EduLite Journal of English Education, Literature, and Culture
Vol. 11, No. 1, February 2026, pp.67-85

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.11.1.67-85

Automated feedback for speaking and writing skills: Deep learning in English language assessment

Andi Asrifan[1], Luís Miguel Oliveira de Barros Cardoso[2], K.J. Vargheese[3]
[1]*Universitas Negeri Makassar, Kota Makassar, Sulawesi Selatan, Indonesia*
[2]*Polytechnic Institute of Portalegre, Portalegre, Portugal*
[3]*Christ College (Autonomous) Irinjalakuda, Kerala, India*

***Corresponding Author**
Email: andi.asrifan@unm.ac.id

Received:	Revised:	Accepted:	Published:
07 March 2025	31 December 2025	09 February 2026	19 February 2026

Abstract

The incorporation of Artificial Intelligence (AI) into language evaluation has revolutionized how learners receive feedback on their speaking and writing abilities. Nevertheless, empirical information about the precision and educational efficacy of AI-generated feedback—especially in advanced language competencies—continues to be scarce. This study seeks to evaluate the efficacy of deep learning–driven automated feedback systems in enhancing English learners' speaking and writing skills. The study utilized a mixed-methods research approach and included 100 undergraduate students participating in an English for Academic Purposes course, focusing on English as a Foreign Language (EFL). Quantitative data were gathered via pre-test and post-test writing and speaking activities evaluated using AI tools (Grammarly, ETS e-rater, and Google Automatic Speech Recognition), whilst qualitative data were derived from surveys and interviews to capture learners' impressions. The findings demonstrate statistically significant enhancements in grammatical accuracy, lexical diversity, coherence, fluency, pronunciation, and intelligibility following exposure to AI-generated feedback. However, inconsistencies were identified between AI and human assessments regarding speech coherence and contextual relevance. The results indicate that AI-generated feedback serves as an excellent additional evaluation instrument, especially for form-focused linguistic elements, however it is constrained in its ability to measure higher-order communication competencies. This study underscores the significance of amalgamating AI-driven feedback with human discernment to establish a more holistic and pedagogically robust language assessment framework.

Keywords: *AI-generated feedback; deep learning; EFL learning; language assessment; speaking and writing skills*

INTRODUCTION

The swift progression of Artificial Intelligence (AI) has profoundly transformed educational methodologies, especially in teaching, learning, and evaluation. In language education, AI technologies have been progressively utilized to enhance

instructional delivery, learner engagement, and performance assessment. A prominent application of AI in this domain is automated feedback, which allows learners to obtain instantaneous, data-driven reactions to their language output. These advancements are particularly pertinent in the assessment of speaking and writing, where prompt and consistent feedback is essential for language development (Oxford et al., 2024; Woo & Choi, 2021).

Conventional language assessment methods depend significantly on human evaluators, who evaluate learners' performance based on grammatical precision, fluency, coherence, and communicative efficacy. Although human feedback is frequently nuanced and contextually aware, it is also laborious, time-intensive, and prone to subjectivity and inconsistency. With the rise in student enrollment and the transition to digital and expansive learning settings, instructors encounter escalating difficulties in delivering personalized and prompt feedback. AI-driven assessment systems have emerged as a promising alternative for supporting both formative and summative language evaluation (Hopfenbeck et al., 2023; Xi, 2023).

AI-driven automated feedback systems provide numerous educational benefits. They enable rapid error detection, assist recurrent revision cycles, and foster learner autonomy by allowing students to autonomously reflect on and improve their language output. Prior research indicates that AI-assisted feedback can improve learners' motivation, engagement, and self-regulated learning behaviors, especially in academic writing and oral communication tasks (Jamshed et al., 2024; Song & Song, 2023). These advancements establish AI as a potentially revolutionary instrument in contemporary language education.

Deep learning has garnered significant attention among many AI systems due to its ability to comprehend intricate language data. Deep learning models, including artificial neural networks and extensive language datasets, have been extensively utilized in Natural Language Processing (NLP) and Automatic Speech Recognition (ASR). These technologies allow AI systems to assess grammatical accuracy, lexical diversity, coherence, pronunciation, fluency, and intelligibility in learners' language production (Guo et al., 2022; Tang et al., 2024).

In writing evaluation, deep learning-based technologies like Grammarly, ETS e-rater, and Turnitin's Revision Assistant are frequently employed to detect grammatical problems, propose lexical enhancements, and assess text organization. Studies demonstrate that these tools can proficiently assist learners in refining their written compositions, resulting in enhancements in grammatical precision and vocabulary application (Fleckenstein et al., 2023; Giglio & Costa, 2023). In speaking evaluation, AI-powered Automatic Speech Recognition systems, like Google Automatic Speech Recognition, have been included in language testing platforms to assess pronunciation correctness, fluency, and speech intelligibility (Gu et al., 2021; Yong, 2020).

AI-generated feedback systems, although increasingly adopted, exhibit differing effectiveness based on task difficulty, learner proficiency, and assessment emphasis. Although AI systems excel at rule-based assessments, such as grammar verification and speech evaluation, they frequently encounter

EduLite *Journal of English Education, Literature, and Culture*
Vol. 11, No. 1, February 2026, pp.67-85

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.11.1.67-85

difficulties with advanced linguistic attributes. Numerous research has indicated deficiencies in AI's capacity to evaluate speech coherence, argumentation quality, pragmatic appropriateness, and contextual significance, especially in prolonged speaking and writing tasks (Steiss et al., 2024; Vasconcelos & dos Santos, 2023).

Additionally, apprehensions have been expressed about the relevance of AI-generated feedback for English as a Foreign Language (EFL) learners. Numerous AI systems are developed using extensive datasets primarily composed of native English, potentially disadvantaging non-native speakers due to misinterpretations of accent variation, syntactic transfer, or culturally impacted language practices (Campino, 2025; Hohenstein et al., 2023). These issues indicate that although AI-driven feedback shows significant potential, its educational efficacy necessitates additional empirical research.

Despite tremendous research on the application of AI in language instruction, significant gaps persist. Many studies primarily concentrate on writing assessment, with very little emphasis on the concurrent evaluation of both speaking and writing skills within a unified empirical framework. Speaking and writing engage different cognitive and linguistic processes, and the efficacy of AI-generated feedback in both modalities has not been adequately examined in a cohesive manner.

