecting sensitive information, and deciding when AI should not be used (Long & Magerko, 2020; Ma et al., 2024; Ng ***et al***., 2021).

Moderate ethical concerns alongside high perceived readiness indicate that institutional support must address more than access to tools. Universities need guidance on privacy, intellectual property, transparency, acceptable assistance, authorship, and responsibility for assessment decisions. Chan (2023) proposes integrating governance, professional learning, and instructional redesign, while Moorhouse ***et al***. (2023) document the variety of institutional responses to generative AI. Crawford ***et al***. (2023) and UNESCO (2023) likewise emphasize human responsibility and clear ethical boundaries.

Policies for EFL academic writing should distinguish among linguistic assistance, editing, translation, idea

generation, content generation, and substitution of student work. These practices involve different educational and integrity risks. Faculty development should therefore allow teachers to test AI on authentic writing, compare its recommendations with course criteria, and design tasks that make the revision process visible. Without this pedagogical layer, favourable intentions may lead to inconsistent practices across courses and instructors.

The linguistic fluency of GenAI outputs can itself create a risk. Well-formed English may conceal generic argumentation, inaccurate information, fabricated references, or revisions that no longer represent the learner's proficiency or authorial voice. Critical AI literacy should help students question what the system privileges, why a suggestion appears plausible, and whether the proposed language serves the intended meaning. Walter (2024) links AI literacy with critical thinking and prompt design, and Tacelosky et al. (2025) show the relevance of critical engagement in language education.

The absence of robust differences by year of high school, gender, AI-use frequency, teacher training, or teaching experience suggests that favourable tendencies were relatively widespread in this dataset. These null findings should not be interpreted as evidence that background characteristics are irrelevant. The teacher sample is small, and the thematic blocks did not show sufficient internal consistency for strong subgroup inference. Future studies should examine whether proficiency, discipline, verified AI literacy, task type, and experience with specific systems explain variation more effectively.

The reliability diagnostics are therefore central to interpreting the study. Several low or negative alpha coefficients show that the project-developed blocks combine heterogeneous content and cannot be treated as established psychometric constructs. This is particularly understandable in the comparative and satisfaction/intention sections, where items intentionally address different sources and purposes. Prioritizing item-level findings avoids giving the descriptive averages a level of measurement precision that the available evidence does not support.

Taken together, the results suggest four requirements for a responsible human–AI feedback model: AI comments should be provisional and verifiable; teachers should retain responsibility for criteria and contextual judgement; students should be required to explain revision decisions; and institutions should provide training and rules concerning privacy, transparency, and authorship. This approach is consistent with human-centred perspectives in which AI extends educational agency without displacing human responsibility (Chiu, 2024; Xia ***et al***., 2024).

For Ecuadorian higher education, the study provides an exploratory diagnosis of conditions that may support hybrid feedback design. Student and teacher openness creates an opportunity for experimentation, but adoption should be accompanied by stronger measurement and direct evidence of learning. Regional work on digital transformation similarly indicates that access to technology must be accompanied by professional competence and institutional support (Escobedo Nicot ***et al***., 2025). In English teaching, the relevant agenda includes feedback quality, learner autonomy, ethical use, and observable writing development (Mendoza Vergara & Macias Sera, 2026).

Implications for Practice

Structure feedback as a sequence rather than as simultaneous comment accumulation. AI can be used during an early revision stage for grammar, vocabulary, clarity, and alternative phrasing, while teacher feedback can concentrate on argumentation, organization, evidence, academic voice, and disciplinary appropriateness.

Make feedback uptake visible. Students should compare AI suggestions with the rubric, explain why they accepted or rejected particular recommendations, and retain a record of substantive revisions. This preserves student decision-making and provides teachers with evidence of how feedback was used.

Develop faculty capacity around pedagogical use, not only tool operation. Professional learning should integrate AI literacy, prompt design, output verification, formative assessment, privacy, academic integrity, and adaptation to different levels of English proficiency.

Clarify institutional expectations. Universities should specify permitted and restricted uses, disclosure requirements, data-protection practices, and responsibility for final assessment decisions. Guidance should distinguish language support from substitutive content generation.

Evaluate future implementations with direct evidence. Perception questionnaires should be complemented with authentic writing tasks, pre-post measures, analytic rubrics, independent raters, and analysis of which teacher- and AI-generated comments students actually use.

Limitations and Future Research

The study is limited by its cross-sectional and self-reported nature. The responses represent perceptions at one point in time and may be influenced by enthusiasm or concern surrounding a rapidly changing technology.

The dataset contains no written texts, pretest-posttest measures, analytic rubrics, or independent ratings; consequently, statements about performance refer only to participants' perceptions.