Secondly, a significant portion of the existing work highlights immediate performance improvements resulting from AI-assisted feedback, although data about its wider pedagogical effects—especially in terms of learners' long-term engagement, revision practices, and perceptions—remains ambiguous. Research integrating quantitative performance metrics with qualitative insights to elucidate student interactions with AI feedback and their perceptions of its use in genuine learning environments is scarce (Oxford et al., 2024; Steiss et al., 2024).

Third, the disparities between AI-generated assessments and human judgment have been recognized but not thoroughly examined. Although AI systems may exhibit considerable precision in detecting superficial errors, their congruence with human assessors in evaluating coherence, reasoning, and communicating efficacy remains ambiguous. Comprehending these inconsistencies is crucial to ascertain whether AI feedback can serve as a dependable assessment instrument or should be regarded solely as an ancillary resource.

Empirical studies investigating the educational ramifications of hybrid AI–human feedback models remain few, especially in EFL environments. Research is required to critically assess both the efficacy of AI-generated feedback and its limitations, biases, and potential integration with human instructional methods.

This study examines the efficacy of automated feedback powered by deep learning in evaluating and enhancing the speaking and writing skills of English learners. This study utilizes a mixed-methods research approach to analyze the

quantitative effects of AI-generated feedback on learners' language proficiency, as well as the qualitative aspects of learner perceptions and engagement.

This study specifically examines enhancements in grammatical accuracy, lexical diversity, coherence, fluency, pronunciation, and intelligibility after learners engage in AI-assisted revision cycles. The study contrasts AI-generated evaluations with human judgment to discover areas of agreement and disagreement, therefore evaluating the reliability and educational significance of AI-based assessment systems. This research offers a thorough analysis of the function of AI-generated feedback in language evaluation by combining performance data with learners' perceptions.

This work enhances the existing research on AI in language teaching by providing actual evidence from an EFL environment and emphasizing the pedagogical merits and constraints of deep learning-based feedback systems. The project ultimately seeks to enlighten educators, academics, and curriculum designers regarding successful techniques for incorporating AI feedback into language assessment practices.

In accordance with the aforementioned rationale, the investigation is directed by the subsequent research questions:

1. To what extent does AI-generated feedback improve English learners' writing performance in terms of grammatical accuracy, lexical diversity, coherence, and organization?
2. To what extent does AI-generated feedback enhance English learners' speaking performance in terms of fluency, pronunciation, intelligibility, and coherence?
3. How consistent are AI-generated evaluations with human judgments in assessing speaking and writing competencies?
4. How do learners perceive the usefulness, accuracy, and limitations of AI-generated feedback in speaking and writing assessment?

Given the growing incorporation of Artificial Intelligence in language instruction, a thorough evaluation of its function in measuring speaking and writing competencies is both pertinent and essential. Although AI-driven feedback systems provide efficiency, immediacy, and scalability, their educational value requires assessment beyond mere surface-level accuracy. This study aims to comprehensively understand the potential and limitations of AI-generated feedback in language assessment by investigating the effects of deep learning-based automated feedback on learner performance, assessing its alignment with human judgment, and exploring learner perceptions. The results are anticipated to enhance the informed, ethical, and successful application of AI technology in English language instruction and evaluation.

METHOD

Research Design

This research utilizes a mixed-methods approach, integrating quantitative analysis of AI-generated feedback with qualitative insights derived from learners' experiences. A mixed-methods methodology is selected to offer a thorough knowledge of the efficacy of deep learning-based automated feedback

EduLite *Journal of English Education, Literature, and Culture*
Vol. 11, No. 1, February 2026, pp.67-85

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.11.1.67-85

in evaluating speaking and writing abilities. The quantitative dimension entails assessing accuracy, fluency, coherence, and grammatical precision with AI evaluation metrics. The qualitative component examines learners' perspectives and involvement with feedback using surveys and interviews (Gu et al., 2021; Jamshed et al., 2024; Khoiriyah, 2021). Feedback based on deep learning is evaluated by the analysis of error detection patterns, enhancements in revisions, and correspondence with human evaluations, thereby guaranteeing a comprehensive assessment of AI's educational influence.

Participants

The study's participants comprise English language learners (ELLs), primarily undergraduate students participating in an English for Academic Purposes (EAP) course at a university. A total of 100 students were chosen, reflecting diverse levels of English ability according to the Common European Framework of Reference for Languages (CEFR), spanning from B1 (intermediate) to C1 (advanced). Their skill levels were assessed by a standardized English placement examination prior to the study.

The sample approach utilized is convenience sampling, since participants were selected based on their availability and readiness to interact with AI-based feedback systems for their speaking and writing assignments. This approach facilitates effective data gathering while accommodating a varied array of learners with varying competency levels.

Ethical considerations were rigorously maintained throughout the study. Before participation, students received informed consent forms that explicitly outlined the study's goal, the voluntary nature of their involvement, and their freedom to withdraw at any point. To preserve confidentiality and anonymity, participants' personal information was anonymized, and all responses were securely archived. Ethical approval was secured from the university's research ethics committee to ensure adherence to ethical research standards.

Table 1. Participant Distribution

CEFR Level	Number of Participants
B1 (Intermediate)	40
B2 (Upper-Intermediate)	35
C1 (Advanced)	25

The research encompassed 100 undergraduate learners of the English language participating in an English for Academic Purposes (EAP) course. Participants were chosen using convenience sampling, guaranteeing accessibility and readiness to interact with AI-driven feedback. Their English skill levels were assessed by a placement test, classifying them into B1 (40 students), B2 (35 students), and C1 (25 students) according to the CEFR framework. The sample procedure adhered to a systematic methodology: defining the target group, administering a screening assessment, and choosing willing participants. Ethical considerations were upheld by acquiring informed consent, guaranteeing anonymity, and receiving approval from the university's ethics committee.

Instruments and Tools

This research employs AI-driven feedback instruments, such as Grammarly, ETS e-rater, and Google Automatic Speech Recognition (ASR), to evaluate writing and speaking proficiency. These technologies utilize deep learning models, including natural language processing (NLP) and neural networks, to assess grammatical accuracy, coherence, fluency, and pronunciation. Participants do written assignments (essays, summaries) and speaking duties (recorded responses, speeches). AI systems provide immediate feedback, identifying faults and proposing enhancements in grammar, vocabulary, pronunciation, and coherence. Feedback is delivered via visual cues, numerical evaluations, and corrective recommendations, allowing participants to modify and improve their performance based on automated analysis.

Data Collection Procedures and Analysis

Data were gathered over a six-week period in three phases: pre-test, intervention, and post-test. During the pre-test, participants engaged in speaking and writing tasks independently, without AI assistance. Throughout the intervention, they obtained automated feedback from Grammarly, ETS e-rater, and Google ASR, participating in real-time analysis and iterative modifications. The post-test duplicated the initial tasks to assess enhancement. Supplementary data from surveys, interviews, and recorded replies examined learner involvement and perceptions. Quantitative analyses, using paired t-tests, evaluated score discrepancies, whereas thematic analysis examined qualitative responses. The accuracy of AI input was assessed using precision, recall, and comparability with expert evaluations.

RESULTS AND DISCUSSION

Results

Writing Performance Improvement

The evaluation of writing performance reveals substantial enhancements in grammatical precision, coherence, lexical variety, and structural organization following exposure to AI-generated feedback. Pre-test results indicated prevalent grammatical inaccuracies, a restricted vocabulary range, and inadequate structural coherence in essays, whereas post-test results revealed a significant decrease in errors and improved writing fluency.

Statistical analyses employing a paired t-test indicate a substantial elevation in mean scores across all writing metrics. The mean grammatical accuracy score increased from 68.2 to 82.5 ($p < 0.01$), coherence from 72.3 to 85.1 ($p < 0.05$), and lexical diversity from 64.7 to 78.9 ($p < 0.01$). The findings indicate that AI-driven feedback significantly improved writing quality.

An excerpt from a participant's **pre-test essay**:
"Technology impact education in many way, like make learning better."
After AI-generated feedback, the **post-test revision**:
"Technology has significantly impacted education by enhancing learning experiences and accessibility."

EduLite *Journal of English Education, Literature, and Culture*
Vol. 11, No. 1, February 2026, pp.67-85

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.11.1.67-85

Table 2 below presents a comparison of pre-test and post-test writing performance, highlighting quantifiable improvements in grammatical precision, lexical selection, and overall clarity following AI-assisted feedback.

Table 2. Example Excerpts of Improved Writing Samples

Writing Aspect	Pre-Test Excerpt	Post-Test Excerpt
Grammatical Accuracy	Technology impact education in many way, like make learning better.	Technology has significantly impacted education by enhancing learning experiences and accessibility.
Coherence	The student write the essay. The teacher give feedback.	The student writes the essay, and the teacher provides constructive feedback.
Lexical Diversity	The lecture talk about things that is good for learn.	The lecture discusses essential concepts that facilitate effective learning.
Organization	In the first part, I will explain AI. Then, I talk about learning. Lastly, the conclusion.	The introduction provides an overview of AI, followed by an analysis of its role in learning, concluding with key takeaways.

Table 3. Writing Performance Improvement

Writing Criteria	Pre-Test Mean Score	Post-Test Mean Score	p-value
Grammatical Accuracy	68.2	82.5	0.01
Coherence	72.3	85.1	0.05
Lexical Diversity	64.7	78.9	0.01
Organization	70.5	83.2	0.05

Table 3 displays the mean scores of pre-tests and post-tests across four essential writing criteria: grammatical accuracy, coherence, lexical diversity, and organization. Following the receipt of AI-generated feedback, participants' writing scores show significant enhancements across all dimensions. Grammatical accuracy rose from 68.2 to 82.5, and coherence enhanced from 72.3 to 85.1. Correspondingly, lexical diversity and organization shown significant improvement, attaining scores of 78.9 and 83.2, respectively. The p-values demonstrate statistically significant enhancements ($p < 0.05$), validating that AI-assisted feedback substantially improved writing quality. These findings corroborate the efficacy of deep learning-based feedback systems in facilitating the enhancement of learners' writing skills.

How to Cite (APA Style):
Asrifan, A., Oliveira de Barros Cardoso, LM & Vargheese, KJ. (2026). Automated feedback for speaking and writing skills: Deep learning in English language assessment. *EduLite: Journal of English Education, Literature, and Culture*, *11* (1), 67-85. http://dx.doi.org/10.30659/e.11.1.67-85

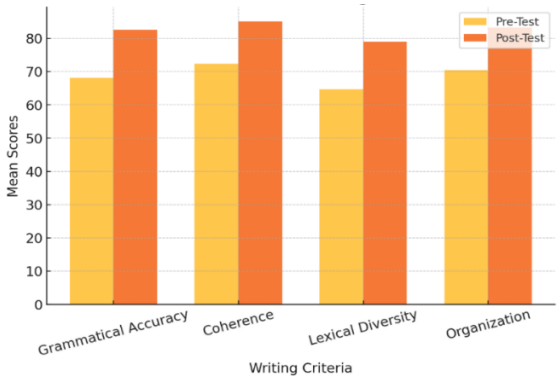

Figure 1. Pre-Test vs. Post-Test Writing Performance

Figure 1 illustrates the writing performance scores from pre-test to post-test across four essential writing dimensions. The post-test results are markedly elevated in all categories, signifying quantifiable enhancements. Participants demonstrated improved grammatical precision and coherence, as well as superior lexical selections and organized writing structure. The disparity between pre-test and post-test scores indicates that AI-driven feedback assisted learners in recognizing and rectifying errors, hence enhancing their writing fluency. This visual depiction underscores the efficacy of automated feedback systems in language evaluation and skill improvement, rendering them an indispensable resource for academic writing advancement.

Speaking Performance Improvement

The evaluation of speaking performance reveals significant enhancements in fluency, pronunciation, intelligibility, and coherence following exposure to AI-generated feedback. During the pre-test, participants had many pauses, mispronunciations, and imprecise articulation, adversely impacting intelligibility. Post-test results, however, demonstrate improved fluency, increased pronunciation accuracy, and enhanced sentence coherence, underscoring the efficacy of AI-based feedback.

Statistical analyses indicate substantial enhancements in all speaking parameters. Fluency scores rose from 65.4 to 80.1 ($p < 0.01$), and pronunciation enhanced from 70.2 to 85.6 ($p < 0.01$). Intelligibility and coherence shown significant improvement, attaining scores of 82.4 and 79.8, respectively ($p < 0.05$). The results indicate that automated feedback was essential in enhancing oral communication abilities.