A second limitation concerns measurement. Both questionnaires were developed for the project, and several thematic blocks produced weak internal-consistency coefficients. The results should therefore be read descriptively and primarily at item level. The small teacher sample further limits subgroup comparisons, while the available dataset does not identify the specific AI tool, task type, duration of use, institution, or sampling procedure needed for more detailed contextual explanation.

Future research should refine the instruments through expert review, cognitive interviews, and pilot testing before evaluating their internal structure with an independent sample. A subsequent effectiveness study should compare clearly defined teacher-only, AI-supported, and hybrid feedback conditions using authentic academic writing tasks, pre- and post-intervention measures, analytic rubrics, at least two blinded raters, and inter-rater agreement. Qualitative analysis of accepted, modified, and rejected comments would also clarify how students exercise judgement during revision.

Conclusions

The findings show that Ecuadorian high school students from educational institutions in Portoviejo, Manabí perceive both teacher and AI-generated feedback as useful resources for EFL academic writing, but they value them for different reasons. AI was associated mainly with objectivity, speed, and accessibility, whereas teacher feedback was associated with personalization, understanding of learner needs, and explanation quality. The absence of a significant difference between the two global ratings should therefore be interpreted as similarity in overall acceptance, not as evidence that the two sources are interchangeable.

Teacher responses reinforce the case for complementarity. Participants expressed favorable views of AI effectiveness, institutional readiness, and future use while maintaining moderate ethical concerns. A responsible model should use AI to expand opportunities for timely revision while preserving teacher judgment, student authorship, verification practices, and institutional safeguards concerning transparency, privacy, and academic integrity.

The contribution of the study lies in identifying how students and teachers distribute perceived strengths across human and automated feedback and in translating those perceptions into conditions for hybrid design. Because the questionnaires require further validation and the study did not directly measure writing outcomes, the results remain exploratory. Future research should test whether carefully sequenced human–AI feedback improves observable writing quality and feedback literacy beyond what either source achieves alone.

References

Ajabshir, Z. F., & Ebadi, S. (2023). The effects of automatic writing evaluation and teacher-focused feedback on CALF measures and overall quality of L2 writing across different genres. Asian-Pacific Journal of Second and Foreign Language Education, 8, 26. https://doi.org/10.1186/s40862-023-00201-9

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. Thinking Skills and Creativity, 56, 101766. https://doi.org/10.1016/j.tsc.2025.101766

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? International Journal of Educational Technology in Higher Education, 21, 23. https://doi.org/10.1186/s41239-024-00455-4

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, 59. https://doi.org/10.1186/s41239-023-00427-0

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, 100745. https://doi.org/10.1016/j.asw.2023.100745

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74. https://doi.org/10.1080/0969595980050102

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

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education, 20, 38. https://doi.org/10.1186/s41239-023-00408-3

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher

education. International Journal of Educational Technology in Higher Education, 20, 43. https://doi.org/10.1186/s41239-023-00411-8

Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative AI. Computers and Education: Artificial Intelligence, 6, 100197. https://doi.org/10.1016/j.caeai.2023.100197

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). Journal of University Teaching & Learning Practice, 20(3), Article 2. https://doi.org/10.53761/1.20.3.02

Escobedo Nicot, M., Rondón Bolúa, Y., & Cano Ortíz, S. D. (2025). Desarrollo de competencias digitales en docentes universitarios: una iniciativa de transformación digital educativa en la Universidad de Oriente. Maestro y Sociedad, 22(1), 312–319. https://maestroysociedad.uo.edu.cu/index.php/MyS/article/view/6781

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2023). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots' impact on university teachers' assessment practices. Assessment & Evaluation in Higher Education. Advance online publication. https://doi.org/10.1080/02602938.2023.2241676

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. Frontiers in Artificial Intelligence, 6, 1162454. https://doi.org/10.3389/frai.2023.1162454

Fu, Q.-K., Zou, D., Xie, H., & Cheng, G. (2024). A review of AWE feedback: Types, learning outcomes, and implications. Computer Assisted Language Learning, 37(1–2), 179–221. https://doi.org/10.1080/09588221.2022.2033787

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. Education and Information Technologies, 29, 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. The Internet and Higher Education, 63, 100962. https://doi.org/10.1016/j.iheduc.2024.100962

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

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537–550. https://doi.org/10.1177/00336882231162868

Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation feedback in ESL writing instruction. Journal of Second Language Writing, 27, 1–18. https://doi.org/10.1016/j.jslw.2014.10.004

Lo, N., Chan, S., & Wong, A. (2025). Evaluating teacher, AI, and hybrid feedback in English language learning: Impact on student motivation, quality, and performance in Hong Kong. SAGE Open, 15(3). https://doi.org/10.1177/21582440251352907

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). ACM. https://doi.org/10.1145/3313831.3376727

Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., & Zhu, X. (2024). Can Chat GPT effectively complement teacher assessment of undergraduate students' academic writing? Assessment & Evaluation in Higher Education, 49(5), 616–633. https://doi.org/10.1080/02602938.2024.2301722