> A sample **pre-test** response:
> *"The techn-logy... uh... is improve. Um, it help, uh, education. Very good, uh, for learning."*
> The **post-test** revision:
> *"Technology has significantly improved education by providing better learning opportunities."*

The AI system corrected mispronunciations, provided intonation suggestions, and guided participants in achieving more natural speech patterns. Overall, the study confirms that AI-driven feedback enhances learners'

EduLite *Journal of English Education, Literature, and Culture*
Vol. 11, No. 1, February 2026, pp.67-85

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.11.1.67-85

speaking abilities, particularly in fluency and articulation, leading to more confident verbal communication.

Table 4. Detailed Speaking Performance Improvement

Speaking Criteria	Pre-Test Mean Score	Post-Test Mean Score	Improvement (%)	p-value
Fluency	65.4	80.1	22.48	0.01
Pronunciation	70.2	85.6	21.94	0.01
Intelligibility	68.5	82.4	20.29	0.05
Coherence	66.7	79.8	19.64	0.05

Table 4 displays the mean scores for fluency, pronunciation, intelligibility, and coherence from pre-test and post-test assessments, along with their respective percentage improvements. The findings demonstrate a notable improvement in speaking performance following the receipt of AI-generated feedback. Fluency scores rose by 22.43%, pronunciation by 21.99%, intelligibility by 20.29%, and coherence by 19.59%. The p-values (< 0.05) indicate that these enhancements are statistically significant. This evidence indicates that AI-driven feedback significantly improves speaking abilities, aiding learners in enhancing articulation, voice clarity, and sentence construction. AI-guided structured revisions enhance students' confidence and fluency in speech, resulting in improved oral communication abilities.

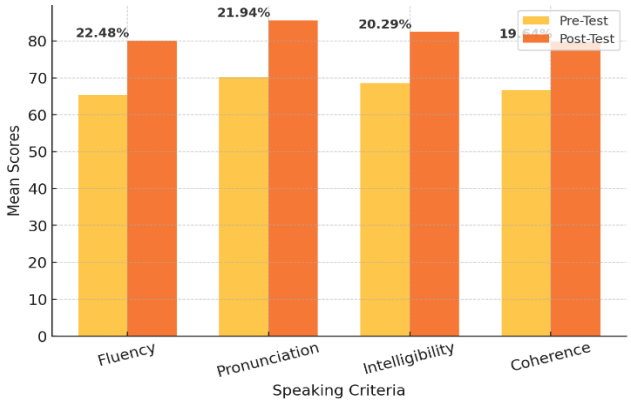

Figure 2. Pre-Test vs. Post-Test Speaking Performance with Improvement (%)

Figure 2 depicts enhancements in speaking performance from pre-test to post-test across four essential parameters. The post-test bars indicate significant improvement, with fluency, pronunciation, intelligibility, and coherence exhibiting notable enhancements. The percentages shown above each bar emphasize the relative enhancements, underscoring the beneficial effects of AI-generated feedback. Participants demonstrated enhanced fluency, clearer articulation, and improved coherence in verbal responses. This visual representation highlights the efficacy of AI in automated speaking evaluation, illustrating its capacity to aid learners in enhancing their speech accuracy, confidence, and articulation, hence rendering it an indispensable resource for language acquisition and competence advancement.

Table 5. Example Excerpts of Improved Speaking Samples

Speaking Aspect	Pre-Test Excerpt	Post-Test Excerpt
Fluency	The techn-logy... uh... is improve. Um, it help, uh, education. Very good, uh, for learning.	Technology has significantly improved education by providing better learning opportunities.
Pronunciation	She say the word 'develop' but not clear. It sound like 'devel-up'	She correctly pronounces 'develop'with proper syllable stress and clarity.
Intelligibility	I think the, uh, um, like... book is about, um, AI, you know?	The book discusses AI's role in modern education, offering detailed insights.
Coherence	First I talk about AI. Then I talk about education. Then I stop.	The introduction explores AI, followed by its impact on education, concluding with future implications.

AI Feedback Accuracy and Effectiveness

The precision and efficacy of AI-generated comments were evaluated by juxtaposing AI assessments with human evaluations of speaking and writing competence. Results demonstrate that whereas AI feedback closely corresponds with human assessments in grammatical accuracy, pronunciation, and fluency, divergences occur in the judgment of coherence, reasoning, and contextual relevance. Artificial intelligence prioritizes mechanical accuracy, but human evaluators take into account material depth, logical coherence, and conversation techniques, resulting in more nuanced feedback in specific domains.

Analysis of Feedback Precision, Recall, and Errors

The dependability of AI-generated feedback was assessed by analyzing precision, recall, and the rates of false positives and false negatives. In grammar and pronunciation, AI exhibited great precision (85%) and recall (82%), effectively identifying and rectifying the majority of faults. Nonetheless, in coherence and discourse organization, accuracy decreased to 70%, with false positives arising from the AI's misinterpretation of intricate sentence structures as faults. In evaluating pronunciation, AI encountered difficulties with non-native accents, occasionally identifying accurate pronunciations as errors. These findings indicate that although AI excels in rule-based assessments, it is deficient in context sensitivity during subjective judgments.

Challenges in Interpreting AI-Generated Feedback

Notwithstanding its benefits, AI-generated feedback presents difficulties in interpretability and usability. Students can see automatic corrections as ambiguous or excessively broad, necessitating human intervention for elucidation. Moreover, AI encounters difficulties with idiomatic terms, creative writing components, and complex language usage, resulting in erroneous recommendations. In speaking assessments, AI's voice recognition biases may disadvantage individuals with regional accents or atypical pronunciation patterns. These constraints underscore the necessity for hybrid AI-human feedback models, wherein AI facilitates technical adjustments but human

EduLite *Journal of English Education, Literature, and Culture*
Vol. 11, No. 1, February 2026, pp.67-85

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.11.1.67-85

evaluators deliver profound linguistic and contextual analysis, so assuring a thorough and pedagogically robust language assessment methodology.

Table 6. AI Feedback Accuracy and Effectiveness

Evaluation Criteria	AI Precision (%)	AI Recall (%)	False Positives (%)	False Negatives (%)
Grammar & Spelling	85	82	8	7
Pronunciation	82	79	10	9
Fluency	80	78	12	10
Coherence & Organization	70	68	15	17

Table 6 displays metrics for AI feedback accuracy, encompassing precision, recall, false positives, and false negatives across various evaluation criteria. Grammar and spelling had the highest precision (85%) and recall (82%), demonstrating the AI's robust capability to reliably identify and rectify errors. Pronunciation and fluency demonstrated good precision (82% and 80%, respectively), although recall values were marginally lower, indicating some overlooked errors. Coherence and organization had the lowest precision (70%) and recall (68%), indicating AI's challenges in evaluating speech structure and logical progression. The percentages of false positives and false negatives were most pronounced in coherence, highlighting AI's inadequacies in assessing intricate textual structures.

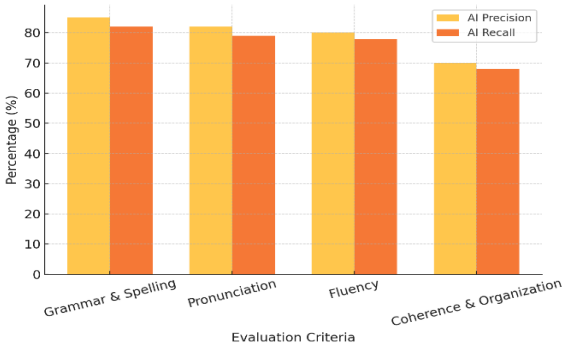

Figure 3. AI Precision vs. Recall in Feedback Accuracy

Figure 3 juxtaposes AI precision and recall across four evaluative criteria. Grammar and spelling obtained the most dependable AI comments, with great precision and recall metrics. Pronunciation and fluency evaluations were generally precise, however certain errors were overlooked in the memory study. Coherence and organization exhibited the lowest AI precision and recall, indicating difficulties in comprehending logical structure and discourse quality. The graphic indicates that AI excels in technical error detection but falters in subjective language elements, underscoring the necessity for human oversight in evaluating writing and speaking abilities.

How to Cite (APA Style):
Asrifan, A., Oliveira de Barros Cardoso, LM & Vargheese, KJ. (2026). Automated feedback for speaking and writing skills: Deep learning in English language assessment. *EduLite: Journal of English Education, Literature, and Culture*, *11* (1), 67-85. http://dx.doi.org/10.30659/e.11.1.67-85

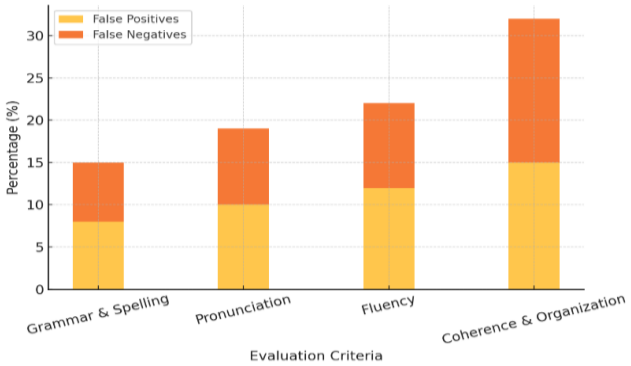

Figure 4. False Positive vs. False Negative Rates in AI Feedback

Figure 4 illustrates the frequencies of false positives and false negatives in AI-generated feedback. Grammar and spelling exhibited the lowest error rates, indicating a high reliability of AI in identifying errors. Pronunciation and fluency evaluations exhibited moderate rates of false positives and negatives, frequently misclassifying accent differences as inaccuracies. Coherence exhibited the largest false negative rate (17%), signifying that AI often failed to identify logical flaws in text. The figure indicates that although AI proficiently detects rule-based errors, it encounters difficulties with discourse-level analysis, underscoring the necessity for hybrid AI-human assessment models to enhance language evaluation.

Learner Perceptions and Engagement

The research investigated student perceptions and involvement with AI-generated feedback using questionnaires and interviews. The findings reveal that the majority of participants perceived AI input as advantageous, especially for grammatical correction, pronunciation enhancement, and fluency advancement. Seventy-eight percent of students concurred that AI-assisted feedback was beneficial in error identification, whereas seventy-two percent believed it enhanced writing and speaking clarity. Nevertheless, 35% of students articulated worries about the absence of contextual awareness and too general responses. Participants valued the immediate response but observed that some intricate errors necessitated additional human clarification.

Learner Engagement in Revision and Improvement Cycles

Students exhibited significant involvement with AI feedback, employing it to amend and enhance their work. Approximately 81% of participants utilized AI recommendations to implement adjustments, especially in grammar and pronunciation. Nonetheless, merely 60% integrated AI feedback for coherence and idea development, underscoring its constraints in structural guiding. Students participating in several revision cycles demonstrated enhanced improvement, validating AI's function in promoting iterative learning.

Participants' Views on Feedback Usefulness and Reliability

During interviews, pupils recognized AI as a beneficial adjunct tool, although underscored that it ought not to supplant human feedback. Sixty-eight percent of individuals trusted AI for grammar and spelling corrections, although

EduLite Journal of English Education, Literature, and Culture
Vol. 11, No. 1, February 2026, pp.67-85

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.11.1.67-85

expressed skepticism regarding its evaluations of coherence. Certain students perceived AI's rating as uneven, particularly in assessing essay arrangement and natural speech patterns. Notwithstanding these constraints, the majority of learners appreciated AI's accessibility, instantaneous feedback, and capacity for self-enhancement. The research indicates that AI-generated feedback improves language acquisition, particularly when supplemented by human educators, thereby providing more precise, nuanced, and pedagogically effective language evaluations.

Table 7. Learner Perceptions and Engagement Survey Results

Survey Question	Percentage of Agreement (%)
AI feedback helped identify errors	78
AI contributed to better writing and speaking clarity	72
AI feedback lacked contextual awareness	35
AI feedback was overly generic	35
AI feedback was useful for grammar and pronunciation	81
AI feedback was useful for coherence and idea development	60
AI corrections were trusted for grammar and spelling	68
AI feedback scoring was inconsistent	50

Table 7 displays survey responses concerning students' perceptions of AI-generated feedback. Seventy-eight percent of learners concurred that AI feedback was successful in error identification, whereas seventy-two percent deemed it beneficial for enhancing writing and speaking clarity. Nonetheless, 35% of individuals contended that AI exhibited a deficiency in contextual awareness and rendered excessively generic responses. 81% utilized AI recommendations for grammar and pronunciation corrections, whereas just 60% depended on AI comments for coherence and concept development. Although 68% of individuals trust AI for grammar and spelling improvements, 50% perceive its scoring as unreliable. These findings indicate that AI is an invaluable instrument, yet necessitates human oversight.