Ma, Q., Crosthwaite, P., Sun, D., & Zou, D. (2024). Exploring ChatGPT literacy in language education: A global perspective and comprehensive approach. Computers and Education: Artificial Intelligence, 7, 100278. https://doi.org/10.1016/j.caeai.2024.100278

Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. Discover Education, 4, 170. https://doi.org/10.1007/s44217-025-00602-7

Mendoza Vergara, C., & Macias Sera, R. (2026). Inteligencia artificial en la enseñanza del inglés. Maestro y Sociedad, 23(1), 331–339. https://maestroysociedad.uo.edu.cu/index.php/MyS/article/view/7420

Moorhouse, B. L., Yeo, M. A., & Wan, Y. W. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. Computers and Education Open, 5, 100151. https://doi.org/10.1016/j.caeo.2023.100151

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Ngo, T. T. N., Chen, H. H. J., & Lai, K. K. W. (2024). The effectiveness of automated writing evaluation in EFL/ESL writing: A three-level meta-analysis. Interactive Learning Environments, 32(2), 727–744. https://doi.org/10.1080/10494820.2022.2096642

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199–218. https://doi.org/10.1080/03075070600572090

Ossa, C., & Willatt, C. (2023). Providing academic writing feedback assisted by generative artificial intelligence in initial teacher education contexts. European Journal of Education and Psychology, 16(2), 1–16. https://doi.org/10.32457/ejep.v16i2.2412

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching & Learning Practice, 20(2), Article 7. https://doi.org/10.53761/1.20.02.07

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. Learning and Instruction, 91, 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. Assessing Writing, 19, 51–65. https://doi.org/10.1016/j.asw.2013.11.007

Tacelosky, K., Kasun, G. S., Shapiro, B. R., Liao, Y.-C., & Harris, K. (2025). Exploring critical AI literacy in language education: A case study. Foreign Language Annals, 58(4), 966–994. https://doi.org/10.1111/flan.70029

Teng, M. F. (2024). ChatGPT is the companion, not enemies: EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. Computers and Education: Artificial Intelligence, 7, 100270. https://doi.org/10.1016/j.caeai.2024.100270

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, 15. https://doi.org/10.1186/s40561-023-00237-x

UNESCO. (2023). Guidance for generative AI in education and research.

Usher, M. (2025). Generative AI vs. instructor vs. peer assessments: A comparison of grading and feedback in higher education. Assessment & Evaluation in Higher Education, 50(6), 912–927. https://doi.org/10.1080/02602938.2025.2487495

Venter, J., Coetzee, S. A., & Schmulian, A. (2025). Exploring the use of artificial intelligence (AI) in the delivery of effective feedback. Assessment & Evaluation in Higher Education, 50(4), 516–536. https://doi.org/10.1080/02602938.2024.2415649

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. International Journal of Educational Technology in Higher Education, 21, 15. https://doi.org/10.1186/s41239-024-00448-3

Xia, Q., Weng, X., Ouyang, F., Lin, T.-J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. International Journal of Educational Technology in Higher Education, 21, 40. https://doi.org/10.1186/s41239-024-00468-z

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. British Journal of Educational Technology, 54(5), 1156–1181. https://doi.org/10.1111/bjet.13370

Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. Innovation in Language Learning and Teaching. Advance online publication. https://doi.org/10.1080/17501229.2025.2503890

Zhang, Z., & 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

Zhan, Y., & Yan, Z. (2025). Students' engagement with ChatGPT feedback: Implications for student feedback literacy in the context of generative artificial intelligence. Assessment & Evaluation in Higher Education. Advance online publication. https://doi.org/10.1080/02602938.2025.2471821

Zou, S., Guo, K., Wang, J., & Liu, Y. (2025). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. Computer Assisted Language Learning. Advance online publication. https://doi.org/10.1080/09588221.2024.2447279

Conflict of interest: The authors declare that they have no conflicts of interest.

Declaration of responsibility of authorship:

We, Edisson Javier Vélez Sánchez, Jostin Javier Fernández Macías, Pamela Nicolle Cedeño Loor, Steeven Josue Valencia Cedeño, and Miguel Macías Loor, authors of the indicated manuscript, DECLARE that we have contributed directly to its intellectual content, as well as to the genesis and analysis of its data; therefore, we are in a position to be made publicly responsible for it and accept that our names appear in the list of authors in the indicated order. We also declare that the ethical requirements of the aforementioned publication have been met, having consulted the Declaration of Ethics and Malpractice in Publication.

Edisson Javier Vélez Sánchez, Jostin Javier Fernández Macías, Pamela Nicolle Cedeño Loor, Steeven Josue Valencia Cedeño, and Miguel Macías Loor: research, data collection, data interpretation and analysis, manuscript drafting, preparation of the abstract, and preparation of the conclusions.