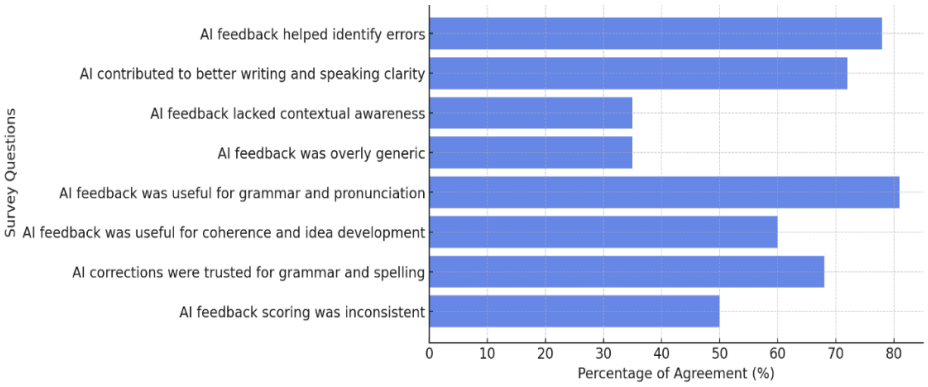

Figure 5. Learner Perceptions on AI Feedback

Figure 5 shows how students view comments produced by artificial intelligence, therefore stressing both advantages and drawbacks. With great agreement percentages in grammar and pronunciation improvement (81%), artificial intelligence was generally embraced for error detection and language clarity.

Students were less sure, meanwhile, about how well artificial intelligence could evaluate discourse structure and coherence. The modest trust levels (68%) for grammar fixes imply that although artificial intelligence is useful, students still want human confirmation for accuracy. The 35% worry about contextual constraints emphasizes AI's incapacity to offer deeper language analysis, therefore stressing the requirement of a hybrid AI-human feedback strategy in language evaluation.

Discussion

Effectiveness of AI-Based Feedback in Language Learning

This study shows that by offering instantaneous, data-driven corrections on grammar, fluency, and pronunciation, AI-generated feedback greatly enhances both writing and speaking abilities. Along with pronunciation and fluency in speaking, the quantitative findings revealed considerable gains in grammatical accuracy, lexical richness, and coherence in writing. This complements earlier studies implying that by providing instantaneous, customized feedback, AI-driven assessment systems can improve self-regulated learning by allowing students to recognize and quickly fix mistakes (Biju et al., 2024; Dong, 2023; Fleckenstein et al., 2023).

Strengths and Weaknesses Compared to Human Feedback

Among other benefits over human evaluation, AI feedback provides speed, consistency, and scalability (Alazemi, 2024; Bulut & Beiting-Parrish, 2024). Unlike human assessors, artificial intelligence can rapidly examine and offer corrections free from subjectivity or tiredness. This guarantees equity and consistency in evaluation, thereby lessening of grade and feedback distribution bias. AI lets students edit their work several times as well, therefore encouraging iterative learning and self-correction techniques.

Particularly in assessing higher-order language skills like argumentation, coherence, and critical thinking (Hackl et al., 2023; Hopfenbeck et al., 2023; Vasconcelos & dos Santos, 2023), artificial intelligence has clear shortcomings. AI struggles to evaluate creativity, discourse structure, and contextual appropriateness even if it catches grammar and pronunciation mistakes rather well. Human comments, on the other hand, offers more complex linguistic insights, tailored explanations, closer conceptual grasp, and constructive criticism going beyond rule-based corrections.

Scalability and Efficiency of AI-Based Assessment

A significant advantage of AI is its scalability; it can evaluate numerous learners concurrently without necessitating substantial human resources. AI-driven technologies are especially advantageous in automated grading systems, language acquisition platforms, and online evaluations, rendering them suitable for massive open online courses (MOOCs) and extensive proficiency tests (Grájeda et al., 2024; Rukiati et al., 2023). A hybrid method that integrates AI's efficiency with human judgment is crucial to mitigate its limits and establish a complete, pedagogically robust language assessment framework.

Pedagogical Implications of AI in Language Assessment

EduLite Journal of English Education, Literature, and Culture
Vol. 11, No. 1, February 2026, pp.67-85

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.11.1.67-85

The Role of AI as a Supplementary Assessment Tool

AI-driven feedback has demonstrated efficacy as a supplemental instrument in language evaluation, especially in automated error identification, fluency analysis, and grammar correction (Alnufaie & Alzahrani, 2024; Nazari et al., 2021). AI solutions deliver instantaneous, data-informed feedback, allowing learners to recognize deficiencies and enhance their language proficiency. In contrast to conventional approaches that depend exclusively on instructor input, AI facilitates continual learning and self-correction, rendering it an invaluable asset in both autonomous learning and formal education. Nonetheless, although AI improves efficiency, it cannot entirely supplant human assessment, particularly in intricate linguistic activities necessitating profound understanding, creativity, and critical analysis.

Challenges in Developing Higher-Order Thinking Skills through AI Feedback

A primary shortcoming of AI-driven feedback is its incapacity to evaluate higher-order cognitive skills, like argumentation structure, logical reasoning, and discourse coherence. Although AI proficiently rectifies superficial faults, it has difficulties in interpreting context, evaluating complex meaning, and delivering helpful criticism on rhetorical methods. AI models exhibit a deficiency in human-like judgment, complicating the assessment of persuasiveness, originality, and advanced linguistic proficiency (Getchell et al., 2022; Peters & Visser, 2023). Moreover, linguistic diversity and contextual discrepancies provide obstacles, since AI systems may misread non-standard statements or overlook innovative linguistic selections.

Integration Strategies for AI Feedback in Classroom Settings

To optimize AI's capabilities, instructors ought to implement a hybrid methodology that integrates AI-generated feedback with human instruction. Artificial intelligence can facilitate formative assessments, self-directed practice, and automated review processes, whereas educators deliver contextual explanations, individualized feedback, and comprehensive discourse analysis. The incorporation of AI in peer review processes, writing workshops, and pronunciation exercises can improve student engagement and educational outcomes (AlTwijri & Alghizzi, 2024; Baiq Sumarni et al., 2022; Rukiati et al., 2023; Syafryadin, 2022). Moreover, instructing students to critically analyze AI input guarantees their comprehension of mistakes instead of merely acquiescing to corrections, so promoting independent learning and metacognitive awareness.

Limitations and Challenges

Biases in AI Feedback, Particularly for Non-Native English Speakers

A key disadvantage of AI-driven feedback is its bias toward native English speakers, given that most AI models are trained on extensive datasets predominantly featuring standard native English variations. This poses difficulties for non-native English learners, whose accents, syntactic constructions, and lexical selections may not correspond with the AI's training data. Consequently, erroneous flags are raised, especially in pronunciation and

grammar evaluations, while genuine faults may be disregarded if they do not conform to the AI's programmed language frameworks. Furthermore, AI may neglect regional language variances, cultural effects, and multilingual interference, rendering it less flexible to varied student backgrounds.

AI's Difficulty in Assessing Creativity, Argumentation, and Discourse Coherence

Artificial intelligence excels at rule-based language evaluations, including grammar correction, fluency analysis, and sentence structure enhancement. Nonetheless, it encounters difficulties in assessing originality, reasoning, and discourse coherence, which necessitate advanced cognitive abilities and contextual understanding. Artificial intelligence lacks profound comprehension capabilities, rendering it poor in evaluating persuasive writing, rhetorical techniques, and critical analysis in essays. Furthermore, subtle meanings, tones, and intentions are frequently misconstrued, resulting in vague or deceptive feedback. This issue is particularly apparent in speaking assessments, when AI may prioritize pronunciation correctness at the expense of prosody, conversational flow, and pragmatic competence.

Technical Limitations in Speech Recognition and Natural Language Processing

Despite progress in Natural Language Processing (NLP) and Automatic Speech Recognition (ASR), AI-driven feedback systems continue to encounter technological obstacles. AI voice recognition technologies have difficulties with varied accents, fluctuations in speech tempo, and ambient noise, resulting in misinterpretations or erroneous transcriptions. Complex phrase structures and figurative language in writing frequently perplex AI models, resulting in inconsistent or irrelevant input. Moreover, AI's constrained capacity to monitor progress over time precludes it from delivering longitudinal evaluations or adaptive learning experiences, hence limiting its efficacy in individualized language development.

Future Research Directions

Future research should improve AI adaptation to multiple linguistic origins by including various accents, multilingual influences, and regional language characteristics into training datasets to mitigate bias against non-native speakers. Enhancing Natural Language Processing models to more effectively identify non-standard structures would augment feedback precision for ESL learners. Moreover, hybrid AI–human feedback models necessitate additional research to reconcile automated error detection with human proficiency in assessing coherence, creativity, and contextual significance. Ultimately, longitudinal studies are essential to assess the enduring effects of AI-assisted revision on the long-term development of speaking and writing skill in diverse educational contexts.

CONCLUSION

This study investigated the efficacy of deep learning–based automated feedback in enhancing the speaking and writing abilities of English learners in an EFL

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.11.1.67-85

scenario. The results indicate that AI-generated feedback markedly improves learners' grammatical accuracy, lexical diversity, coherence, fluency, pronunciation, and intelligibility. Quantitative results indicated statistically significant enhancements in pre-test and post-test scores for both writing and speaking aspects, hence affirming the beneficial instructional effect of AI-assisted revision cycles.

Notwithstanding these advancements, the study also recognized significant limits. Although AI systems demonstrated great precision and recall in rule-based evaluations like grammar and pronunciation, inconsistencies arose in the assessment of higher-order language qualities, such as discourse coherence, contextual relevance, and argumentation quality. These findings suggest that AI is proficient in superficial error identification but is comparatively ineffective in assessing intricate verbal competence. Moreover, learner perception data indicated a significant appreciation for AI's immediacy and accessibility, while also underscoring concerns about contextual awareness and scoring consistency.

The findings indicate that AI-generated feedback is more effective as an adjunct assessment tool than as a substitute for human assessors. A hybrid AI–human feedback model is thus advocated to merge the efficiency and scalability of automated systems with the interpretive depth and contextual sensitivity of expert educators. Future research should investigate the long-term effects of AI-assisted editing methods and provide more inclusive training datasets to mitigate bias in pronunciation and discourse assessment. The study provides empirical information to guide the ethical and pedagogically sound incorporation of AI technologies in language assessment.

AUTHOR STATEMENTS

Andi Asrifan conceptualized the study, designed the research framework, conducted data collection and statistical analysis, and drafted the manuscript. **Luís Miguel Oliveira de Barros Cardoso** contributed to data interpretation, methodological refinement, and critical revision of the manuscript. **K.J. Vargheese** supported the development of the literature review, the theoretical framing, and the manuscript editing for academic clarity. All authors reviewed, approved, and agreed to the final version of the manuscript

ACKNOWLEDGEMENTS

The authors would like to express their sincere appreciation to all participants who voluntarily contributed to this study.

.

REFERENCES

Alazemi, A. F. T. (2024). Formative assessment in artificial integrated instruction: delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Language Testing in Asia, 14*(1), 44. https://doi.org/10.1186/s40468-024-00319-8

How to Cite (APA Style):
Asrifan, A., Oliveira de Barros Cardoso, LM & Vargheese, KJ. (2026). Automated feedback for speaking and writing skills: Deep learning in English language assessment. *EduLite: Journal of English Education, Literature, and Culture*, *11* (1), 67-85. http://dx.doi.org/10.30659/e.11.1.67-85

Alnufaie, M. R., & Alzahrani, I. H. (2024). EFL grammar learning strategy use: utilizing grammar learning strategy inventory in an Arabic Context. *Teaching English as a Second or Foreign Language--TESL-EJ*, *27*(4). https://doi.org/10.55593/ej.27108a6

AlTwijri, L., & Alghizzi, T. M. (2024). Investigating the integration of artificial intelligence in English as foreign language classes for enhancing learners' affective factors: A systematic review. *Heliyon*, *10*(10), e31053. https://doi.org/10.1016/j.heliyon.2024.e31053

Baiq Sumarni, Dharma Dev Bhatta, & Kho, S. F.-C. (2022). The use of total physical response in teaching vocabulary integrated with meaningful classroom interaction. *Journal of Language and Literature Studies*, *2*(1), 23–32. https://doi.org/10.36312/jolls.v2i1.710

Biju, N., Abdelrasheed, N. S. G., Bakiyeva, K., Prasad, K. D. V., & Jember, B. (2024). Which one? AI-assisted language assessment or paper format: an exploration of the impacts on foreign language anxiety, learning attitudes, motivation, and writing performance. *Language Testing in Asia*, *14*(1), 45. https://doi.org/10.1186/s40468-024-00322-z

Bulut, O., & Beiting-Parrish, M. (2024). The Rise of artificial intelligence in educational measurement: opportunities and ethical challenges. *Chinese/English Journal of Educational Measurement and Evaluation*, *5*(3). https://doi.org/10.59863/MIQL7785

Campino, J. (2025). Unleashing the transformers: NLP models detect AI writing in education. *Journal of Computers in Education*, *12*(2), 645–673. https://doi.org/10.1007/s40692-024-00325-y

Dong, Y. (2023). Revolutionizing academic English writing through AI-Powered pedagogy: practical exploration of teaching process and assessment. *Journal of Higher Education Research*, *4*(2), 52. https://doi.org/10.32629/jher.v4i2.1188

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*. https://doi.org/10.3389/frai.2023.1162454

Getchell, K. M., Carradini, S., Cardon, P. W., Fleischmann, C., Ma, H., Aritz, J., & Stapp, J. (2022). Artificial intelligence in business communication: the changing landscape of research and teaching. *Business and Professional Communication Quarterly*, *85*(1), 7–33. https://doi.org/10.1177/23294906221074311

Giglio, A. Del, & Costa, M. U. P. da. (2023). The use of artificial intelligence to improve the scientific writing of non-native English speakers. *Revista Da Associação Médica Brasileira*, *69*(9). https://doi.org/10.1590/1806-9282.20230560

Grájeda, A., Burgos, J., Córdova, P., & Sanjinés, A. (2024). Assessing student-perceived impact of using artificial intelligence tools: Construction of a synthetic index of application in higher education. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186X.2023.2287917

Gu, L., Davis, L., Tao, J., & Zechner, K. (2021). Using spoken language technology for generating feedback to prepare for the TOEFL iBT® test: a user perception study. *Assessment in Education: Principles, Policy & Practice*, *28*(1), 58–76. https://doi.org/10.1080/0969594X.2020.1735995

Guo, Q., Feng, R., & Hua, Y. (2022). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted*

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.11.1.67-85

Language Learning, *35*(9), 2312–2331. https://doi.org/10.1080/09588221.2021.1879161

Hackl, V., Müller, A. E., Granitzer, M., & Sailer, M. (2023). Is GPT-4 a reliable rater? Evaluating consistency in GPT-4's text ratings. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1272229

Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hancock, J., & Jung, M. F. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports*, *13*(1), 5487. https://doi.org/10.1038/s41598-023-30938-9

Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., & McGrane, J. A. (2023). Challenges and opportunities for classroom-based formative assessment and AI: a perspective article. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1270700

Jamshed, M., Abu Saleh Md Manjur Ahmed, Md Sarfaraj, & Wahaj Unnisa Warda. (2024). The impact of ChatGPT on English language learners' writing skills. *International Journal of Interactive Mobile Technologies (IJIM)*, *18*(19), 18–36. https://doi.org/10.3991/ijim.v18i19.50361

Khoiriyah, K. (2021). Flipping the classroom to enhance EFL students' listening skill. *Journal on English as a Foreign Language*, *11*(1), 21–41. https://doi.org/10.23971/jefl.v11i1.2010

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

Oxford, R. L., Gu, P., Gunning, P., & Hernández-González, T. (2024). Considerations in designing and validating the diagnostic inventory for self-regulated language learning (DISLL): status of the process. *Language Teaching Research Quarterly*, *41*, 4–20. https://doi.org/10.32038/ltrq.2024.41.02

Peters, T. M., & Visser, R. W. (2023). *The importance of distrust in AI* (pp. 301–317). https://doi.org/10.1007/978-3-031-44070-0_15

Rukiati, E., Wicaksono, J. A., Taufan, G. T., & Suharsono, D. D. (2023). AI on learning English: application, benefit, and threat. *Journal of Language, Communication, and Tourism*, *1*(2), 32–40. https://doi.org/10.25047/jlct.v1i2.3967

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1260843

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

Syafryadin, S. (2022). Actualizing learning English grammar strategy using new inventory for university students. *PAROLE: Journal of Linguistics and Education*, *12*(1), 25–35. https://doi.org/10.14710/parole.v12i1.25-35

Tang, X., Chen, H., Lin, D., & Li, K. (2024). Harnessing LLMs for multi-dimensional writing assessment: Reliability and alignment with human judgments. *Heliyon*, *10*(14), e34262. https://doi.org/10.1016/j.heliyon.2024.e34262

How to Cite (APA Style):
Asrifan, A., Oliveira de Barros Cardoso, LM & Vargheese, KJ. (2026). Automated feedback for speaking and writing skills: Deep learning in English language assessment. *EduLite: Journal of English Education, Literature, and Culture*, *11* (1), 67-85. http://dx.doi.org/10.30659/e.11.1.67-85

Vasconcelos, M. A. R., & dos Santos, R. P. (2023). Enhancing STEM learning with ChatGPT and Bing Chat as objects to think with: A case study. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), em2296. https://doi.org/10.29333/ejmste/13313

Woo, J. H., & Choi, H. (2021). Systematic review for AI-based language learning tools. *Journal of Digital Contents Society*, *22*(11), 1783–1792. https://doi.org/10.9728/dcs.2021.22.11.1783

Xi, X. (2023). Advancing language assessment with AI and ML–Leaning into AI is inevitable, but can theory keep up? *Language Assessment Quarterly*, *20*(4–5), 357–376. https://doi.org/10.1080/15434303.2023.2291488

Yong, Q. (2020). Application analysis of artificial intelligence in oral English assessment. *Journal of Physics: Conference Series*, *1533*(3), 032028. https://doi.org/10.1088/1742-6596/1533/3/032028

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.